UNIVERSITY OF CALIFORNIA
RIVERSIDE

CRISPR/Cas9-enabled Functional Genomic Editing in the Thermotolerant Yeast *Kluyveromyces Marxianus*

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemical and Environmental Engineering

by

Mengwan Li

December 2022

Dissertation Committee:
Dr. Ian Wheeldon, Chairperson
Dr. Robert Jinkerson
Dr. Yanran Li

The Dissertation of Mengwan Li is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgement

Chapter 2 of this dissertation has been published in the journal Biotechnology for Biofuels in 2020. I want to express my gratitude to the corresponding author of this paper, Dr. Ian Wheeldon, who also directed and supervised all the other projects in my PhD program. Without his support and guidance, I can never reach this point of qualified PhD graduation. I appreciate every face-to-face discussion we had about research projects before COVID and all of the virtual ones since 2020. They were very helpful to brainstorm new ideas to fulfill a comprehensive story, and more importantly, to keep my research on track. I learned a lot from him about how to efficiently convey a story to the audience and readers when we made figures and wrote papers together. Although he directed and supervised all my research, I was given spaces to outline papers I was interested in writing especially after promoted to senior PhD candidate. It made me become an independent researcher I am today who had opportunities to practice taking the major responsibility of a project before graduation. I am sure I will benefit from that for a life-long time. I would also like to thank the co-author Dr. Xuye Lang, who helped characterize a set of native-derived promoters. Thank you to co-authors Marcos Moran Cabrera, Sawyer De Keyser, and Dr. Xiyan Sun for all supporting experiments. Thank you to our wonderful collaborator Dr. Nancy Da Silva for her inspiring suggestions and comments to make this paper a real success. I would also like to acknowledge fundings from the National Science Foundation and Department of Energy to support all works in this dissertation.

Pursuing a PhD is never an easy thing. However, because my lab colleagues and friends always had my back, five years went very fast and all that remained were good memories. I benefited a lot from internal lab collaboration. Thank you to Adithya Ramesh, Varun Trivedi, and Aida Tafrishi for every constructive conversation we had together to develop the genome-wide functional screening platform and corresponding analysis pipelines mentioned in chapter 4 and 5. Thank you to Shuang Wei, my dear friend. She is always supportive whenever I need her hand. Our friendship lasts for five years and will not stop because of graduation. Thank you to my CEE class 2022, Musen Zhou, Qi Li, Ning Yu, Jiayan Shi, Ming Lei, Cheng Tan, Soham Shah, and Daniel White. It is my fortune to have you as companions along my PhD journey. I would also love to thank you to the UCR GPP-E program where my life in the US began. Thank you to Jun Wang, who made it possible for me to get to know this fantastic program. Thank you to Jun Yuan. Having you in the past five years makes my PhD life an unforgettable experience that I will cherish in the rest of my life.

Last but not the least, I want to sincerely express my deepest gratitude to my parents in China, for their love for me without expecting anything in return. They are open-minded and thoughtful, showing great respect to my life plan. I can never become the person of who I am today without being their daughter!

ABSTRACT OF THE DISSERTATION

CRISPR/Cas9-enabled Functional Genomic Editing in the Thermotolerant Yeast *Kluyveromyces Marxianus*

by

Mengwan Li

Doctor of Philosophy, Graduate Program in Chemical and Environmental Engineering
University of California, Riverside, December 2022
Dr. Ian Wheeldon, Chairperson

A multigene integration tool takes advantage of type II CRISPR/Cas9 induced breaks on the genome as a selection for rapid one-step integration of up to three gene expression cassettes. Using this tool, a full factorial library of *KmARO4*, *KmARO7*, and *KmPHA2*, each driven by three different promoters that span a wide expression range, was constructed for Shikimate pathway refactoring. With high expressions of the tyrosine-deregulated $KmARO4^{K221L}$, native *KmPHA2* and *KmARO10*, and medium expression of feedback insensitive $KmARO7^{G141S}$, the best strain achieves 1943 ± 63 mg/L 2-phenyl ethanol.

Our highly efficient chemical and heat shock induced method achieving 6×10^5 transformants and electroporation for 1.5×10^6, facilitate functional genomic studies by enabling adequate representation of a complex screening library with less host cells and transformed DNA. With the CRISPR/Cas9 system and efficient DNA transformation, we characterized a 4-fold single-guide RNA (sgRNA) library targeting both promoters and coding regions of the *K. marxianus* CBS 6556 genome. With culturing library

transformants in three carbon sources, glucose, lactose, and xylose, the differentially induced editing efficiencies of the same sgRNA make the genome-wide CRISPR/Cas9 library a powerful tool to elucidate gene regulations involved in glucose repression.

Knocking out *ALPHA3* and *KAT1* significantly avoids single cell mating type switching because of homothallism, which enables *K. marxianus* a much more robust platform for high throughput screens heavily dependent on differentially accumulated gene editing fitness from a population growing to cell confluence. An efficient CRISPR RNA (crRNA) library with adequate gene coverage was obtained using six CRISPR guide activity prediction algorithms provided by CHOPCHOP v3. In-house uniqueness assessment criteria were applied to minimize off-target effects and save the highly active to the maximum possible extent. This optimized library enabled us to identify the first set of essential genes for *Kluyveromyces* genus.

Overall, this dissertation developed stable *α*/*a* haploid strains for *K. marxianus* CBS 6556, advanced CRISPR-mediated genetic tools, and highly efficient transformation protocols. They all facilitate genetic engineering and system metabolic engineering in *K. marxianus*, as well as better understanding of its genetics and metabolism.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1 Background

Total domestic US revenues generated by biotechnology in 2012 achieved more than $324 billion, accounting for over 2% of the US GDP. Revenues from the industrial biotechnology reached at least $105 billion, with $66 billion in biochemicals. [1] After a decade, the industrial biotech market expanded to $364 billion in 2021, with the compound annual growth rate (CAGR) of about 15%, and is forecasted to remain a high CAGR of 10% in the next five years. The biochemical segment has been dominant in the industrial biotech market for the past decade, and is expected to command a market share of a third by 2027. [2]

Industrial biotechnology, enabling the use of microorganisms and their enzymes to generate chemicals and energy from renewable biowastes of forestry and agriculture, significantly benefits the conventional chemical industry because of reduced carbon emission. There is no doubt that industrial biotech will become an essential technology paradigm for sustainable economic growth to meet carbon neutrality. However, the current competitiveness of bioprocessing is rather limited because of the requirement for expensive feedstocks, high energy and water use to maintain optimal producing host growth. The ease of productivity loss due to contamination also makes it less preferable. [3] These technical challenges have been proven difficult to overcome because microbial metabolism unfortunately does not evolve to suit the practical outcomes desired by

human needs. Thus, when microbes were isolated from nature, their efficiency in producing a particular molecule in high quantity and purity was rather low.

Metabolic engineering has played an exceptional role in strain genetic improvement to enhance bioproduction. It enables rational genetic modification without accumulation of unfavorable mutations, as well as pinpoints the genetic change(s) causing a variation in phenotype. It also allows introduction of genes from foreign organisms to equip the host with novel traits. [4] While both bacteria and yeasts have a long history of application in the biotech industry, it is easier to functionally express certain enzymes, for example, heterologous cytochrome P450 anchored on the endoplasmic reticulum (ER) membrane from animal and plant cells, in yeasts than bacterial hosts, since yeasts share the same basic set of membrane-enclosed organelles with other more advanced eukaryotes. [5] Yeasts also provide more robust production platforms than bacteria, especially under typical stressful culturing environment in an industrial setting, such as low pH. [6] Moreover, biotic factors like phage contamination in bioreactors can cause detrimental consequences on biological processing if the host is a bacterial species such as *E. coli*. [7]

Due to its economic importance in traditional biotechnology such as baking, brewing, and winemaking, there was a historically long-lasting research focus on the yeast *Saccharomyces cerevisiae*. It was the first eukaryotic organism whose complete genome sequence was determined, and the first yeast species having a genome-wide functional deletion collection [8,9]. As a result, in-depth knowledge of its genetics, physiology, and biochemistry, together with a wealth of genetic engineering

technologies, have accumulated over the time. The availability of numerous specialized expression plasmids, genome editing toolboxes, optimized phenotypically selectable gene tags, efficient DNA transformation for high throughput screens, and high homologous recombination success for precise gene integration, has made *S. cerevisiae* a well-established eukaryotic model organism for fundamental studies in biological process to enhance specific native metabolisms or introduce heterologous pathways for special industrial applications. [10,11,12,13,14,15,16,17]

Significant success has been achieved to extend the boundaries of bioprocessing competitiveness by engineering *S. cerevisiae* [5]. Beside continuous strain optimization of this model yeast, non-conventional hosts are also coming into people's attention nowadays as an alternate solution to the remaining bioproduction conundrums despite of lacking understanding of their genetics and metabolism, since many of them have inherent beneficial phenotypes for industrial biochemical production, such as stress tolerance and flexible carbon source utilization. [3,18,19]

The advent of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) mediated genetic and metabolic engineering tools have revolutionized and expedited the non-conventional yeast domestication to obtain highly efficient and focused bio-factories that produce considerable amount of biochemicals of interest [19,20]. The flexible, efficient, and multiplexed functional type II CRISPR system allowed rational strain engineering of various non-model yeasts regardless of their intrinsic inefficient homologous recombination, or the lack of a good set of well-characterized native promoters and selection markers. [21,22,23,24,25,26,27,28,29]

The budding yeast *Kluyveromyces marxianus* is a good example for benefits of using non-conventional yeasts in industrial bioproduction [30]. Same as *S. cerevisiae*, it is also classified as a generally regarded as safe (GRAS) organism [31]. However, unlike the model yeast *S. cerevisiae*, toxic or unintended ethanol formation under aerobic condition can be avoided in *K. marxianus* even with high concentration of glucose because it is Crabtree-negative [32]. *K. marxianus* is thermotolerant to ~ 50 °C [33,34,35,36], and a growth rate double that of *S. cerevisiae* at 30 °C is beneficial to produce growth or biomass formation associated products with high TCA cycle flux for complete carbon oxidation [23,37,38]. Its broad substrate utilization spectrum enables sufficient conversion of sugars and organic acids hydrolyzed from lignocellulosic biomass to products with much higher commercial values [39,40]. Finally, *K. marxianus* is able to use some genetic parts, including promoters and terminators, from *S. cerevisiae*, both of which are part of the saccharomyces subgenera within hemiascomycetes [23,41,42,43].

1.2 Thesis Organization

One of the pressing challenges in using *K. marxianus* as an industrial host is the lack of well- characterized synthetic biology tools to precisely edit the genome and rapidly engineer multi-enzymatic pathways. In chapter 2 we developed a multigene integration tool that uses CRISPR/Cas9 induced breaks on the genome as a selection for one-step integration of an insert that encodes one, two, or three gene expression cassettes. This metabolic engineering tool is applied to construct a full factorial library of

KmARO4, *KmARO7*, and *KmPHA2* on Shikimate pathway driven by three characterized native-derived *K. marxianus* promoters that span a wide expression range at variable temperatures to successfully enhance the bioproduction of an aromatic amino acid derived alcohol, 2-phenylethanol (2-PE).

The flexible, efficient, and multiplexed functional type II CRISPR system revolutionized and expedited functional genomic studies in non-conventional yeasts. Genome-wide screening necessitates highly efficient transformation methods. Therefore, in chapter 3 we optimized the chemical and heat shock induced transformation to obtain a seven-times higher efficiency compared with the current, for 6×10^5 transformants per reaction in average, using 1.04 pmol DNA (equal to 4 µg for 6,275-bp plasmids) and 5×10^8 of cells. The optimized electroporation achieved 1.5×10^6 transformants per reaction on average, using 0.26 pmol DNA (equal to 1 µg for 6,275-bp plasmids) and 8.4×10^8 of cells.

Chapter 4 describes our first work of CRISPR/Cas9 mediated genome-wide screening in *K. marxianus*. It showed that highly efficient crRNA sequences only account for a small portion of all 20-nt sequences with an NGG motif at the 3' end. We showed that the variation of apparent CRISPR/Cas9 editing efficiencies using the same guide sequence is a good metric to quantify chromatin accessibility, thus can be a sensitive probe for specific epigenetic effects. Because of the NGG requirement as well as crucial crRNA-DNA binding within the seed region (NGG-proximal 10 to 12 bp) for functional Cas9 cleavage, the differentially CRISPR/Cas9 enabled gene editing efficiencies can be

used as a screening tool to elucidate transcriptional regulation mediated by zinc finger proteins.

In chapter 5 we designed and validated an optimized whole genome crRNA library in *K. marxianus*. The on-target efficiencies were predicted by CHOPCHOP v3. As for the evaluation of guide specificity, we enhanced uniqueness criteria by adding "seed_MM0". Given that *K. marxianus* CBS 6556 has gene duplication and repeated elements even within some unique genes, reasonable and necessary uniqueness exemption was also applied in these special cases to keep as many highly active crRNAs as possible. Stable *a* haploids were also constructed and validated to be the robust screening platform. The first set of essential genes for the *Kluyveromyces* genus was obtained from our genome-wide deletion fitness screening.

1.3 References

[1] Carlson R. Estimating the biotech sector's contribution to the US economy. Nat Biotechnol. 2016;34:247-255.

[2] https://www.maximizemarketresearch.com/market-report/global-white-biotechnology-market/98059/ (last accessible date: September 03, 2022)

[3] Thorwall S, et al. Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. Nat Chem Biol. 2020;16(2),113–121.

[4] Nevoigt E. Progress in Metabolic Engineering of *Saccharomyces cerevisiae*. Microbiol Mol. 2008;72(3):379-412.

[5] Paddon CJ and Keasling JD. Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. Nat Rev Micro. 2014;12:355-367.

[6] Fletcher E, et al. RNA-seq analysis of *Pichia anomala* reveals important mechanisms required for survival at low pH. Microb Cell Fact. 2015;14:143.

[7] González-García VA, et al. Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor. J Biol Chem. 2015;290:jbc.M114.614222.

[8] Goffeau A, et al. Life with 6000 genes. Science. 1996;274(5287):546-567.

[9] Winzeler EA, et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science. 1999;285:901-906.

[10] Parent SA and Bostian KA. Recombinant DNA technology: yeast vectors, p121-178. In A. E. Wheals, A. H. Rose, and J. S. Harrison (ed.), 1995. The yeasts, vol. 6. Yeast genetics. Academic Press, London, United Kingdom.

[11] Gueldener UJ, et al. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. Nucleic Acids Res. 2002;30,e23.

[12] Guldener US, et al. A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res. 1996;24:2519-2524.

[13] Janke C, et al. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast. 2004;21:947-962.

[14] Sheff MA and Thorn KS. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. Yeast. 2004;21:661-670.

[15] Gietz RD and Woods RA. Genetic Transformation of Yeast. BioTechniques. 2001;30(4): 816-831.

[16] Klinner U and Schafer B. Genetic aspects of targeted insertion mutagenesis in yeasts. FEMS Microbiol Rev. 2004;28, 201-223.

[17] Hong KK and Nielsen J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. Cell Mol Life Sci. 2012;69:2671-2690.

[18] Wagner JM and Alper HS. Synthetic biology and molecular genetics in non-conventional yeasts: current tools and future advances. Fungal Genet Biol. 2016;89:126-136.

[19] Löbs AK, et al. Genome and metabolic engineering in non-conventional yeasts: current advances and applications. Synth Syst Biotechnol. 2017;2(3):198-207.

[20] Donohoue PD, et al. Advances in Industrial Biotechnology Using CRISPR-Cas Systems. Trends Biotechnol. 2018;36(2):134-146.

[21] Schwartz C, et al. Standardized Markerless Gene Integration for Pathway Engineering in *Yarrowia lipolytica*. ACS Synth Biol. 2016;6(3):402-409.

[22] Schwartz C, et al. CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*. Biotechnol. Bioeng. 2017;114(12):2896-2906.

[23] Löbs AK, et al. CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol Biofuels. 2017;10(164).

[24] Löbs AK, et al. Highly Multiplexed CRISPRi Repression of Respiratory Functions Enhances Mitochondrial Localized Ethyl Acetate Biosynthesis in *Kluyveromyces marxianus*. ACS Synth Biol. 2018;7, 2647-2655.

[25] Schwartz C, et al. Multiplexed CRISPR Activation of Cryptic Sugar Metabolism Enables *Yarrowia Lipolytica* Growth on Cellobiose. Biotechnol J. 2018;13(9):1700584.

[26] Cao M, et al. CRISPR–Mediated Genome Editing and Gene Repression in *Scheffersomyces stipitis*. Biotechnol J. 2018;13(9):1700598.

[27] Schwartz C, et al. Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast *Yarrowia lipolytica*. Metab Eng. 2019;55:102-110.

[28] Tran, V. G., et al. (2019). Development of a CRISPR/Cas9-Based Tool for Gene Deletion in *Issatchenkia orientalis*. mSphere. 2019;4(3):e00345-00319.

[29] Dalvie NC, et al. Host-Informed Expression of CRISPR Guide RNA for Genomic Engineering in *Komagataella phaffii*. ACS Synth Biol. 2020;9(1):26-35.

[30] Fonseca GG, et al. The yeast *Kluyveromyces marxianus* and its biotechnological potential. Appl Microbiol Biotechnol. 2008;79:339-354.

[31] Lane M and Morrissey J. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. Fungal Biol Rev. 2010;24:17-26.

[32] Malina C, et al. Adaptations in metabolism and protein translation give rise to the Crabtree effect in yeast. Proc. Natl. Acad. Sci. U.S.A. 2021;118(51):e2112836118.

[33] Abdel-Banat BMA, et al. High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? Appl Microb Biotechnol. 2010;85(4):861-867.

[34] Nonklang S, et al. High-Temperature Ethanol Fermentation and Transformation with Linear DNA in the Thermotolerant Yeast *Kluyveromyces marxianus* DMKU3-1042. Appl and Environ Microbiol. 2008;74(24):7514.

[35] Medeiros ABP, et al. Aroma compounds produced by *Kluyveromyces marxianus* in solid state fermentation on a packed bed column bioreactor. World J Microbiol Biotechnol. 2001;17(8):767-771.

[36] Nambu-Nishida Y, et al. Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. Sci. Rep. 2017;7 (8993).

[37] Groeneveld P, et al. Super life – how and why 'cell selection' leads to the fastest-growing eukaryote. FEBS J. 2009;276(1):254-270.

[38] Fonseca GG, et al. Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with glucose as the sole carbon source. FEMS Yeast Res. 2007;7(3):422-435.

[39] Löbs AK, et al. High throughput, colorimetric screening of microbial ester biosynthesis reveals high ethyl acetate production from *Kluyveromyces marxianus* on C5, C6, and C12 carbon sources. Biotechnol J. 2016;11(10):1274-1281.

[40] Hillman ET, et al. Hydrolysis of lignocellulose by anaerobic fungi produces free sugars and organic acids for two-stage fine chemical production with *Kluyveromyces marxianus*. Biotechnol Progress. 2021;37(5):e3172.

[41] Lee KS, et al. Characterization of *Saccharomyces cerevisiae* promoters for heterologous gene expression in *Kluyveromyces marxianus*. Appl Microbiol Biotechnol. 2013;97(5), 2029-2041.

[42] Lang X, et al. Developing a broad-range promoter set for metabolic engineering in the thermotolerant yeast *Kluyveromyces marxianus.* Metab Eng Commun. 2020;11:e00145.

[43] Belloch C, et al. Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome-c oxidase II gene. Int J Syst Evol Microbiol. 2000;50(1):405-416.

Chapter 2: CRISPR-mediated Multigene Integration for 2-PE Biosynthesis Enhancement

2.1 Abstract

We develop a multigene integration tool that uses CRISPR-Cas9 induced breaks on the genome as a selection for the one-step integration of an insert that encodes one, two, or three gene expression cassettes. Integration of a 5-kbp insert containing three overexpression cassettes successfully occurs with an efficiency of 51 ± 9% at the *ABZ1* locus and was used to create a library of *K. marxianus* CBS 6556 strains with refactored Shikimate pathway genes. The 3^3-factorial library includes all combinations of *KmARO4*, *KmARO7*, and *KmPHA2*, each driven by three different promoters that span a wide expression range. Analysis of the refactored pathway library reveals that high expression of the tyrosine-deregulated $KmARO4^{K221L}$ and native *KmPHA2*, with the medium expression of feedback insensitive $KmARO7^{G141S}$, results in the highest increase in 2-PE biosynthesis, producing 684 ± 73 mg/L. Ehrlich pathway engineering by overexpression of *KmARO10* and disruption of *KmEAT1* further increases 2-PE production to 766 ± 6 mg/L. The best strain achieves 1943 ± 63 mg/L 2-PE after 120 h of fed-batch operation in shake flask cultures.

2.2 Introduction

Like many esters and alcohols produced during yeast fermentation, the aromatic alcohol 2-phenylethanol (2-PE) is used in a wide variety of applications. Its rose-like aroma is used to add flavor and fragrance to food, perfumes, and cosmetics [1]. The high

energy density and compatibility with gasoline blends also makes 2-PE a promising next-generation biofuel [2]. The worldwide flavor and fragrance market was valued at upward of $US20 billion in 2018 and is expected to expand over the next decade (https://www.grandviewresearch.com/industry-analysis/flavors-fragrances-market). High purity, food-grade 2-PE is still produced mainly by isolation from rose petals, which presents a technically challenging separation problem, and the supply is subject to annual fluctuations in crop yields [3]. The end result is a high market price of ~$US1000 per kg. Synthesis by chemical catalysis is possible, but typically produces 2-PE that sells for less than $US4 per kg because it is not suitable for human consumption or use [4]. The chemical catalysis route also creates environmental and health challenges due to the reliance on petrochemical reagents [5]. As a metabolic intermediate formed via the Shikimate and Ehrlich pathways, 2-PE produced by microbial systems is a potentially sustainable alternative to fossil fuel-based production and isolation from native plants.

The Shikimate pathway is broadly conserved across various plants and microbes, producing aromatic amino acids from erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP; Figure 2.1). Chorismate is the first branch point, with one branch leading to tryptophan and the other to prephenate. One pathway output from prephenate is L-tyrosine, synthesized by prephenate dehydrogenase and aromatic aminotransferase. L-phenylalanine biosynthesis begins with prephenate as well, but prephenate is first converted to phenylpyruvate (PP) prior to transamination. Decarboxylation of phenylpyruvate, the beginning of the Ehrlich pathway, produces phenylacetaldehyde, with subsequent catalysis by aromatic alcohol dehydrogenase

resulting in 2-PE. Acetylation synthesizes the corresponding ester, 2-phenylethyl acetate (2-PEAc), by alcohol acetyltransferase (ATTase or Atf) [6-8]. These biosynthetic pathways have been exploited to overproduce both 2-PE and 2-PEAc. Relieving tyrosine feedback inhibition at key nodes in the Shikimate pathway has produced substantial gains in 2-PE and 2-PEAc titers [9]. Overexpression of homologous and heterologous genes of the Ehrlich pathway, including phenylpyruvate decarboxylase and alcohol dehydrogenase (or phenylacetaldehyde reductase, PAR) is known to increase titer [10-12]. Finally, *in situ* product removal also has been shown to increase 2-PE titers beyond 2 g/L by alleviating product toxicity [13]. *Kluyveromyces marxianus* natively has strong metabolic pathways to synthesize aromatic compounds [14], thus we use *K. marxianus* as the host for 2-PE biosynthesis in this work.

As is common for many non-conventional yeasts, there has been substantially less effort put towards creating advanced genetic and metabolic engineering tools than there has been for *S. cerevisiae*. Targeted gene integration is hampered by a low capacity for homologous recombination (HR); however, the widespread adoption of type II CRISPR systems for genome editing has helped alleviate this problem [15]. Targeted gene disruption, integration, and regulation have been made possible by CRISPR-Cas9 and CRISPR activation and interference (CRISPRa/i) systems for various non-conventional yeasts [16-20]. CRISPR-based genome editing has also been demonstrated in *K. marxianus*, but efficient multigene integration for pathway refactoring [21, 22] has not yet been established [7, 23, 24, 25].

Here, we engineered new strains of *K. marxianus* CBS 6556 with increased flux along the Shikimate and Ehrlich pathways, and high titer biosynthesis of 2-PE. To accomplish this, we first developed a CRISPR-Cas9-mediated multigene integration system that enabled simultaneous integration of up to three expression cassettes into a single, targeted genomic locus. Multigene integration was used to create a 3^3 combinatorial library with variable expression of three key Shikimate pathway genes, *ARO4*, *ARO7* and *PHA2*. The elimination of tyrosine feedback inhibition to *KmARO4* and *KmARO7*, the deletion of parasitic pathways including *KmEAT1* acetylation of 2-PE to 2-PEAc, and the overexpression of *KmARO10* belonging to the Ehrlich pathway, resulted in enhanced 2-PE production. Given the broad temperature range of *K. marxianus*, we also explored the effects of increased temperature. The highest 2-PE titer was achieved at 30 °C, with high expression of $KmARO4^{K221L}$, *KmPHA2*, and *KmARO10*, medium-level expression of $KmARO7^{G141S}$, disruption of functional *KmEAT1*, and cultures operated in fed-batch mode.

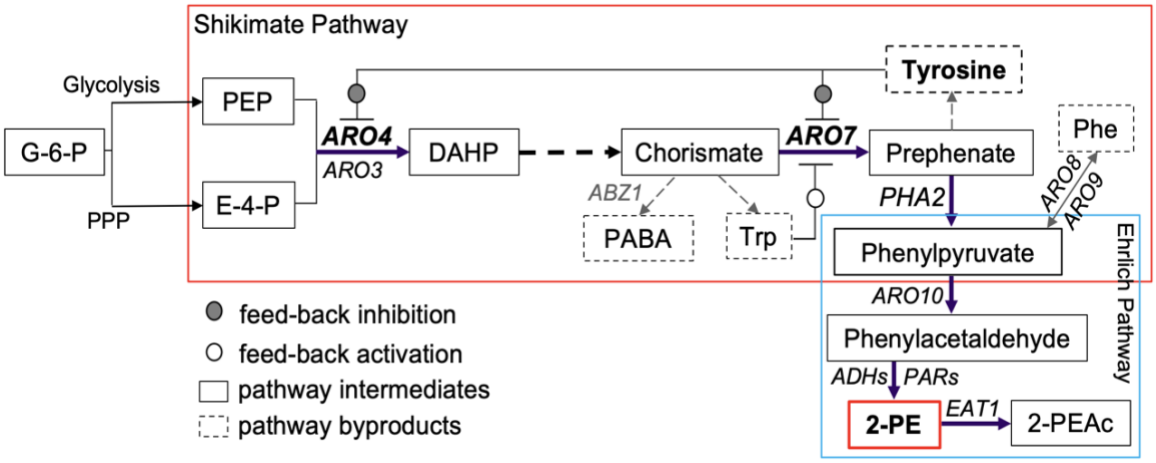

Figure 2.1. Simplified diagram of the Shikimate and Ehrlich pathways in *K. marxianus* for 2-phenylethanol biosynthesis. Arrows with dashed lines represent more than one reaction. G-6-P: glucose 6-phosphate; PEP: phosphoenolpyruvate; E-4-P: erythrose-4-phosphate; DAHP: 3-deoxy-D-arabinoheptulosonate-7-phosphate; PABA: para-aminobenzoate; 2-PE: 2-phenylethanol; 2-PEAc: 2-phenylethyl acetate. *ARO3/ARO4*: DAHP synthase; *ABZ1*: PABA synthase; *ARO7*: chorismate mutase; *PHA2*: prephenate dehydratase; *ARO8/ARO9*: aromatic aminotransferase; *ARO10*: phenylpyruvate decarboxylase; *ADH*: alcohol dehydrogenase; *PAR*: phenylacetaldehyde reductase; *EAT1*: ethanol acetyltransferase.

2.3 Results

Given the importance of the Shikimate pathway to 2-PE production, we sought to increase flux to phenylpyruvate (PP), the precursor to the Ehrlich pathway, via Shikimate pathway refactoring. To accomplish this in *K. marxianus*, we first needed to identify a series of variable strength promoters, and secondly, create a multigene integration tool. We recently designed and characterized a set of *K. marxianu*s promoters with a wide expression range under glucose metabolism, and here elected P_{KmTEF3}, P_{KmPGK} and P_{KmTDH3} to access high, medium, and low levels of gene expression in the refactored library [26].

Based on the design of a CRISPR-Cas9-mediated gene integration tool that we previously created for the oleaginous yeast *Y. lipolytica*, we designed a two-plasmid system for targeted gene integration in *K. marxianus* [27]. The CRISPR plasmid, pCRISPR, expressed a single guide RNA (sgRNA) along with Cas9. The homology donor plasmid, pHD, with 700 bp up- and down-stream homology to the targeted site was used for the integration of one, two, and three genes into a single locus. A schematic of the system, and the homology donor constructs for single, dual, and triple gene integration are shown in Figure 2.2A. Also shown is the three-primer verification test for chromosomal integration.

Prior to expanding the system for multiple genes, we identified five different genomic sites that were suitable for high efficiency integration. *URA3*, *XYL2*, *ABZ1*, *SDL1* and *LYS1* were selected, and at least one sgRNA sequence that resulted in high efficiency integration was identified. sgRNA sequences for each targeted gene are presented in Table S2.1. The *URA3* locus was selected because wild type function was previously eliminated when creating the base strain *K. marxianus* CBS 6556 *ura3Δ his3Δ* [23]. *ABZ1* encodes aminodeoxychorismate synthase, which activates a parasitic reaction diverting chorismate to para-aminobenzoate (PABA; Figure 2.1). *LYS1* was selected because it was previously identified as a hotspot for linear DNA integration [28], while *SDL1* is a non-essential gene in *K. marxianus*, and the gene knockout is insensitive to growth at elevated temperature [29]. Finally, *XYL2* was included in the set of potential integration sites because it facilitates rapid phenotypic screening for successful integrations and we have previously identified an sgRNA with high efficiency [23].

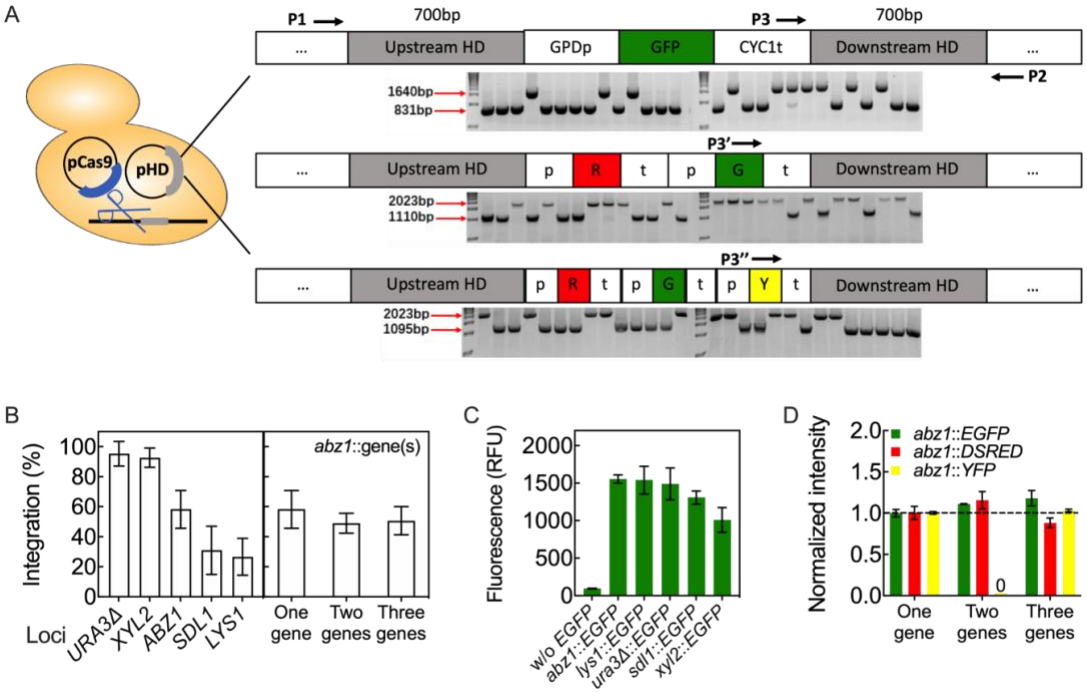

Figure 2.2. Markerless CRISPR-Cas9-mediated multigene integration in *K. marxianus* CBS 6556. (A) Schematic representation of the three-primer colony PCR test and example gels for single and multiple gene integrations into the *ABZ1* locus. The length of the up- and down-stream homology regions is 700 bp. (B) Integration efficiency of single (*EGFP*), dual (*EGFP*, *DSRED*), and triple (*EGFP*, *YFP*, and *DSRED*) genes. (C) Relative fluorescence intensity of *EGFP* expressed from different loci as measured by fluorescence microplate reader. (D) Normalized expression level of fluorescent proteins when produced individually or simultaneously with one or two additional fluorescent proteins. All experiments were performed in biological triplicate. Bars represent the mean, while error bars represent the standard deviation.

Heterologous gene integration at each of the selected loci was tested using fluorescent reporters *EGFP*, *YFP*, and *DSRED* (Figure 2.2B). The highest integration efficiency of a single gene was 95 ± 8%, observed at the *URA3* locus. The second highest, 93 ± 6%, occurred at *XYL2*. For *ABZ1*, *SDL1* and *LYS1*, gene integration was achieved with efficiencies of 58 ± 13%, 31 ± 16%, and 27 ± 12%, respectively. Given these results, *ABZ1* was chosen as the primary site to develop the multigene integration

system for Shikimate pathway refactoring because it is parasitic to flux along the pathway and because integrations were achieved with an efficiency of nearly 60%. Integration of an insert encoding expression cassettes for *EGFP* and *DSRED* occurred with 49 ± 7% of tested colonies, while an insert encoding all three reporter genes occurred at 51 ± 9% efficiency. Of note, shorter homology lengths down to 300 bp led to a 68 ± 11% decrease of integration efficiency (Figure S2.1); therefore, 700 bp up- and downstream homology was used for all subsequent gene integrations.

Previous reports suggested that heterologous gene expression can differ depending on integration site within the yeast genome [27, 30]. Figure 2.2C compares the *EGFP* fluorescence from integration at the five tested sites, including *ABZ1*, *LYS1*, *URA3*, *SDL1*, and *XYL2*. One-way ANOVA suggested that there was a significant effect due to integration locus, with the effect driven primarily from low expression at the *XYL2* locus ($p = 0.0183$, $n = 3$, Welch's test). Expression from single, dual, and triple integrations was also tested (Figure 2.2D). Based on fluorescence measurements, expression of a single gene was equivalent to the expression of the same gene when integrated along with other reporter genes at the same locus, thus providing a genome editing tool that can be used to simultaneously refactor up to three genes of a desired pathway.

It has been shown that *S. cerevisiae ARO4* and *ARO7* are feedback inhibited by tyrosine, thus limiting Shikimate pathway flux. The inhibition effect is eliminated in the feedback insensitive variants $ScARO4^{K229L}$ and $ScARO7^{G141S}$ [9]. To assess potential feedback inhibition in *K. marxianus*, we overexpressed a series of wild type and mutant

ARO4 and *ARO7* genes from both *S. cerevisiae* BY4742 and *K. marxianus* CBS 6556. In synthetic defined medium, *K. marxianus* CBS 6556 *ura3Δ his3Δ* accumulated more ester (2-PEAc) than alcohol (2-PE; Figure S2). As such, for all plasmid-based overexpression experiments we used 2-PEAc titer to quantify the effect of mutations found in *S. cerevisiae* on the corresponding *K. marxianus* variants. 2-PEAc concentration reached a maximum after 18 h of cultivation in the selective medium at 30 °C, this time point was consequently used for enzyme screening. A K229L mutation in *ScARO4* alleviated the inhibition effect and resulted in an increase in 2-PEAc biosynthesis from 31 ± 7 to 209 ± 5 mg/L when overexpressed in *K. marxianus* CBS 6556 *ura3Δ his3Δ* (Figure 2.3). The *ScARO4* mutation mapped to K221L in *KmARO4*, overexpression of which also produced an increasing amount of 2-PEAc (202 ± 16 mg/L). Feedback inhibition alleviation of *ARO7* was tested with $ScARO7^{G141S}$ and $KmARO7^{G141S}$; however, no increase in 2-PEAc production was observed. Given that there was no difference in 2-PEAc biosynthesis between the overexpression of *K. marxianus* and *S. cerevisiae* homologs of *ARO4*, *ARO7*, and their corresponding mutants, we decided to use the regulation-insensitive *K. marxianus* variants, $KmARO4^{K221L}$ and $KmARO7^{G141S}$, in the refactoring experiments along with wild-type *KmPHA2*.

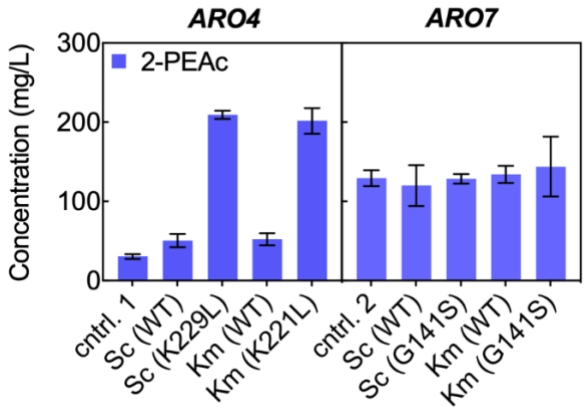

Figure 2.3. Shikimate pathway enzyme screening. *ARO4* and *ARO7* variants were overexpressed from plasmids with *S. cerevisiae TDH3* promoter (P_{ScTDH3}) and were screened for 2-phenylethyl acetate (2-PEAc) biosynthesis. 2-PEAc was quantified by GC-FID after 18 h cultivation at 30 °C from an initial OD_{600} of 0.05 in 25 mL SD-H medium. The first control strain (cntrl.1) was *K. marxianus* CBS 6556 *ura3Δ his3Δ* harboring a low-copy number empty vector. The second control strain (cntrl.2) is *K. marxianus* CBS 6556 *ura3Δ his3Δ abz1*::(P_{ScTDH3})*KmARO4*^{K221L}-(P_{ScTDH3})*KmARO10* harboring the empty vector. All experiments were performed in biological triplicates. Bars represent the mean, while error bars represent the standard deviation.

Pathway refactoring to balance and increase pathway flux requires promoters of variable strengths; as such, we selected three previously characterized promoters with high, medium, and low expression [26]. Promoters were cloned from the 700 bp upstream of start codons (ATG) of native *K. marxianus TEF3*, *PGK*, and *TDH3* genes to produce P_{KmTEF3}, P_{KmPGK}, and P_{KmTDH3}. Figure 2.4A shows the resulting *EGFP* fluorescence as measured by flow cytometry from plasmid overexpression at 30 °C. P_{KmTDH3} produced the lowest fluorescence, P_{KmPGK} exhibited a 2.6-fold increase in *EGFP* expression, and P_{KmTEF3} produced the highest expression, 7.8-fold above that produced from P_{KmTDH3}. A negative control without *EGFP* expression demonstrated a low fluorescence background, while a positive control with *EGFP* constitutively expressed from P_{ScTDH3} is provided for

comparison with a known high-level *S. cerevisiae* promoter that is functional in *K. marxianus* [31].

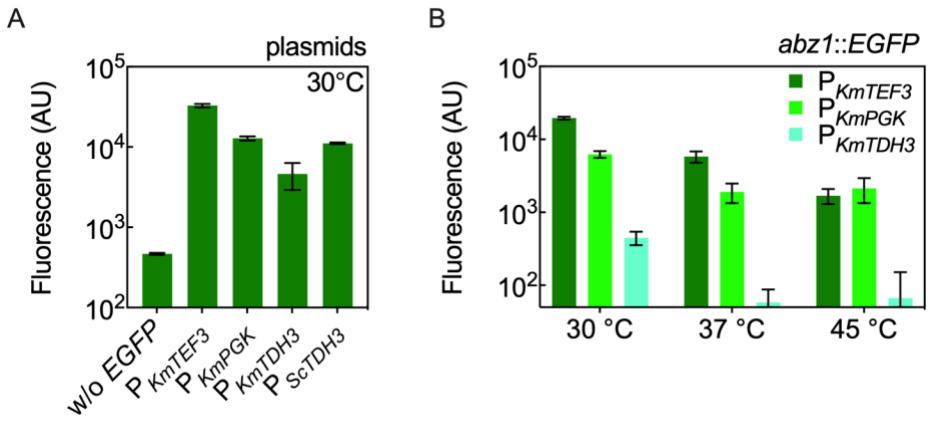

Figure 2.4. Promoters to construct the Shikimate pathway refactoring library. (A) Transcriptional strengths of constitutive promoters from *K. marxianus* and *S. cerevisiae* were quantified by overexpression of *EGFP* on a low copy-number plasmid. Three characterized promoters from *K. marxianus* CBS 6556 (P_{KmTEF3}, P_{KmPGK}, and P_{KmTDH3}) were used to construct the Shikimate pathway library. (B) The temperature effect on chromosomal expression of *EGFP* from variable strength promoters. Expression cassettes of *EGFP* driven by P_{KmTEF3}, P_{KmPGK} and P_{KmTDH3} were integrated into the genome of *K. marxianus* CBS 6556 *ura3Δ his3Δ* at the *ABZ1* locus. All fluorescence measurements were taken using flow cytometry. All experiments were performed in biological triplicates. Bars represent the mean of background subtracted fluorescence, while error bars represent the standard deviation.

Preliminary Shikimate pathway refactoring experiments suggested that 2-PEAc formation was repressed in *K. marxianus* CBS 6556 *ura3Δ his3Δ* when cultured in rich medium (YPD; Figure S2.2). In addition, plasmid expression significantly reduced growth rate at elevated temperatures (Figure S2.3). For these reasons we conducted all refactoring experiments for 2-PE biosynthesis in YPD medium, thus necessitating chromosomal gene integration. Importantly, the relative expression from the promoter set was maintained after integration (Figure 2.4B). With respect to evaluating promoters at higher temperatures, *EGFP* was found to be insensitive to temperature between 30 to 45

°C, and served as a reliable expression reporter for temperature effects (Figure S2.4). At 37 °C, the strengths of P_{KmTEF3} and P_{KmPGK} decreased by ~70%, while P_{KmTDH3} decreased expression by 88% from that observed at 30 °C. Heterologous gene expression driven by P_{KmTEF3} was further reduced at 45 °C, exhibiting only 10% of that at 30 °C. The P_{KmPGK} and P_{KmTDH3} expressions at 45 °C were comparable to those at 37 °C. Despite decreased expression at elevated temperatures, the promoter set provided a substantial range at 30 and 37 °C. A reasonable range in promoter expression was also maintained at 45 °C, but the effect was limited to differences between P_{KmPGK} and P_{KmTDH3}, as P_{KmPGK} and P_{KmTEF3} produced similar expression levels.

Given a validated promoter set with variable expression, we set out to create a series of *K. marxianus* strains with a refactored Shikimate pathway. The library included 27 unique combinations of variable expressions of $KmARO4^{K221L}$, $KmARO7^{G141S}$, and *KmPHA2*. As a first step, integration plasmids targeting the *ABZ1* locus with a three-gene insert were constructed. Expression cassettes and inserts are depicted in Figure 2.5A. All plasmids and resulting integrated strains were verified by multiple Sanger sequencing experiments. Of note, the refactored pathway was integrated into the genome leaving intact the native Shikimate pathway. This was done so as to not disrupt native Shikimate pathway functions and because the wildtype pathway already produces upward of ~150 mg/L of 2-PE, thus providing a strong starting point for pathway engineering.

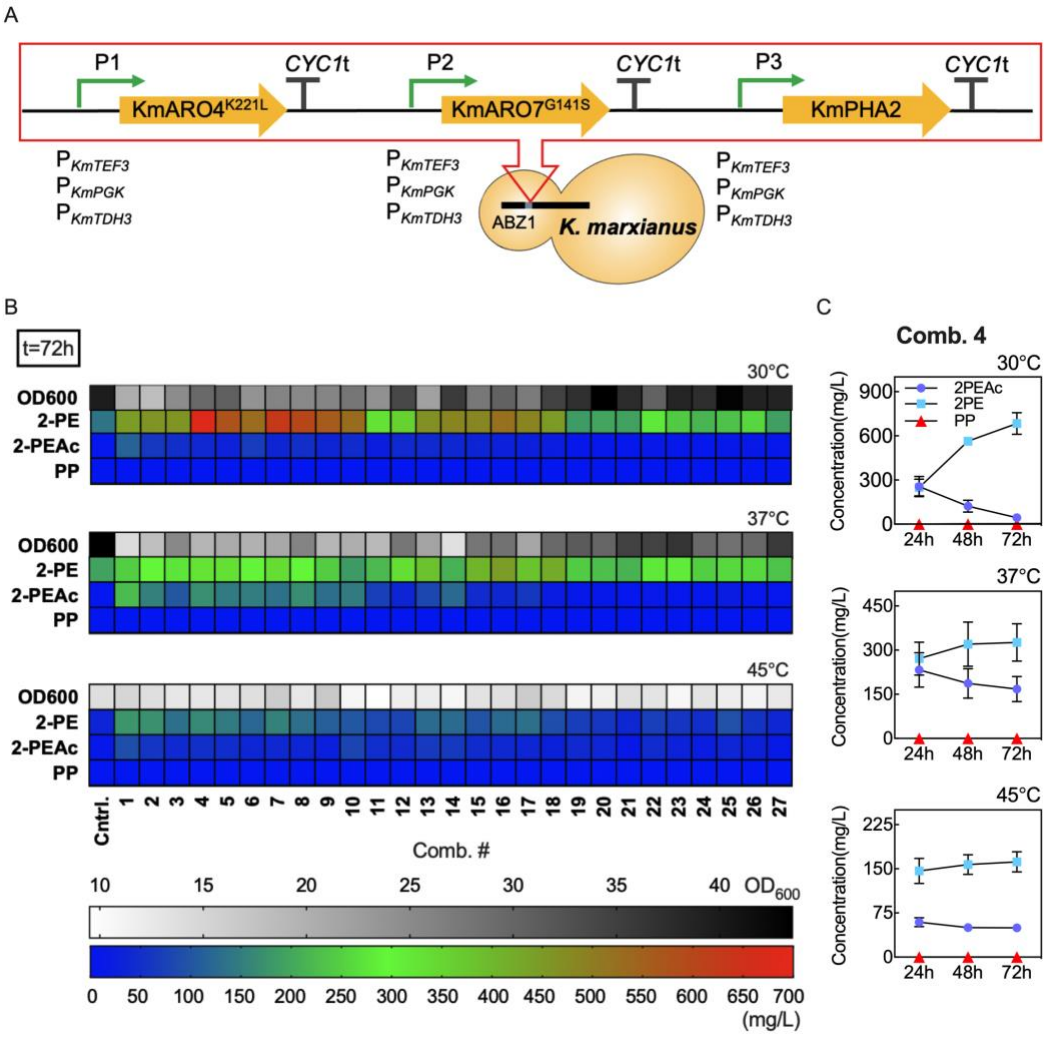

Figure 2.5. Shikimate pathway refactoring for enhanced 2-PE biosynthesis. (A) Schematic representation of pathway constructs for the library consisting of all 27 combinations of P_{KmTEF3}, P_{KmPGK}, and P_{KmTDH3} promoters for expression of *KmARO4*^{K221L}, *KmARO7*^{G141S}, and *KmPHA2*. (B) OD_{600}, extracellular accumulation of 2-phenylethanol (2-PE), 2-phenylethyl acetate (2-PEAc) and phenylpyruvate (PP) in all refactored pathway combinations after 72 h of cultivation at 30, 37 and 45 °C in 25 mL YPD medium. The negative control strain (Cntrl.) is *K. marxianus* CBS 6556 *ura3Δ his3Δ*. (C) Extracellular formation of 2-PE, 2-PEAc and PP over time and at different temperatures for *K. marxianus* CBS 6556 *ura3Δ his3Δ* with refactored Shikimate pathway combination 4 (Comb.4). All experiments were performed in biological triplicates. Bars represent the mean, while error bars represent the standard deviation.

Figure 2.5B presents the OD_{600} along with the titer of 2-PE, 2-PEAc, and PP produced by each refactored pathway after 72 h of culture at 30, 37, and 45 °C. This time

point was used for pathway screening because all strains, including those with a refactored pathway as well as the base strain CBS6556 *ura3Δ his3Δ*, showed the highest extracellular concentration of 2-PE after 72 h cultivation with 20 g/L D-glucose (Figure S2.5). The same trend was not observed for 2-PEAc formation, which decreased by 81 ± 11% on average from 24 to 72 h at 30 °C. The concentration of PP was found to be nearly undetectable and unchanging throughout the course of experiments at all temperatures.

Multilinear regression of 2-PE concentration resulting from the refactored pathway combinations at 30 °C revealed statistically significant effects from the overexpression of each gene. The magnitude of the $KmARO4^{K221L}$ effect was the largest, 4.0- and 3.7-fold greater than that of $KmARO7^{G141S}$ and *KmPHA2* overexpression. Interaction effects between genes were not found to be significant. Post-hoc analysis by multiple t tests showed that pathway combination 4 (Comb.4) with P_{KmTEF3} driving expression of $KmARO4^{K221L}$ and *KmPHA2*, and P_{KmPGK} for overexpression of $KmARO7^{G141S}$ resulted in the highest 2-PE titer, 684 ± 73 mg/L (Figure 2.5B, direct comparison shown in Figure S2.6). The large effect of $KmARO4^{K221L}$ overexpression was observable in the combinations which have high or medium expression of this gene. At 30 °C, Comb.1-9 and 10-18 produced on average 563 ± 121 and 493 ± 88 mg/L of 2-PE, respectively, while Comb.19-27 only produced 257 ± 45 mg/L of 2-PE (Figure S2.7).

Analysis of the data collected at 30 °C also revealed another trend, that of decreased biomass accumulation with increased 2-PE production. Correlation analysis between specific 2-PE titer (mg L^{-1} OD^{-1}) and OD_{600} after 72 h of cultivation resulted in a

Pearson coefficient, r, of -0.89 (Figure S2.8). A similar correlation with a Pearson coefficient, r, of -0.74 was also found between total 2-PE titer (mg/L) and OD_{600}.

Given that *K. marxianus* CBS 6556 *ura3Δ his3Δ* exhibited comparable growth rate at 30 and 45 °C (Figure S2.3), we used the refactored pathway library to investigate any potential temperature effects on both biomass accumulation and 2-PE biosynthesis. Figures 2.5B and S2.7 show that OD_{600} of the negative control strain at 45 °C was only half of that observed at 30 and 37 °C. 2-PE production, both in terms of titer (mg/L) and specific productivity (mg L^{-1} OD^{-1}), was highest at 30 °C and decreased as temperature increased to 37 and 45 °C (Figures 2.5B, S2.5, and S2.9). For example, 2-PE production from Comb.4 at 72 h decreased from a high of 23 ± 2 mg L^{-1} OD^{-1} at 30 °C to 17 ± 3 mg L^{-1} OD^{-1} at 37 °C and 9 ± 1 mg L^{-1} OD^{-1} at 45 °C (Figure S2.9). The time course production of the three relevant metabolites (2-PEAc, 2-PE, and PP) at 30, 37, and 45 °C for the highest producing strain also demonstrated that 30 °C was the optimal temperature to minimize the acetylation of 2-PE (Figure 2.5C). As such, we only considered 30 °C cultures for additional analysis and subsequent mutations of the engineered strain with Comb.4 integrated into the genome to enhance 2-PE biosynthesis.

To further increase 2-PE biosynthesis, we explored additional mutations to increase Ehrlich pathway flux and reduce ester production (Figure 2.6A). We first disrupted *KmARO8*, knockout of which is known in *S. cerevisiae* to upregulate *ARO10* expression and increase Ehrlich pathway flux [32]. In the engineered strain with integrated Comb.4, loss of *KmARO8* function had little to no effect on 2-PE and 2-PEAc biosynthesis after 72 h of cultivation. Conversely, P_{KmTEF3}-driven overexpression of

KmARO10 integrated at the *URA3* locus increased 2-PE and 2-PEAc production by 10% and 146% over Comb.4, producing 746 ± 31 and 110 ± 29 mg/L, respectively. Disruption of the ethanol acetyltransferase *KmEAT1* increased the ratio of 2-PE to 2-PEAc by preventing conversion of alcohol to ester but did not completely eliminate 2-PEAc biosynthesis. The combined effect of *KmARO10* overexpression and *KmEAT1* disruption produced the highest amount of 2-PE, 766 ± 6 mg/L, from 20 g/L D-glucose, with only 10 ± 1 mg/L 2-PEAc at 30 °C after 72 h of cultivation.

During time-course analysis of the pathway refactoring experiments, we observed that D-glucose was depleted by 24 h in batch culture at 30 °C (Figure S2.10). As such, we tested the highest producing strain (*K. marxianus* CBS 6556 *his3Δ eat1Δ abz1*::(P_{KmTEF3})*KmARO4*^{K221L}- (P_{KmPGK})*KmARO7*^{G141S}-(P_{KmTEF3})*KmPHA2*, *ura3Δ*::(P_{KmTEF3})*KmARO10*) under fed-batch conditions with repeated addition of D-glucose every 24 h, replenishing the culture to 20 g/L (Figure 2.6B). 2-PE titer kept increasing until 120 h, reaching a plateau at 1943 ± 63 mg/L. Until 120 h of cultivation, PP remained undetectable and 2-PEAc remained below 110 mg/L. At longer times, 2-PE biosynthesis stalled, glucose consumption decreased, biomass accumulation plateaued, and PP titer increased, suggesting a potential end point to the process.

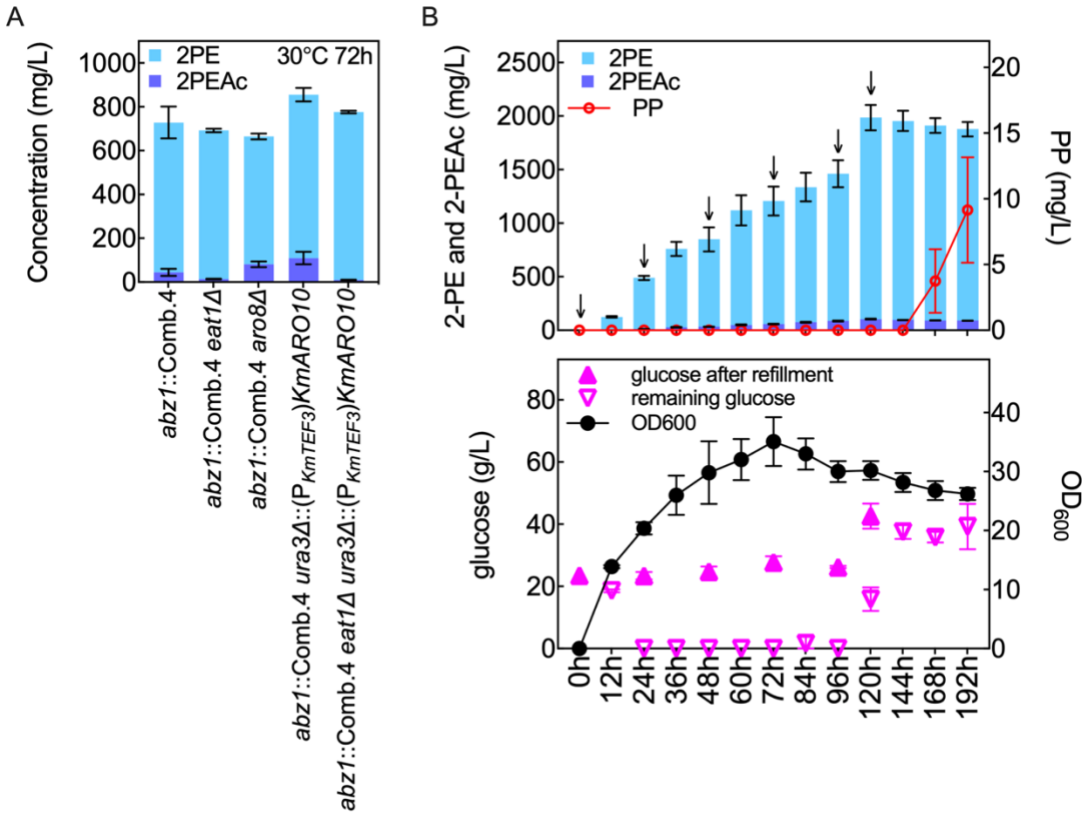

Figure 2.6. Ehrlich pathway engineering and cultivation optimization. (A) Screening of Ehrlich pathway mutations for enhanced 2-PE biosynthesis. (B) 2-phenylethanol (2-PE), 2-phenylethyl acetate (2-PEAc), phenylpyruvate (PP) and biomass (OD_{600}) accumulation from D-glucose in fed-batch operation of the optimal strain, *K. marxianus* CBS 6556 *eat1Δ his3Δ ura3Δ*::(P_{KmTEF3})*KmARO10*, *abz1*::(P_{KmTEF3})*KmARO4*^{K221L}-(P_{KmPGK})*KmARO7*^{G141S}-(P_{KmTEF3})*KmPHA2*. Arrows indicated glucose addition. All experiments were performed in biological triplicates. Bars represent the mean, while error bars represent the standard deviation.

2.4 Discussion

A central challenge in creating new high-producing strains for the industrial bioproduction of chemicals is the rapid creation of new strains with optimized metabolism [15, 33]. One critical component of this is moving beyond plasmid-based expression and creating new strains with integrated optimized pathways in the genome, thus facilitating translation from laboratory screens to industrial use. In this work, we set

out to create a multigene integration tool that facilitates pathway refactoring in the stress-tolerant yeast *K. marxianus*. Our work expands the *K. marxianus* metabolic engineering toolbox, which includes sequential untargeted gene integration using the *URA3* blaster selection cassette [34], random simultaneous integration of multiple genes [35], genome-editing by CRISPR-Cas9 [23, 24, 25], and gene regulation by CRISPRi [7]. Missing from these techniques is the ability to integrate multiple genes into desired sites on the genome at high efficiency. With a CRISPR-Cas9 induced double-stranded break at a targeting locus, we demonstrated that up to three unique expression cassettes concatenated into a single insert can be integrated with an efficiency that is comparable to single gene integration. Efficiency was observed to be as high as 95% for a single gene integration into the *URA3* locus. Three-gene integration into the *ABZ1* locus using a single genetic insert which is 4994 bp in length occurred at an efficiency of 58.3%, statistically similar to the efficiency of integrating a single-gene insert (1,730 bp) and a two-gene insert (3,343bp) at the same site (Figure 2.2B). Moreover, the CRISPR-mediated system developed here shortens the experimental time required to create multigene insertions. A triple-gene integration into a targeted genomic site can be accomplished within 4 days from transformation to confirmation (Figure S2.11). Given a promoter set with a wide transcriptional range, we used this new genomic integration tool to create the full factorial library of three Shikimate pathway genes *ARO4*, *ARO7*, and *PHA2* at three different expression levels (Figures 2.4 and 2.5).

2-PE is a valuable compound for a variety of different applications including next-generation biofuels and flavor and fragrance additives for food, beverages,

perfumes, and personal care products. As a potential alternative to chemical synthesis and isolation from natural sources, bioproduction from engineered yeast and bacteria has been investigated by a number of researchers [36-39]. One of these works was able to increase 2-PE production in *K. marxianus* to upward of 1.1 g/L by growth selection in the presence of a toxic L-phenylalanine paralog, p-fluorophenylalanine [37]. The concept of the screen was to identify strains that overproduce phenylalanine, a major product of the Shikimate pathway and a precursor to 2-PE via the Ehrlich pathway. While mutations to the high producing strain were not identified in the growth screen, the study suggested that rational engineering to increase flux along the Shikimate pathway should prove valuable. Our Shikimate pathway refactoring results support this conclusion: increased flux to 2-PE was achieved by increasing and balancing the expression of mutant variants of *Km*ARO4, *Km*ARO7, and wild type *Km*PHA2 (Figure 2.5).

Previous work in *S. cerevisiae* suggested that *ARO4* and *ARO7* are negatively regulated by tyrosine, thus reducing flux through the Shikimate pathway [9, 40-42]. A K229L mutation alleviated this feed-back inhibition in *ScARO4* as evidenced by an ~80-fold increase in 2-PE, reaching a titer of 32 mg/L with glucose as a feedstock [9]. A G141S or T226I mutation to *ScARO7* also relieved the suppression induced by tyrosine [41, 42]. Co-overexpressing *ScARO7*^{G141S} and *ScARO4*^{K221L} increased flux along the Shikimate pathway and boosted 2-PE formation by 200-fold, from 0.4 to ~80 mg/L [9]. In the *K. marxianus* homologs, the K229L mutation in *ScARO4* maps to *KmARO4*^{K221L}, while the *ScARO7* mutation occurs at the same position in the *K. marxianus* variant, G141S. Plasmid-based overexpression of the putatively feedback-insensitive

KmARO4^{K221L} in *K. maxianus* CBS 6556 *ura3Δ his3Δ* increased the 2-PEAc titer from 31 ± 7 to 202 ± 16 mg/L, a nearly 7-fold increase (Figure 2.3). The overexpression of *KmARO7*^{G141S} in *K. marxianus* CBS 6556 *ura3Δ his3Δ abz1*::(P_{ScTDH3})*KmARO4*^{K221L}-(P_{ScTDH3})*KmARO10* did not further enhance the 2-PEAc biosynthesis (Figure 2.3, *ARO7*). Overexpressing the *S. cerevisiae* variants in *K. marxianu*s produced similar results, that is, alleviating *ARO4* inhibition was beneficial but overexpression of the *ARO7* mutant did not increase pathway flux (Figure 2.3). Taken together, these results suggest key differences between *K. marxianus* and *S. cerevisiae* in Shikimate pathway metabolism: (1) *KmARO4* still suffers from but appears to be less sensitive to feedback inhibition, and (2) feedback inhibition suppressing *KmARO7* activity was not observed.

By refactoring the Shikimate pathway, we were able to identify a condition that enhanced 2-PE biosynthesis with high expression of *KmARO4*^{K221L} and *KmPHA2*, and medium expression of *KmARO7*^{G141S}. Additional gains in 2-PE biosynthesis were not achieved with higher *ARO7* expression potentially due to the interplay between tyrosine feedback inhibition, which was alleviated in *KmARO4* ^{K221L} and *KmARO7*^{G141S}, and the concentration of tyrosine precursors, specifically the ARO7 product prephenate. The optimized pathway produced 684 ± 73 mg/L of 2-PE after 72 h of shaking at 30 °C in YPD medium, achieving 62% of that in the previously reported toxic phenylalanine paralog screen [37]. This result suggests that while increasing Shikimate pathway flux is important, other mutations might also be beneficial including Ehrlich pathway engineering. Overexpression of the decarboxylation step encoded by *KmARO10* further increased 2-PE titer to 746 ± 31 mg/L, but similar success was not achieved in

overexpressing genes encoding putative alcohol dehydrogenases. Specifically, screening various native and heterologous *ADH* variants did not increase 2-PE production (Figure S2.12), suggesting that *K. marxianus* CBS 6556 *ura3Δ his3Δ* maintains a high native capacity to reduce the intermediate, phenylacetaldehyde. The terminal Ehrlich pathway step acetylates 2-PE into 2-PEAc primarily via Eat1 activity [7, 43]. With the disruption of *KmEAT1* the extracellular formation of 2-PEAc decreased by 91%, from 110 ± 29 to 10 ± 0.5 mg/L. We previously identified *KmATF1* in *K. marxianus* as an alcohol acetyltransferase that is partially responsible for ethyl acetate production [23]. Promiscuous activity of *KmATF1* towards 2-PE may account for 2-PEAc biosynthesis in the absence of *KmEAT1*. The final strain with optimized Shikimate pathway, *KmARO10* overexpression, and *KmEAT1* disruption produced a high 766 ± 6 mg/L 2-PE titer from 20 g/L of D-glucose in 25 mL YPD medium with shake flask experiments.

K. marxianus' fast growth and thermotolerance to temperature upward of 45-50 °C are valuable traits beneficial to bioprocess design [33]. As such, we investigated the effect of increased culture temperature on 2-PE biosynthesis using the refactored pathway library. Data presented in Figure 2.5 demonstrated that at elevated temperature 2-PE production decreased and the effect of pathway refactoring was minimized. The best Shikimate pathway combination (Comb.1: (P_{KmTEF3})*KmARO4*^{K221L}-(P_{KmTEF3})*KmARO7*^{G141S}-($P_{Km\ TEF3}$)*KmPHA2*) at 45 °C was limited to 177 ± 20 mg/L of 2-PE, 74% lower than the optimal combination (Comb.4: (P_{KmTEF3})*KmARO4*^{K221L}-(P_{KmPGK})*KmARO7*^{G141S}-($P_{Km\ TEF3}$)*KmPHA2*) at 30 °C (Figures 5 and S5). Production was higher at 37 °C: the best combination (Comb. 16: (P_{KmPGK})*KmARO4*^{K221L}-

(P_{KmTDH3})*KmARO7*^{G141S}-(P_{KmTEF3})*KmPHA2*) produced 444 ± 196 mg/L, a decrease of 35% compared to the Comb.4 at 30 °C (Figures 2.5 and S2.5).

While we observed a negative correlation between temperature and 2-PE titer in this work, others have reported increased production in metabolically engineered *K. marxianus* with increasing temperature. For example, an engineered pathway for triacetic acid lactone (TAL) in *K. marxianus* CBS 712 produced the highest titer, 1.24 g/L, at 37 °C, ~2-fold more than that at 30 °C [44]. Two key differences are notable between these experiments: (1) our data showed that the expression level of the promoters used to refactor the Shikimate pathway decreased with increasing temperature (Figure 2.4), thus resulting in less impact on flux regulation along the Shikimate pathway at elevated temperatures; and, (2) the TAL experiments were conducted with xylose instead of glucose as the carbon source. Previous transcriptional analysis of temperature effects in *K. marxianus* suggested that glycolysis is down-regulated while the pentose phosphate pathway is up-regulated at 45 °C [29, 45]. As for *K. marxianus* CBS 712, the enzymatic activity of xylose reductase and xylitol dehydrogenase was not hampered by high temperature upward to 41 °C [46]. The decrease in glycolytic activity in our experiments was a likely factor in the loss of 2-PE production and is supported by the significant decrease in biomass accumulation observed at 45 °C (Figures 2.5B and S2.7).

With respect to process design, we further increased 2-PE production through fed-batch operation adding glucose every 24 h over a 120-h experiment (Figure 2.6B). Similar to other reports, in our long-term cultures we observed that 2-PE was toxic to *K. marxianus* [13, 47]. This toxicity is observed after 120 h of fed-batch cultivation, when

PP began to accumulate and the 2-PE titer plateaued. The effect can be reduced and/or eliminated through *in situ* removal in a two-phase bioreactor [13]. Such experiments are the focus of on-going process development.

In this work, we focused on refactoring the Shikimate pathway, but other studies have suggested additional metabolic engineering targets that have potential to make gains in 2-PE production. Two recent studies using *S. cerevisiae* and *Y. lipolytica* as the production host both demonstrated that suppressing the consumption of PEP, one of the substrates of Aro4, to form pyruvate by reducing or even blocking the pyruvate kinase, led to enhancement of 2-PE biosynthesis [36, 38]. In *Y. lipolytica*, overexpression of the homologous transketolase *ylTKT* and heterologous phosphoketolases *BbxfpK* and *AcxpkA*, which aimed to increase E4P formation from the pentose phosphate pathway, also increased the 2-PE production [38]. However, in *S. cerevisiae*, enhancing the phosphoketolase pathway did not result in higher 2-PE titer. Instead, elimination of the parasitic pathway forming p-hydroxyphenylethanol significantly increased the 2-PE production [36].

2.5 Materials and Methods

2.5.1 Molecular Cloning and Reagents

All molecular cloning reagents and enzymes were purchased from NEB. Plasmids created and used in this work are shown in Table S2.2. Plasmids were constructed using the following molecular biology reagents: Q5® High-Fidelity DNA polymerase was used for DNA amplification, NEBuilder® HiFi DNA Assembly Master Mix was used for

plasmid assembly, and selected endonucleases were used for plasmid digestion as described in subsequent sections. Primers for gene amplification and mutation were identified in the *K. marxianus* CBS 6556 strain by translated BLAST (tblastn) of annotated genes in the *K. marxianus* DMKU3-1042 or *Saccharomyces cerevisiae* S288C. All primers were purchased from IDT™. All molecular cloning was accomplished in TOP 10 competent *E. coli* (Thermo Fisher Scientific).

All other reagents used in this work were purchased from Fisher Scientific, Sigma-Aldrich or as otherwise noted: BD Difco™ Yeast Nitrogen Base without Amino Acids (Fisher Scientific), CSM-His powder (Sunrise Science Products), CSM-His-Ura powder (Sunrise Science Products), Yeast Synthetic Drop-out Medium Supplements without uracil (Sigma-Aldrich), D-glucose (Fisher Scientific), Gibco™ Bacto™ Yeast Extract (Fisher Scientific), Gibco™ Bacto™ Peptone (Fisher Scientific), 5-FOA (Fisher Scientific), Miller's LB powder (Sigma-Aldrich), Ampicillin sodium salt (Fisher Scientific), Pfaltz & Bauer Polyethylene glycol 3350 (Fisher Scientific), Lithium acetate dihydrate (Sigma-Aldrich), EDTA disodium salt dihydrate (Sigma-Aldrich), Dithiothreitol (Fisher Scientific), R&D Systems™ Salmon Sperm DNA (Fisher Scientific), 2-Phenylethanol (99.0% min by GC, Millipore Sigma), 2-phenylethyl acetate (99.0% min by GC, Millipore Sigma), Phenylpyruvic acid (97.0% min by HPLC, Millipore Sigma), and Cyclohexane (99.0% min by GC, Fisher Scientific).

2.5.2 *K. marxianus* Strains, Media, and Cultivation

K. marxianus CBS 6556 *ura3Δ his3Δ* was used as a starting strain for all experiments described in this work. All constructed strains are listed in Table S2.2.

Synthetic defined (SD) media was used for all plasmid-based expression experiments. The SD-U medium is defined as 6.7 g/L BD Difco™ Yeast Nitrogen Base without Amino Acids, 1.92 g/L Yeast Synthetic Drop-out Medium Supplements without uracil, and 20 g/L D-glucose, while SD-H and SD-H-U are similar defined but with 0.75 g/L of CSM-His and CSM-His-Ura, respectively. For all pathway refactoring experiments and 2-PE biosynthesis analysis, *K. marxianus* strains were cultivated rich YPD medium (YPD: 10 g/L Gibco™ Bacto™ Yeast Extract, 20 g/L Gibco™ Bacto™ Peptone, 20 g/L D-glucose). 20 g/L agar was added to make solid agar plates. All yeast cultures were conducted in 250 mL baffled shake flasks containing 25 mL of appropriate media. Culturing was conducted in an INFORS HT Multitron incubation shaker with temperature control set to 30, 37 and 45 °C as needed.

2.5.3 CRISPR-Cas9 Gene Disruption and Integration Plasmids

CRISPR plasmids (pCRISPR) were constructed using pIW601 (Addgene ID 98907) linearized at PspXI and re-assembled with a 60 bp insert containing 20 bp upstream and downstream homology as well as the 20 bp target sequence by Gibson assembly. Each insert was constructed by annealing the two complementary 60 bp primers. For each locus, five crRNA sequence candidates were chosen based on the on-target cutting efficiency using the sgRNA design tool hosted by Broad Institute (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) and the uniqueness across the *K. marxianus* genome using a CRISPR/Cas9 target online predictor, CCTop (https://crispr.cos.uni-heidelberg.de/).

Homology donor plasmids (pHD) for gene integration were constructed from pIW578. pIW578 was digested with two restriction enzymes, NotI and EagI, to sequentially insert 700 bp up- and down-stream homology of the 20 bp crRNA targeting locus on the *K. marxianus* CBS 6556 *ura3Δ his3Δ* genome. The homology donor vectors were then digested at XmaI and XhoI to insert gene expression cassette(s) for targeted integration. For example, pHDs with single- (*EGFP*), dual- (*DSRED* and *EGFP*), and triple-gene (*DSRED*, *YFP*, and *EGFP*) cassettes were created as listed in Table S2.2.

2.5.4 Enzyme Screening and Shikimate Pathway Refactoring Plasmids

All plasmids used for enzyme screening were constructed by replacing *EGFP* in the pIW1135 with the new gene (Table S2.2). As for the expression plasmids for single-amino acid mutated variants of *ScARO4*, *ScARO7*, *KmARO4*, and *KmARO7*, a 53-bp insert containing the mutation flanked by 25 bp up- and down-stream homology of the to-be-replaced triplet of bases was designed for each of them. The vector expressing the corresponding native protein was linearized by a pair of primers designed to eliminate the exact codon to be substituted.

Refactored Shikimate pathways were integrated into *K. marxianus* CBS6556 *ura3Δ his3Δ* in a three-step process. First, variable strength promoters were validated, second, plasmids with multiple gene expression cassettes were constructed, and finally, the 27 multigene expression combinations were integrated into the *ABZ1* locus in parallel. Briefly, promoter validation plasmids were constructed by replacing the P_{ScTDH3} used to drive *EGFP* expression in pIW1135 with P_{KmTEF3}, P_{KmPGK}, and P_{KmTDH3} [26] by Gibson assembly. Subsequently, nine expression cassettes using each of the promoters

separately for *KmARO4*^{K221L}, *KmARO7*^{G141S} and *KmPHA2* were constructed by replacing the *EGFP* gene. These expression cassettes were then concatenated into multigene inserts by designing primers with overlapping ends for Gibson assembly. The result was 27 unique integration plasmids with 700 bp up and down-stream homology to *ABZ1* flanking a three-gene overexpression insert. Integration was accomplished as described below.

2.5.5 *K. marxianus* Transformation

K. marxianus transformations were accomplished using a modified protocol described in [23]. Briefly, the desired starting *K. marxianus* strain was cultivated in 2 mL YPD for 16 h at 30 °C. 1 mL of the cell culture was harvested by centrifugation at 5,000 g for 1 min in a sterile 1.5 mL Eppendorf tube. After washing twice with 1 mL of sterile ddH_2O, cells were mixed with 10 µL of 10 mg/mL carrier DNA (R&D Systems™ salmon sperm DNA) by gently vortexing. For gene integration, 1 µg pCRISPR and 0.8 µg pHD were then added. For CRISPR-mediated gene disruption or plasmid-based gene expression, 1 µg pCRISPR or 0.8 µg overexpression plasmid was used. After adding 500 µL of transformation buffer, all components in the tube were mixed thoroughly by pipetting and were incubated at room temperature for 15 min. Subsequently, the mixture was heat shocked at 47 °C for 5 min using a solid heat block containing diH_2O in each well. Cells were then harvested by centrifugation at 5,000 g for 30 s. Supernatant was removed and cell pellets were resuspended in 500 µL selective media based on the selection marker of the transformed plasmid(s). The transformation buffer contains 40% w/w autoclave sterilized polyethylene glycol 3350 (PEG), 0.45-µm filtered 0.1 M lithium acetate (LiAc), 10 mM Tris–HCl (pH 7.5) with 1 mM EDTA (1× TE buffer), and 10 mM

DTT. Transformants were selected by inoculated in the appropriate selective liquid medium or by plating directly on selective solid agar medium as described in next section.

2.5.6 Strain Selection and Validation

For gene integration, 50 µL of transformed cell resuspension was first cultivated in 2 mL SD-H-U medium at 37 °C for 36 h. To enrich colonies with expected chromosomal insertion, cells were reinoculated in fresh medium overnight at 30 °C, and subsequently plated on YPD for single colonies (see Figure S2.11). The efficiency of genomic integration was determined after 16 h incubation on YPD plates at 37 °C. In each case, 28 colonies were randomly picked and screened by a three-primer colony PCR and subsequent gel electrophoresis. Two of the designed primers bind to the genome outside the homology arms (one upstream and one downstream). The third primer binds to the integration cassette (see Figure 2.2A). Positive integrations were confirmed with amplicons produced from the upstream primer that binds to the insert and downstream genomic primer. Positive integration amplicons vary in size considering primer uniqueness; therefore, the binding location of the upstream genomic primer was adjusted on purpose in response to maintain band size difference for the positive discovery. For single-gene integration, in the absence of integration, amplification occurred with the two genomic primers produced a 1,640 bp amplicon while successful gene integrations gave an 831 bp amplicon. Positive multigene integration produced a ~1,110 bp amplicon and the negative exhibited a 2,023 bp band. Plasmids were removed from selected colonies that gave positive results in the colony PCR analysis by culturing in 1 mL YPD

containing 1 g/L 5-Fluoroorotic acid (5-FOA) at 30 °C for 16 h. Single colonies were isolated by plating again on solid YPD. The inserted gene expression cassette(s) were confirmed by Sanger sequencing after genome isolation (YeaStar™ Genomic DNA Kit), fragment amplification and purification (DNA Clean & Concentrator®-5). The colony selection for plasmid-based gene overexpression was achieved by outgrowth of the transformed cells in the appropriate selective liquid medium for two times to enrich the number of cells harboring the expected plasmid, and then plated on selective agar plates for single colonies ready for use.

To confirm gene disruptions, 50 µL of transformed cell resuspension was inoculated in 2 mL SD-U. After reaching confluency at 37 °C (usually after 24 h cultivation), cells were reinoculated in another 1 mL SD-U for outgrowth at 30 °C overnight and appropriately diluted cell cultures were plated on SD-U to obtain single colonies. After 16 h of growth at 37 °C, 14 randomly picked colonies were screened by amplifying a ~500 bp fragment containing the CRISPR-Cas9 targeted site by colony PCR. PCR amplicons were isolated by gel electrophoresis and were sent for Sanger sequencing after column-purification (Zymoclean™ Gel DNA Recovery Kit).

2.5.7 Microtiter Plate Measurements

A BioTek®Synergy™ Neo2 multi-mode microplate reader was used to measure the relative fluorescence intensity of yeast cells. Optimization studies revealed that maximum fluorescence was observed when cells reached late exponential phase, approximately 14 h of growth at 30 °C in YPD medium. Due to differences in quantum yield and absolute fluorescence produced from *EGFP*, *YFP*, and *DSRED*, fluorescence

measurements varied between samples. 500 µL of cell culture was collected, pelletized, and washed twice with PBS. For *EGFP*, the washed cell pellets were resuspended in 500 µL of PBS and diluted 10-fold prior to measurement (Ex/Em 488/511). For *YFP*, cell pellets were resuspended in 500 µL of PBS and measured directly (Ex/Em 520/540). Finally, for *DSRED*, 100 µL PBS was used to resuspend harvested cells before fluorescence quantification (Ex/Em 555/604). All measurements were conducted on 100 µL of resuspended cells in a black Corning™ 96-well flat-bottom microplate (Fischer Scientific).

2.5.8 Flow Cytometry

Single cell fluorescence quantification was conducted using a BD Accuri™ C6 Plus flow cytometer. Briefly, 500 µL of cell culture was transferred into a 1.5 mL tube and centrifuged at 5,000 g for 1 min. Supernatants were discarded and cell pellets were washed twice with 1 mL PBS. Cell pellets were then resuspended in 500 µL sterile ddH_2O. Two-microliter of the cell solutions was further diluted in 100 µL ddH_2O and placed into one well of a clear Corning™ 96-well flat-bottom polystyrene microtiter plate for auto-loading. Standard manufacturer settings of laser, filter, and detector to quantify *EGFP* fluorescence were used (Ex/Em 488/533). Each run was limited by collecting 10,000 events, and the fluidics rate was 14 µL/min. Histograms of flow cytometry events for data presented in Figure 2.4 are shown in Figures S2.14-S2.17. For strains with plasmid-based *EGFP* overexpression cultured in selective medium, samples were taken after 16 h of cultivation at 30 °C. For chromosomal expression in the rich YPD medium,

the sampling time for 30 and 37 °C was 14 h, while for 45 °C, samples were analyzed at 10 h.

2.5.9 Gas Chromatography

Analysis of 2-phenylethanol (2-PE), 2-phenylethyl acetate (2-PEAc), and phenylpyruvate (PP) was carried out on a Shimadzu GC-2010 Plus equipped with a Shimadzu AOC-20s autosampler and a Shimadzu AOC-20i auto-injector. The GC suite was coupled to a flame ionization detector (FID). Compounds were separated on an Agilent J&W DB-WAX Ultra Inert column (length: 30 m; inner diameter: 0.32 mm; film thickness: 0.5 µm), using a 21-min temperature program as follows: start temperature of 100 °C, 20 °C/min to 140 °C, 10 °C/min to 150 °C, 5 °C/min to 160 °C, hold for 2 min then increase by 1 °C/min to 170 °C, hold for 2 min, and finally 25 °C/min to 220 °C. Helium was used as the carrier gas at a flow rate of 1.9 mL/min. The sample injection volume was 1 µL. Split mode was used for injection and the ratio was 20:1. The following retention times were determined using standard samples: PP, 4.5 min; 2-PEAc, 8.1 min; 2-PE, 10.1 min.

For sample preparation, 700 µL cell culture was centrifuged for 1 min at 5,000 g, 500 µL supernatant was collected and transferred to a clean 1.5 mL tube with an equal volume of cyclohexane. The mixture was vortexed thoroughly for 30 min. After centrifugation at 10,000 g for 1 min, 100 µL of the organic layer was transferred into a 2 mL clear Agilent GC vial with glass insert. Standard curves depicting the linear correlation between the concentration of three compounds (PP, 2-PEAc, 2-PE) with the area of peaks from FID were obtained respectively to quantify extracellular metabolite

accumulation by different strains. A series of YPD solutions of 2-PE, 2-PEAc and PP with known concentrations were made and extracted by cyclohexane following the same procedure which was used to extract these three compounds from the supernatants of cell cultures.

2.5.10 Reverse Transcription Quantitative PCR (RT-qPCR)

The relative expression strengths of promoter P_{KmTEF3}, P_{KmPGK} and P_{KmTDH3} were quantified by RT-qPCR according to the *KmARO4*^{K221L} expression in three *K. marxianus* CBS 6556 strains with engineered Shikimate pathway (see the refactored combination 1, 10, and 19). These strains differ only in the promoter that drives *KmARO4*^{K221L} expression. Total RNA was extracted using the YeaStar™ RNA Kit (Zymo Research). 10 µg RNA was DNAse treated (DNAse I, New England Biolabs) in a 100 µL reaction. The mixture was subsequently purified using the RNA Clean & Concentrator™-25 Kit (Zymo Research). 400 ng column-purified RNA was used for a 20 µL reverse transcription reaction (iScript™ Reverse Transcription Supermix for RT-qPCR, Bio-Rad) and 2 µL cDNA synthesis reaction was directly used for SYBR Green qPCR (SsoAdvanced™ Universal SYBR® Green Supermix, Bio-Rad) conducted on the Bio-Rad CFX Connect™. Fold change of the number of *KmARO4*^{K221L} transcripts of each engineered strain (with Comb.1, Comb.10, or Comb.19 integrated at locus *ABZ1*) versus the negative control (*K. marxianus* CBS 6556 *ura3Δ his3Δ*) was calculated using transcription level normalized to *GAPDH* expression (Figure S2.13).

2.5.11 Glucose Assay

The concentration of residual glucose was quantified using a commercially available glucose assay kit (GAGO-20; Sigma-Aldrich). 200 µL of sample was mixed with 400 µL assay reagents as directed. After reaction for 30 min at 37 °C, the reaction was stopped by addition of 400 µL 12 N H_2SO_4. 100 µL of the final solution was then transferred into a clear Corning™ 96-well flat-bottom polystyrene microplate and absorbance was determined at 540 nm using BioTek®Synergy™ Neo2 multi-mode microplate Reader. The region where absorbance is proportional to glucose concentration was identified by the calibration curve. A 0.1 g/L glucose solution stock was made by adding 5 µL YPD containing 20 g/L glucose into 995 µL ddH_2O and mixing thoroughly. A series of dilution were obtained at 0.04, 0.02, 0.01, and 0.005 g/L to plot a standard curve of absorbance at 540 nm and glucose concentration (Figure S2.18). For sample preparation, 500 µL cell cultures were centrifuged down at 5,000 g for 1min and the supernatant was filtered using a Corning® syringe filter with 0.2 µm-pore membrane. 5 µL of the supernatant was diluted in 995 µL ddH_2O. If the absorbance of this dilution at 540 nm is below 0.046, the concentration of glucose remaining in the media will be less than 0.2 g/L, which in this work, was considered as “glucose depleted”.

2.5.12 Statistical Analysis

Brown-Forsythe and Welch one-way ANOVA tests were used to compare GFP fluorescence from various genomic integration sites. Multiple linear regression was used to analyze the pathway refactoring experiments. Comparison between 2-PE titers of any two pathway refactoring combinations was accomplished by multiple t-tests. The Pearson

coefficient, r, was used to represent the correlation between biomass accumulation and 2-PE titer or specific 2-PE production. All data plots and statistical analysis were performed using GraphPad Prism 8 software with a P value < 0.05 taken as significant.

2.6 Supporting Information

Table S2.1. Elected integration loci on the genome of *K. marxianus* CBS6556 ura3Δ his3Δ. The annotated function of the gene, gene ID (*K. marxianus* DMKU3-1042), * protein sequence ID (*S. cerevisiae* S288C) and the targeted sequence of the sgRNA are shown.

Locus	Gene function	Gene ID	sgRNA targeting sequence (5' to 3')
URA3	Disrupted in *K. marxianus* CBS 6556 *ura3Δ his3Δ*	34714385	TTGCCGAGTTATCGTCCAAG
XYL2	Xylose metabolism	* NP_013171.1	ACGATCGCCAACCTTGACCA
ABZ1	para-aminobenzoic acid (PABA) biosynthesis	34717950	AGAATCTTGCTGTTCGAAGC
LYS1	Lysine biosynthesis	34716354	AAAGAGATACAGCTTCTGTT
SDL1	Serine metabolism	34715098	ATTTGATAATGCAACCCTCT
ARO8	Phenylalanine biosynthesis and metabolism	34714734	GATCCACGACATGCAATACG
EAT1	Acetylation	* NP_011529.1	GAGCCTTCAACCAACAAAAA

Table S2.2. Plasmids and strains used in this study

Plasmids	Description	Reference (Addgene #)
pIW601	P_{ScTEF1}-KmCas9-SV40-*ScCYC1*t, P_{KmPRP1}-$tRNA^{Gly}$-PspXI recognition site-SUP4	Löbs et al. [23] (98907)
pIW578	Replacing the *URA3* marker in the pJSKM316-GPD with *HIS3* marker	Lee et al. [31]
pIW538	pIW601with *URA3* targeting sgRNA	This chapter
pIW447	pIW601with *XYL2* targeting sgRNA	This chapter
pIW1134	pIW601with *ABZ1* targeting sgRNA	This chapter
pIW1137	pIW601with *LYS1* targeting sgRNA	This chapter
pIW1132	pIW601with *SDL1* targeting sgRNA	This chapter
pIW21 (pATF1G)	pET-28(+) derivative with *ScATF1-EGFP* insertion	Zhu et al. [48]
pIW1198	700bp *URA3* up and downstream, pIW578 with *EGFP*	This chapter
pIW1136	700bp *XYL2* up and downstream, pIW578 with *EGFP*	This chapter
pIW1135	700bp *ABZ1* up and downstream, pIW578 with *EGFP*	This chapter
pIW1146	700bp *LYS1* up and downstream, pIW578 with *EGFP*	This chapter
pIW1149	700bp *SDL1* up and downstream, pIW578 with *EGFP*	This chapter
pIW247	P_{ScPGK1}-*ATF1-DSRED-ScTEF2*t, P_{ScTDH2}-*ALD6-YFP*-Doc(Cp)-*ScHXT7*t	Lin et al. [49]
pIW1189	Replacing the *EGFP* in pIW1135 with *DSRED*	This chapter
pIW1190	Replacing the *EGFP* in pIW1135 with *YFP*	This chapter
pIW1147	pIW1135 with P_{ScTDH3}-*DSRED-ScCYC1*t	This chapter
pIW1148	pIW1147 with P_{ScTDH3}-*YFP-ScCYC1*t	This chapter
pIW1199	Replacing the *EGFP* in pIW1135 with *ScARO4*	This chapter
pIW1200	Replacing the *EGFP* in pIW1135 with *ScARO7*	This chapter
pIW1201	Replacing the *EGFP* in pIW1135 with *KmARO4*	This chapter
pIW1202	Replacing the *EGFP* in pIW1135 with *KmARO7*	This chapter
pIW1203	Replacing the *EGFP* in pIW1135 with $ScARO4^{K229L}$	This chapter
pIW1204	Replacing the *EGFP* in pIW1135 with $ScARO7^{G141S}$	This chapter
pIW1205	Replacing the *EGFP* in pIW1135 with $KmARO4^{K221L}$	This chapter
pIW1206	Replacing the *EGFP* in pIW1135 with $KmARO7^{G141S}$	This chapter
pIW1222	pIW1205 with P_{ScTDH3}-*KmARO10-ScCYC1*t	This chapter
pIW1246	Replacing the P_{ScTDH3} in pIW1135 with P_{KmTEF3}	This chapter
pIW1247	Replacing the P_{ScTDH3} in pIW1135 with P_{KmPGK}	This chapter
pIW1248	Replacing the P_{ScTDH3} in pIW1135 with P_{KmTDH3}	This chapter
pIW1234	A blank vector made from pIW1135 for individual insertion of the 27 refactored overexpression combinations	This chapter
pIW1235	Replacing the *EGFP* in pIW1246 with $KmARO4^{K221L}$	This chapter
pIW1236	Replacing the *EGFP* in pIW1247 with $KmARO4^{K221L}$	This chapter
pIW1237	Replacing the *EGFP* in pIW1248 with $KmARO4^{K221L}$	This chapter
pIW1229	Replacing the *EGFP* in pIW1246 with $KmARO7^{G141S}$	This chapter
pIW1238	Replacing the *EGFP* in pIW1247 with $KmARO7^{G141S}$	This chapter
pIW1239	Replacing the *EGFP* in pIW1248 with $KmARO7^{G141S}$	This chapter
pIW1230	Replacing the *EGFP* in pIW1246 with *KmPHA2*	This chapter
pIW1240	Replacing the *EGFP* in pIW1247 with *KmPHA2*	This chapter
pIW1241	Replacing the *EGFP* in pIW1248 with *KmPHA2*	This chapter

Plasmids	Description	Reference (Addgene #)
pIW1251	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
pIW1252	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
pIW1253	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
pIW1254	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
pIW1255	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
pIW1256	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
pIW1257	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
pIW1258	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
pIW1259	pIW1234 with P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
pIW1260	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
pIW1261	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2- ScCYC1*t	This chapter
pIW1262	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2- ScCYC1*t	This chapter
pIW1263	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
pIW1264	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
pIW1265	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
pIW1266	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
pIW1267	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
pIW1268	pIW1234 with P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
pIW1269	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}- *KmPHA2-ScCYC1*t	This chapter
pIW1270	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
pIW1271	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
pIW1272	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
pIW1273	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1t*	This chapter
pIW1274	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}- *ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter

Plasmids	Description	Reference (Addgene #)
pIW1275	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}- *ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
pIW1276	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
pIW1277	pIW1234 with P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
pIW869	pIW601 with *EAT1* targeting sgRNA	Löbs et al. [7]
pIW1278	pIW601 with *ARO8* targeting sgRNA	This chapter
pIW1279	Replacing the P_{ScTDH3}-*EGFP* in pIW1198 with P_{KmTEF3}-*KmARO10*	This chapter
pIW1280	Replacing the *EGFP* in pIW1135 with *KmADH1*	This chapter
pIW1281	Replacing the *EGFP* in pIW1135 with *KmADH2*	This chapter
pIW1282	Replacing the *EGFP* in pIW1135 with *KmADH5*	This chapter
pIW1283	Replacing the *EGFP* in pIW1135 with *KmADH6*	This chapter
pIW1284	Replacing the *EGFP* in pIW1135 with *KmADH7*	This chapter
pIW1285	Replacing the *EGFP* in pIW1135 with *KmSFA1*	This chapter
pIW1286	Replacing the *EGFP* in pIW1135 with *ScADH1*	This chapter
pIW1287	Replacing the *EGFP* in pIW1135 with *ScADH2*	This chapter
pIW1288	Replacing the *EGFP* in pIW1135 with *ScPAR1*	This chapter
pIW1289	Replacing the *EGFP* in pIW1135 with *ScPAR2*	This chapter
pIW1290	Replacing the *EGFP* in pIW1135 with *ScPAR3*	This chapter
pIW1291	Replacing the *EGFP* in pIW1135 with *ScPAR4*	This chapter
pIW1292	Replacing the *EGFP* in pIW1135 with *KmPAR1*	This chapter
pIW1293	Replacing the *EGFP* in pIW1135 with *KmPAR2*	This chapter
pIW1294	Replacing the *EGFP* in pIW1135 with *KmPAR3*	This chapter
pIW1295	Replacing the P_{ScTDH3} in pIW1198 with P_{KmNC1}[7]	This chapter
pIW1296	Replacing the *EGFP* in pIW1295 with codon-optimized *rosePAR*	This chapter

Strains	Description	Reference
E. coli TOP 10	F- *mcrA* Δ (*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ (araleu)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Thermo Fisher Scientific
YS4	*S. cerevisiae* BY4742	Shao et al. [50]
YS626	*K. marxianus* CBS 6556 *ura3Δ his3Δ*	Lang et al. [26]
YS1155	YS626 *ura3Δ*::P_{ScTDH3}-*EGFP-ScCYC1*t	This chapter
YS1156	YS626 *xyl2*::P_{ScTDH3}-*EGFP-ScCYC1*t	This chapter
YS1157	YS626 *abz1*::P_{ScTDH3}-*EGFP-ScCYC1*t	This chapter
YS1158	YS626 *lys1*::P_{ScTDH3}-*EGFP-ScCYC1*t	This chapter
YS1159	YS626 *sdl1*::P_{ScTDH3}-*EGFP-ScCYC1*t	This chapter
YS1160	YS626 *abz1*::P_{ScTDH3}-*DSRED-ScCYC1*t	This chapter
YS1161	YS626 *abz1*::P_{ScTDH3}-*YFP-ScCYC1*t	This chapter
YS1162	YS626 *abz1*::P_{ScTDH3}-*DSRED-ScCYC1*t, P_{ScTDH3}-*EGFP-ScCYC1*t	This chapter
YS1163	YS626 *abz1*::P_{ScTDH3}-*DSRED-ScCYC1*t, P_{ScTDH3}-*EGFP-ScCYC1*t, P_{ScTDH3}-*YFP-ScCYC1*t	This chapter

Strains	Description	Reference
YS1164	YS626 *abz1*::P_{ScTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{ScTDH3}-*KmARO10*-*ScCYC1*t	This chapter
YS1165	YS626 *abz1*::P_{KmTEF3}-E*GFP*-*ScCYC1*t	This chapter
YS1166	YS626 *abz1*::P_{KmPGK}-*EGFP*-*ScCYC1*t	This chapter
YS1167	YS626 *abz1*::P_{KmTDH3}-*EGFP*-*ScCYC1*t	This chapter
YS1168	YS626 *abz1*::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2*-*ScCYC1t*	This chapter
YS1169	YS626 *abz1*::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2*-*ScCYC1*t	This chapter
YS1170	YS626 abz1::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1171	YS626 *abz1*::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1172	YS626 *abz1*::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}- *ScCYC1*t, P_{KmPGK}-*KmPHA2*-*ScCYC1*t	This chapter
YS1173	YS626 *abz1*::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1174	YS626 *abz1*::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1175	YS626 *abz1*::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2*-*ScCYC1*t	This chapter
YS1176	YS626 *abz1*::P_{KmTEF3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1177	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1178	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2*- *ScCYC1*t	This chapter
YS1179	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2*- *ScCYC1*t	This chapter
YS1180	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1181	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2*-*ScCYC1*t	This chapter
YS1182	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1183	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1184	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2*-*ScCYC1*t	This chapter
YS1185	YS626 *abz1*::P_{KmPGK}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1186	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}- *KmPHA2*-*ScCYC1*t	This chapter
YS1187	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2*-*ScCYC1*t	This chapter
YS1188	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTEF3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2*-*ScCYC1*t	This chapter
YS1189	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTEF3}-*KmPHA2*-*ScCYC1*t	This chapter

Strains	Description	Reference
YS1190	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
YS1191	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmPGK}-*KmARO7*^{G141S}- *ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
YS1192	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}- *ScCYC1*t, P_{KmTEF3}-*KmPHA2-ScCYC1*t	This chapter
YS1193	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmPGK}-*KmPHA2-ScCYC1*t	This chapter
YS1194	YS626 *abz1*::P_{KmTDH3}-*KmARO4*^{K221L}-*ScCYC1*t, P_{KmTDH3}-*KmARO7*^{G141S}-*ScCYC1*t, P_{KmTDH3}-*KmPHA2-ScCYC1*t	This chapter
YS1195	YS1171 *eat1Δ*	This chapter
YS1196	YS1171 *aro8Δ*	This chapter
YS1197	YS1171 *ura3Δ*::P_{KmTEF3}-*KmARO10-ScCYC1*t	This chapter
YS1198	YS1197 *eat1Δ*	This chapter
YS1199	YS1164 *ura3Δ*::P_{KmNC1}-*rosePAR-ScCYC1*t	This chapter

Table S2.3. Primers used for cloning in this study.

Primers	Sequence (5' to 3', priming parts are underlined)	Use
CasML_125	GAATCCCGTCAGTGTCAACCTTGCCGAGTTATCGTCCAAGGGTTTTAGAGCTAGAAATAG	URA3_f_pCRISPR
CasML_126	CTATTTCTAGCTCTAAAACCCTTGGACGATAACTCGGCAAGGTTGACACTGACGGGATTC	URA3_r_pCRISPR
CasML_045	GAATCCCGTCAGTGTCAACCAGAATCTTGCTGTTCGAAGCGGTTTTAGAGCTAGAAATAG	ABZ1_f_pCRISPR
CasML_046	CTATTTCTAGCTCTAAAACCGCTTCGAACAGCAAGATTCTGGTTGACACTGACGGGATTC	ABZ1_r_pCRISPR
CasML_043	GAATCCCGTCAGTGTCAACCAAAGAGATACAGCTTCTGTTGGTTTTAGAGCTAGAAATAG	LYS1_f_pCRISPR
CasML_044	CTATTTCTAGCTCTAAAACCAACAGAAGCTGTATCTCTTTGGTTGACACTGACGGGATTC	LYS1_r_pCRISPR
CasML_039	GAATCCCGTCAGTGTCAACCATTTGATAATGCAACCCTCTGGTTTTAGAGCTAGAAATAG	SDL1_f_pCRISPR
CasML_040	CTATTTCTAGCTCTAAAACCAGAGGGTTGCATTATCAAATGGTTGACACTGACGGGATTC	SDL1_r_pCRISPR
CasML_141	GAATCCCGTCAGTGTCAACCGATCCACGACATGCAATACGGGTTTTAGAGCTAGAAATAG	ARO8_f_pCRISPR
CasML_142	CTATTTCTAGCTCTAAAACCCGTATTGCATGTCGTGGATCGGTTGACACTGACGGGATTC	ARO8_f_pCRISPR
ML_165	TGGAGCTCCACCGCGGTGGCCTAGATCTATTATGCATTATAATTAATAGTTGTAGC	URA3_pHDup_f
ML_156	TATTGATAATGATAAACTGCGCATTAACAACCCTCTAGGTT	URA3_pHDup_r
ML_105	TGGAGCTCCACCGCGGTGGCGAAGGCCAATTAGTGGAGAACTTGACAT	XYL2_pHDup_f
ML_092	TATTGATAATGATAAACTGCGGGTAACTGCTTCGCCTACTTC	XYL2_pHDup_r
ML_039	TGGAGCTCCACCGCGGTGGCGCCACTCGATGCGATGACTTG	ABZ1_pHDup_f
ML_074	TATTGATAATGATAAACTGCGGGTGGTATTGGACCCCATAC	ABZ1_pHDup_r
ML_077	TGGAGCTCCACCGCGGTGGCAACGGGAATGGTGTGTTATATGACCTTG	LYS1_pHDup_f
ML_070	TATTGATAATGATAAACTGCGGGGAAGTTGTTCCAAGGAAGG	LYS1_pHDup_r
ML_115	TGGAGCTCCACCGCGGTGGCTAGCAACTTTTGCTAGAAGAAAAAACTCAA	SDL1_pHDup_f
ML_100	TATTGATAATGATAAACTGCGGATGAACATATAACTTTTTCGAATATGTAGATTGC	SDL1_pHDup_r
ML_147	GCTCGAAGGCTTTAATTTGCGGGTCTCTAGCGCACGGTGAATA	URA3_pHDdown_f
ML_158	CAAAAGCTGGGTACCGGCCGAATAAAAATTGACTCTTGTATCTTTCTTCACACTAGTAGT	URA3_pHDdown_r
ML_109	GCTCGAAGGCTTTAATTTGCGTCGCTGTAGAGCCAGGTGTTC	XYL2_pHDdown_f

Primers	Sequence (5' to 3', priming parts are underlined)	Use
ML_094	CAAAAGCTGGGTACCGGCCGTCAAAGCTTTGGCATTGACTTTGCCTTC	XYL2_pHDdown_r
ML_091	GCTCGAAGGCTTTAATTTGCTTGGTGGGAAATTAATCCAGAATTTTATTAAATTGGCGAC	ABZ1_pHDdown_f
ML_078	CAAAAGCTGGGTACCGGCCGGTACAGTTTCCCCTCCTTATGGTTAATCAAAATGC	ABZ1_pHDdown_r
ML_081	GCTCGAAGGCTTTAATTTGCTGGGTCAAGGCAAGAAAGCTCTATGAGA	LYS1_pHDdown_f
ML_072	CAAAAGCTGGGTACCGGCCGAGAACCCGAAGTCGAAGAAGAAGAAGAAGAAG	LYS1_pHDdown_r
ML_119	GCTCGAAGGCTTTAATTTGCGGGAGGGCCATTCTGAGATGG	SDL1_pHDdown_f
ML_102	CAAAAGCTGGGTACCGGCCGCACCATTCCCGAGACAAAGAGATTCTCT	SDL1_pHDdown_r
ML_001	TTCTAGAACTAGTGGATCCCCCGGGATGGCTAGCATGACTGGTGGACA	EGFP_f
ML_002	GACATAACTAATTACATGACTCGAGTCATTTGTATAGTTCATCCATGCCATGTGTAATCC	EGFP_r
ML_121	GATTCTAGAACTAGTGGATCCCATGGACAACACCGAGGACGTC	DSRED_f
ML_106	GCGTGACATAACTAATTACATGACTACTGGGAGCCGGAGTGG	DSRED_r
ML_013	GATTCTAGAACTAGTGGATCCCATGGTGAGCAAAGGCGAAGA	YFP_f
ML_014	GCGTGACATAACTAATTACATGACTTATAGAGCTCGTTCATGCCCTC	YFP_r
ML_177	GACGGATTCTAGAACTAGTGGATCCC	DSRED-CYC1t_f
ML_170	GCAAATTAAAGCCTTCGAGCGTCC	DSRED-CYC1t_r
ML_003	ACGCTCGAAGGCTTTAATTTGCGGCCGCAGTTTATCATTATCAATACTCGC	P_{ScTDH3}_f, P_{ScTDH3}-YFP_f
ML_004	TCCACCAGTCATGCTAGCCATC	P_{ScTDH3}-EGFP_r, P_{KmTEF3}-EGFP_r
ML_005	GGCATGGATGAACTATACAAATGACTCGAGTCATGTAATTAGTTATGTCACGCT	CYC1t_f
ML_006	GCAAATTAAAGCCTTCGAGCG	CYC1t_r
ML_016	GAATGTAAGCGTGACATAACTAATTACATGAC	P_{ScTDH3}-YFP_r
ML_181	TCTAGAACTAGTGGATCCCCCGGGATGAGTGAATCTCCAATGTTCGCTGC	ScARO4_f
ML_174	ACATAACTAATTACATGACTCGAGCTATTTCTTGTTAACTTCTCTTCTTTGTCTGACAGC	ScARO4_r
ML_183	TCTAGAACTAGTGGATCCCCCGGGATGGATTTCACAAAACCAGAAACTG	ScARO7_f
ML_176	ACATAACTAATTACATGACTCGAGTTACTCTTCCAACCTTCTTAGCAAG	ScARO7_r
ML_185	TCTAGAACTAGTGGATCCCCCGGGATGTCAGCTACACCACAACCTATG	KmARO4_f
ML_178	ACATAACTAATTACATGACTCGAGTTATTTAGCGGCCTTCTTTTTTAGTTCTCT	KmARO4_r

Primers	Sequence (5' to 3', priming parts are underlined)	Use
ML_187	TCTAGAACTAGTGGATCCCCCGGGATGGATTTTTTTAAACCAGAAACTGTTC	KmARO7_f
ML_180	ACATAACTAATTACATGACTCGAGTCATTTCTCTTCATCCTCCAAC	KmARO7_r
ML_239	TTCTAGAACTAGTGGATCCCCCGGGATGGTTAAAGTGCTGTATCTAGGGCC	KmPHA2_f
ML_232	GACATAACTAATTACATGACTCGAGCTAAGACACCTGGTAATACGAAGGATTTCT	KmPHA2_r
ML_241	GAGGTTAAATATGTGTTCTATTTCCTCTAGACCGTTCCATGCGGACTCCCTTG	KmPHA2_S252_TCG_to_TCT_f
ML_234	CAAGGGAGTCCGCATGGAACGGTCTAGAGGAAATAGAACACATATTTAACCTC	KmPHA2_S252_TCG_to_TCT_r
ML_219	TTCCATGCGGACTCCCTTGC	KmPHA2_linear_f
ML_212	GAACACATATTTAACCTCTGATCCTTCAATACGT	KmPHA2_linear_r
ML_237	TTCTCACCATTTCATGGGTGTTACTTTGCATGGTGTTGCTGCTATCACCACTA	ScARO4^{K229L} insert_f
ML_230	TAGTGGTGATAGCAGCAACACCATGCAAAGTAACACCCATGAAATGGTGAGAA	ScARO4^{K229L} insert_r
ML_189	CATGGTGTTGCTGCTATCACCACTAC	ScARO4 linear_f
ML_182	AGTAACACCCATGAAATGGTGAGAATGAG	ScARO4 linear_r
ML_225	AGATGGTGATGATAAGAATAACTTCTCTTCTGTTGCCACTAGAGATATAGAAT	ScARO7^{G141S} insert_f
ML_220	ATTCTATATCTCTAGTGGCAACAGAAGAGAAGTTATTCTTATCATCACCATCT	ScARO7^{G141S} insert_r
ML_223	TCTGTTGCCACTAGAGATATAGAATGTTTG	ScARO7 linear_f
ML_218	GAAGTTATTCTTATCATCACCATCTCTTTTCG	ScARO7 linear_r
ML_211	CCCACATCACTTCATGGGTGTTACCTTGCACGGTGTTGCTGCCATCACCACCA	KmARO4^{K221L} insert_f
ML_206	TGGTGGTGATGGCAGCAACACCGTGCAAGGTAACACCCATGAAGTGATGTGGG	KmARO4^{K221L} insert_r
ML_207	CACGGTGTTGCTGCCATCAC	KmARO4 linear_f
ML_202	GGTAACACCCATGAAGTGATGTGG	KmARO4 linear_r
ML_191	AGAGGGAAACACATCTGAGAATTTTTCTAGTGTGGCTACGAGGGATATCGAAA	KmARO7^{G141S} insert_f
ML_184	TTTCGATATCCCTCGTAGCCACACTAGAAAAATTCTCAGATGTGTTTCCCTCT	KmARO7^{G141S} insert_r
ML_193	AGTGTGGCTACGAGGGATATCG	KmARO7 linear_f
ML_186	AAAATTCTCAGATGTGTTTCCCTCTCTCT	KmARO7 linear_r
ML_265	TCTAGAACTAGTGGATCCCCCGGGATGGCTCCAGTAGTTCTAGACGA	KmARO10_f

Primers	Sequence (5' to 3', priming parts are underlined)	Use
ML_254	ACATAACTAATTACATGACTCGAGTTATTTTG GCTTACCATTTACCAACATAGC	KmARO10_r
ML_259	GAAGAGAACTAAAAAAGAAGGCCGC	KmARO4^{K221L}-CYC1t_f
ML_261	TCGAAGGCTTTAATTTGCGCAGTTTATCATTAT CAATACTCGC	P_{ScTDH3}-ARO10_f
ML_252	AAGCGTGACATAACTAATTACATGAC	P_{ScTDH3}-ARO10_r
ML_245	GGCGAATTGGAGCTCCACCG	ABZ1_upHD _f, URA3_upHD_f
ML_036	GGGTGGTATTGGACCCCATACC	ABZ1_upHD _r
XL_001	TATGGGGTCCAATACCACCCAACACCGATGAA GCAAAGAAG	P_{KmTEF3}_f
XL_002	CACCAGTCATGCTAGCCATCCCGGGCTTTAAT GTTACTTCTCTTGGAGTTAGAAC	P_{KmTEF3}_r
XL_003	TATGGGGTCCAATACCACCCTTACCCTCACTCT TTCACATTAC	P_{KmPGK}_f
XL_004	CACCAGTCATGCTAGCCATCCCGGGTTTTGTA TCTTTATATAGGTAGTGTGTAT	P_{KmPGK}_r
XL_005	TATGGGGTCCAATACCACCCCACTATATCAGG CCTCCACTATTCCA	P_{KmTDH3}_f
XL_006	CACCAGTCATGCTAGCCATCCCGGGTGTGAAT GTGTAAAAGTGTGTGTGTACTGTTG	P_{KmTDH3}_r
ML_331	AAGAGAAGTAACATTAAAGCATGTCAGCTAC ACCACAACCTATG	P_{KmTEF3}-KmARO4^{K221L}_f
ML_320	GACATAACTAATTACATGACTCGAGTTATTTA GCGG	KmARO4^{K221L}_r, KmPHA2_r
ML_333	AAGAGAAGTAACATTAAAGCATGGATTTTTTT AAACCAGAAACTGTTC	P_{KmTEF3}-KmARO7^{G141S}_f
ML_324	AGCGTGACATAACTAATTACATGAC	KmARO7^{G141S}_r
ML_303	AAGAGAAGTAACATTAAAGCATGGTTAAAGT GCTGTATCTAGGGCC	P_{KmTEF3}-KmPHA2_f
ML_263	CCTATATAAAGATACAAAACATGTCAGCTACA CCACAACCTATG	P_{KmPGK}-KmARO4^{K221L}_f
ML_307	CCTATATAAAGATACAAAACATGGATTTTTTT AAACCAGAAACTGTTC	P_{KmPGK}-KmARO7^{G141S}_f
ML_297	CCTATATAAAGATACAAAACATGGTTAAAGTG CTGTATCTAGGGCC	P_{KmPGK}-KmPHA2_f
ML_293	ACACTTTTACACATTCACACATGTCAGCTACA CCACAACCTATG	P_{KmTDH3}-KmARO4^{K221L}_f
ML_309	ACACTTTTACACATTCACACATGGATTTTTTTA AACCAGAAACTGTTC	P_{KmTDH3}-KmARO7^{G141S}_f
ML_311	ACACTTTTACACATTCACACATGGTTAAAGTG CTGTATCTAGGGCC	P_{KmTDH3}-KmPHA2_f
ML_007	GGCATGGATGAACTATACAAATGAC	CYC1t_611bp_ABZ1_downHD_f
ML_192	AGGGAACAAAAGCTGGTACCGGCCGGACCGT AAACTGACCCATTTCATATC	611bp_ABZ1_downHD_r
ML_215	CTGGCCCCAGTGCTGCAATGATACCGCGACTC CCACGCTCACCGGCTCCAGATTTATCA	AmpR_S240_TCT_to_AGT_insert_f

Primers	Sequence (5' to 3', priming parts are underlined)	Use
ML_210	TGATAAATCTGGAGCCGGTGAGCGTGGGAGTCGCGGTATCATTGCAGCACTGGGGCCAG	AmpR_S240_TCT_to_AGT_insert _r
ML_213	TATCATTGCAGCACTGGGGC	611bp_ABZ1_downHD_linear_f
ML_208	TCACCGGCTCCAGATTTATCAGC	611bp_ABZ1_downHD_linear_r
ML_233	GGGTCCAATACCACCCTACCGCGAGACCCACGGGTCTCTGAATCTTGGTGGGAAATTAAT	BsaI(2)_insert_f
ML_226	ATTAATTTCCCACCAAGATTCAGAGACCCGTGGGTCTCGCGGTAGGGTGGTATTGGACCC	BsaI(2)_insert_r
ML_231	TTGGTGGGAAATTAATCCAGAATTTTATTAAATTGGCG	611bp_ABZ1_downHD_AmpR_S240_TCT_to_AGT_linear_f
ML_298	GGGTGGTATTGGACCCCATAC	611bp_ABZ1_downHD_AmpR_S240_TCT_to_AGT_linear_r
ML_377	GGGTCCAATACCACCCTAACACCGATGAAGCAAAGAAGTAACAGC	P_{KmTEF3}-KmARO4^{K221L}-CYC1t_f
ML_366	CGAATGTCTGTTCGACACTGGCAAATTAAAGCCTTCGAGCGTCC	KmARO4^{K221L}-CYC1t_r
ML_365	CAGTGTCGAACAGACATTCGAACACCGATGAAGCAAAGAAGTAACAGC	P_{KmTEF3}-KmARO7^{G141S}-CYC1t_f
ML_368	TCGATACTGGTACTAATGCGGCAAATTAAAGCCTTCGAGCGTCC	KmARO7^{G141S}-CYC1t_r
ML_371	CGCATTAGTACCAGTATCGAAACACCGATGAAGCAAAGAAGTAACAGC	P_{KmTEF3}-KmPHA2-CYC1t_f
ML_370	GTCGCCAATTTAATAAAATTCTGGATTAATTTCCCACCAA	KmPHA2-CYC1t_r
ML_373	CGCATTAGTACCAGTATCGATTACCCTCACTCTTTCACATTACCCTCC	P_{KmPGK}-KmPHA2-CYC1t_f
ML_375	CGCATTAGTACCAGTATCGACACTATATCAGGCCTCCACTATTCCAGG	P_{KmTDH3}-KmPHA2-CYC1t_f
ML_367	CAGTGTCGAACAGACATTCGTTACCCTCACTCTTTCACATTACCCTCC	P_{KmPGK}-KmARO7^{G141S}-CYC1t_f
ML_369	CAGTGTCGAACAGACATTCGCACTATATCAGGCCTCCACTATTCCAGG	P_{KmTDH3}-KmARO7^{G141S}-CYC1t_f
ML_379	GGGTCCAATACCACCCTTTACCCTCACTCTTTCACATTACCCTCC	P_{KmPGK}-KmARO4^{K221L}-CYC1t_f
ML_381	GGGTCCAATACCACCCTCACTATATCAGGCCTCCACTATTCCAGG	P_{KmTDH3}-KmARO4^{K221L}-CYC1t_f
ML_238	GCATTAACAACCCTCTAGGTTCTTTCGTAACTTC	URA3_upHD_r
ML_247	ACCTAGAGGGTTGTTAATGCAACACCGATGAAGCAAAGAAGTAACAGC	URA3_upHD-P_{KmTEF3}_f
ML_015	TCTAGAACTAGTGGATCCCCCGGGATGGCTATTCCAGAAACTCAA	KmADH1_f
ML_018	ACATAACTAATTACATGACTCGAGTTATTTGGAAGTGTCAACGACAA	KmADH1_r

Primers	Sequence (5’ to 3’, priming parts are underlined)	Use
ML_017	TCTAGAACTAGTGGATCCCCCGGGATGTCTATTCCAACTACTCAAAAGG	KmADH2_f
ML_020	ACATAACTAATTACATGACTCGAGTTATTTGGAAGTGTCAACAACGT	KmADH2_r
ML_019	TCTAGAACTAGTGGATCCCCCGGGATGTTTCATAGAAGAGCATTGAAG	KmADH5_f
ML_022	ACATAACTAATTACATGACTCGAGTTAGCATTCATAGGCCTGTCTGA	KmADH5_r
ML_021	TCTAGAACTAGTGGATCCCCCGGGATGTCCTACCCAGATAGTTTCC	KmADH6_f
ML_024	ACATAACTAATTACATGACTCGAGTTATTTTTGAGCCTTGAACTCTCC	KmADH6_r
ML_023	TCTAGAACTAGTGGATCCCCCGGGATGTTTCGTAAGGTCACATCTG	KmADH7_f
ML_026	ACATAACTAATTACATGACTCGAGTTAAAAGTTAATAATAAGTTTCATAGCCTTTT	KmADH7_r
ML_025	TCTAGAACTAGTGGATCCCCCGGGATGTCATCCGAAACCGCAGG	KmSFA1_f
ML_028	ACATAACTAATTACATGACTCGAGTTATTTCTCCAAATCCAAGACGG	KmSFA1_r
ML_027	TCTAGAACTAGTGGATCCCCCGGGATGTCTATCCCAGAAACTCAAAA	ScADH1_f
ML_030	ACATAACTAATTACATGACTCGAGTTATTTAGAAGTGTCAACAACGTATCTAC	ScADH1_r
ML_029	TCTAGAACTAGTGGATCCCCCGGGATGTCTATTCCAGAAACTCAAAAAG	ScADH2_f
ML_032	ACATAACTAATTACATGACTCGAGTTATTTAGAAGTGTCAACAACGTATCTACC	ScADH2_r
ML_269	TCTAGAACTAGTGGATCCCCCGGGATGACTACTGAAAAAACCGTTGTTTTTGTTTCTGG	ScPAR1_f
ML_258	ACATAACTAATTACATGACTCGAGTTAGCTTTTACTTTGAACTTCTAGTAATTGCGAGGC	ScPAR1_r
ML_271	TCTAGAACTAGTGGATCCCCCGGGATGTCTAATACAGTTCTAGTTTCTGGCGCTTC	ScPAR2_f
ML_260	ACATAACTAATTACATGACTCGAGTTATAATCTGTTCTGCTTCTTCAAAATTTGGGCAG	ScPAR2_r
ML_273	TCTAGAACTAGTGGATCCCCCGGGATGACTACTGATACCACTGTTTTCGTTTCTGG	ScPAR3_f
ML_262	ACATAACTAATTACATGACTCGAGTTAGGCTTCATTTTGAACTTCTAACATTTGCGC	ScPAR3_r
ML_275	TCTAGAACTAGTGGATCCCCCGGGATGTCAGTTTTCGTTTCAGGTGCTAACG	ScPAR4_f
ML_264	ACATAACTAATTACATGACTCGAGTTATATTCTGCCCTCAAATTTTAAAATTTGGGAGGC	ScPAR4_r
ML_277	TCTAGAACTAGTGGATCCCCCGGGATGACGTACGTTGTGGTTACTGGTG	KmPAR1_f
ML_266	ACATAACTAATTACATGACTCGAGTTAGTTGTTAGCCTTTAGTATTTGGGCAACAGTATC	KmPAR1_r

Primers	Sequence (5’ to 3’, priming parts are underlined)	Use
ML_279	TCTAGAACTAGTGGATCCCCCGGGATGACATA TACAGTGGTGACAGGCG	KmPAR2_f
ML_268	ACATAACTAATTACATGACTCGAG TACTTACCCACGGTACGCGC	KmPAR2_r
ML_281	TCTAGAACTAGTGGATCCCCCGGGATGTCATA CACGGTAATTACAGGCGC	KmPAR3_f
ML_270	ACATAACTAATTACATGACTCGAGTTAATTCT GAGGAATTCCCCTGACTCTCAAAAG	KmPAR3_r
ML_425	ACCTAGAGGGTTGTTAATGCCCACGCAGTGTG AATGGACTTT	URA3_upHD-P_{KmNC1}_f
ML_414	CACCAGTCATGCTAGCCATCCCGGGTTTTGAT TTGTGTTTAAGCGAGTGACTGAAG	URA3_upHD-P_{KmNC1}_f

Table S2.4. Sequences of *KmARO4*, *KmARO4*^{K221L}, *KmARO7*, *KmARO7*^{G141S}, *KmPHA2*, *KmARO10*, and the codon-optimized *rosePAR*

KmARO4
ATGTCAGCTACACCACAACCTATGTTCCATGAACAAGAAGACGTTAGAATCTTGGGTTAC GATCCATTAGTGTCACCAGCACTACTTCAGGCCCAAGTTCCAGCTTCTCCTGAATGTCTAG CAACAGCTCAAAGAGGTAGAAAGGAGTCTGTTGATATTATTACCGGTAAAGATGACAGA GTGCTGGTTATCGTCGGTCCATGCTCCATCCACGATTTGGATCAAGCACAAGAGTACGCT AAGATGTTGAAGGCTCTTTCGGATGAGTTGAAGGACGATTTGTGTATCATAATGAGAGCT TACTTAGAAAAGCCAAGAACCACTGTTGGCTGGAAGGGTTTGATCAACGATCCAGATGTC GACAACACTTTCAACATCAACAAAGGTTTGCAAGTATCCAGACAATTGTTTGTAAATTTG ACCAGTCTGGGATTACCAATTGGTTCTGAGATGTTGGACACTATCTCTCCTCAATTCTTGG CTGATTTGCTATCTTTCGGTGCCATTGGTGCTAGAACTACCGAGTCTCAATTGCACAGAG AATTGGCCTCTGGTTTATCCTTCCCAGTCGGTTTCAAGAACGGTACCGACGGTACATTGG GTGTTGCCGTCGATGCTGTCCAAGCTGCCTCTCACCCACATCACTTCATGGGTGTTACCAA GCACGGTGTTGCTGCCATCACCACCACCAAGGGTAATGAACACTGCTTCGTTATTCTAAG AGGTGGTAAGAAGGGAACAAACTACGACCCTGCTTCCGTCGCTGAAGCCAAGGCTCAGT TGCCTGAAAAGGGTGTTCTAATGATCGACTACTCCCACGGTAACTCCAACAAGGACTTCA GAAACCAACCTAAGGTTAACGATGTCGTCTGTGAACAAATTGCCAATGGTGAAGATAAG ATCATCGGTGTCATGATCGAATCTAACATCAACGAAGGTAAGCAATGCATTCCTCCAGAA GGTAAGGCTGGCTTGAAATACGGTGTCTCCATCACTGACGGATGCATAAGTTTCGAAACA ACCACCGAGGTCCTACGTAAGCTAGCTGCTGCCGTAAGAGCTAGAAGAGAACTAAAAAA GAAGGCCGCTAAATAA
KmARO4^{K221L}
ATGTCAGCTACACCACAACCTATGTTCCATGAACAAGAAGACGTTAGAATCTTGGGTTAC GATCCATTAGTGTCACCAGCACTACTTCAGGCCCAAGTTCCAGCTTCTCCTGAATGTCTAG CAACAGCTCAAAGAGGTAGAAAGGAGTCTGTTGATATTATTACCGGTAAAGATGACAGA GTGCTGGTTATCGTCGGTCCATGCTCCATCCACGATTTGGATCAAGCACAAGAGTACGCT AAGATGTTGAAGGCTCTTTCGGATGAGTTGAAGGACGATTTGTGTATCATAATGAGAGCT TACTTAGAAAAGCCAAGAACCACTGTTGGCTGGAAGGGTTTGATCAACGATCCAGATGTC GACAACACTTTCAACATCAACAAAGGTTTGCAAGTATCCAGACAATTGTTTGTAAATTTG ACCAGTCTGGGATTACCAATTGGTTCTGAGATGTTGGACACTATCTCTCCTCAATTCTTGG CTGATTTGCTATCTTTCGGTGCCATTGGTGCTAGAACTACCGAGTCTCAATTGCACAGAG AATTGGCCTCTGGTTTATCCTTCCCAGTCGGTTTCAAGAACGGTACCGACGGTACATTGG GTGTTGCCGTCGATGCTGTCCAAGCTGCCTCTCACCCACATCACTTCATGGGTGTTACCTT GCACGGTGTTGCTGCCATCACCACCACCAAGGGTAATGAACACTGCTTCGTTATTCTAAG AGGTGGTAAGAAGGGAACAAACTACGACCCTGCTTCCGTCGCTGAAGCCAAGGCTCAGT TGCCTGAAAAGGGTGTTCTAATGATCGACTACTCCCACGGTAACTCCAACAAGGACTTCA GAAACCAACCTAAGGTTAACGATGTCGTCTGTGAACAAATTGCCAATGGTGAAGATAAG ATCATCGGTGTCATGATCGAATCTAACATCAACGAAGGTAAGCAATGCATTCCTCCAGAA GGTAAGGCTGGCTTGAAATACGGTGTCTCCATCACTGACGGATGCATAAGTTTCGAAACA ACCACCGAGGTCCTACGTAAGCTAGCTGCTGCCGTAAGAGCTAGAAGAGAACTAAAAAA GAAGGCCGCTAAATAA

KmARO7
ATGGATTTTTTTAAACCAGAAACTGTTCTAAATTTGCAGAACATTAGAGATGAGCTCGTG AAGATGGAAGATACGATCATCTTCAATTTCATTGAGAGATCACACTTTGCCACATGTCCC AGTGTGTACCAAAACAAAGATCCATTAGTGAATCTCCCTGACTTTGATGGTAGTTTCTTA GATTGGGCCTTAATGCATGTGGAAATTGTGCAATCTCAATTGAGACGGTTCGAGTCACCA GACGAAACACCCTTCTTTCCAGATAAAATTCTCAAGCCAATCATTCCTAGCTTGAACTAC CCCAAGATTCTGGCGTCCTATTCGAGTCAAATAAACTATACTGACAAAATCAAGAAGATT TACATCGAAACTATTGTACCCTTGATATCCAAGAGAGAGGGAAACACATCTGAGAATTTT GGGAGTGTGGCTACGAGGGATATCGAAACCTTGCATGCTCTAAGCAGAAGGATTCACTTT GGTAAATTTGTAGCTGAAGCAAAGTTTCAAAGTGAAAAGGAAAAATACACTGAGCTTAT CCGTAACAAAGATACAGAAGGAATTATGAAGGCTATCACTAACTCTGCAGTAGAGGAGA AAATTTTGAAAAGACTACAAGTCAAAGCCGAAGTCTACGGTGTTGACCCTACTAATGCAC AGGGCGATAGAAAGATCACACCTGAGTATTTGGTTAGGATTTATAAAGAAATAGTTATCC CCATTACAAAAGAGGTAGAAGTCGAATATCTTTTGAGGAGGTTGGAGGATGAAGAGAAA TGA
KmARO7^{G141S}
ATGGATTTTTTTAAACCAGAAACTGTTCTAAATTTGCAGAACATTAGAGATGAGCTCGTG AAGATGGAAGATACGATCATCTTCAATTTCATTGAGAGATCACACTTTGCCACATGTCCC AGTGTGTACCAAAACAAAGATCCATTAGTGAATCTCCCTGACTTTGATGGTAGTTTCTTA GATTGGGCCTTAATGCATGTGGAAATTGTGCAATCTCAATTGAGACGGTTCGAGTCACCA GACGAAACACCCTTCTTTCCAGATAAAATTCTCAAGCCAATCATTCCTAGCTTGAACTAC CCCAAGATTCTGGCGTCCTATTCGAGTCAAATAAACTATACTGACAAAATCAAGAAGATT TACATCGAAACTATTGTACCCTTGATATCCAAGAGAGAGGGAAACACATCTGAGAATTTT TCTAGTGTGGCTACGAGGGATATCGAAACCTTGCATGCTCTAAGCAGAAGGATTCACTTT GGTAAATTTGTAGCTGAAGCAAAGTTTCAAAGTGAAAAGGAAAAATACACTGAGCTTAT CCGTAACAAAGATACAGAAGGAATTATGAAGGCTATCACTAACTCTGCAGTAGAGGAGA AAATTTTGAAAAGACTACAAGTCAAAGCCGAAGTCTACGGTGTTGACCCTACTAATGCAC AGGGCGATAGAAAGATCACACCTGAGTATTTGGTTAGGATTTATAAAGAAATAGTTATCC CCATTACAAAAGAGGTAGAAGTCGAATATCTTTTGAGGAGGTTGGAGGATGAAGAGAAA TGA
KmPHA2
ATGGTTAAAGTGCTGTATCTAGGGCCTGCTGGCACATACTCACACCAAGCTGTGCTGCAA CAATTCGCCGATGAGGAACTAATTCCAACGAATTCAATCCCTTCTTGCTTCAAAACGCTC ATTGAAGATGAAGATATAGACTATGGGGTTGTGCCATTGGAGAATTCAACGAATGGACA AGTTGTGTTCACATACGATCTTTTCAGAGATTTTATGCAGTCCGAGGAGGAACCAAAGCT TGAAGTTGTTGGGGAACAATACGTTGATATCGCACATTGTCTCATCGCACCAGCACCATT GGAGGTAGACAAGCTCAGAAAAATCGGCACAATATATTCGCATCCACAGGTGTGGGGTC AAGTGAAGGACTATCTTGCTGATCTTGAGAAGAAGTACGGGAAGTTTACAAAAATTGAC TGTAATTCCACATCTGAGGCTGTAACGAAATGCATTCAAGAGTGGGATTCAAAAGAAAG GATAACTCTTGCCATTGGAAGTCGTGCTAGTGCAAAATTAAACAAAGGTTTCATAGTAGA TAGTGGAATTAATGACATAAAGGGGAATACAACAAGGTTTTTGATATTGAAAAGGCGTA AGCCGAATAATCGACTAGTGACAGATTGTCTACCTCCCAACACGGACAGTAAGAAGGTG AATTTAGTGACCTTTGTAACCAAACAGGACGACCCAGGTTCTTTGGTGGACGTTTTAAAC GTATTGAAGGATCAGAGGTTAAATATGTGTTCTATTTCCTCTAGACCGTTCCATGCGGACT CCCTTGCAGGTCGAAAATGGCAGTACTGTTTCTTTATTGAGTTTTATCATAGTGAAACAAC GGACTATGAGTTGTTAATGGAGAAGTTTGACGAGTACTGTGCGAAATGGATTCATTGGGG CAGATTCTACAGAAATCCTTCGTATTACCAGGTGTCTTAG

KmARO10

ATGGCTCCAGTAGTTCTAGACGATAAATCTGCATCTGAATCACCAAGAAGCGACTCTCCA
GTCCACGGCCTCTCCAGTGTTGTCAAGGACATCACTCTTGGTCGTTATGTGTTCGAGAGAT
TGCTCAACTGTGGTTCCAAGACCATTTTCGGTGTTCCAGGTGACTTCAACTTGCCTTTGCT
AGAATACTTGTACGAAGAAGAGTTGGTGCAAAACGGGCTCCAATGGGTTGGTACTTGTA
ACGAATTGAACGCTGCATATGCTGCGGACGGTTACTCCCGTTACACTAGTAAAATCGGTT
GTGTAATCACCACATTTGGTGTCGGTGAATTGTCCGCTTTGAACGGTATTTCTGGTGCCTT
TGCGGAAGATGTTAAGGTCTTGCACATTGTCGGTGTCAGTCCAACCAAGTTCAGAAAGAA
TGACAAGTTCCGTTCCCACAATGTTCACCACTTGGTTCCAGACTTGGATGGTGACAAGGA
ACCAAACCACGAAGTTTACTTCGATATGATTAAGGACAGAGTTTCTTGCTCTAGCGCTTT
CTTGCACGATGTCAAATCTGCACCAGAAAAGATCGACAAGGTTATTGCTGATATTTACAA
GTACAGTAAGCCAGGTTACATCTTTATTCCAGCTGATTTCGCTGATGAAATGGTCTCTAAT
AAGAACTTGGTTGAGACCCCAGTTATAGACTTGCCTTATGTCATTGAAAACACAACCTCT
AAGGATGCTACCAAAAAGGCCGGTGACAAAATCTTGCAATGGCTATACGAATCTAAGAC
TCCATCTGTCTTTTCCGACGCTCTAGTTGGTAGATTTAACTTGAACAAGGATATTCGTGAG
CTAATCAACAAGGTTGACATGTGGAACTTCACCACTGCCATGGCTAAGTCTTCTTTGGAC
GAACATCACCCAAAGCATTTGGGTGTCTACAAGGGTGCTGAAACTGGTAAGGAAATGCA
ATCCATTGTGGAAATGAGTGACTTGATCCTACACTTTGGTCCTTGTAAGAATGAAATCAA
CTTTGGTTACTACACTTTCCGTTACAACGACAACGCCAGAGTTGTTGAACTTTCGAAGGA
CAAGATCACATTCTTCGAAACTAAGGCTAATGGTTCAGTTGAAGTTCAAGAAGCTAACTT
TGCTGCCGTTTTGAAATACATGAATGAAAACCTAGATGTCTCCAAGATTTCCACCGCTTA
CCCATCTGTCTCTAGAACTAGAGAACACATCGAATTTGGTGAAGATGATGAAATTAGTCA
ACAGAGTTTGAAGCGTATAGTGGAGAATTTCTACAACCCTGGTGATGTTTTGGTTGTTGA
AACTGGTTCTTTCCAATTTAACTTGATCAACATGAAATTCGCTCCAGAAATGAAATATAT
GACACAAACATTCTACCTTTCTATTGGTATGGCTTTGCCAGCTGCATTAGGTGTCGGCTGT
GGTATGCGCGACTACCCAAGATCTCACATTATCAACCAATCTGCTGTCCCAGCAGACTAT
GTTCCAAAGTTGATCTTGTGTGAAGGTGATGGTGCTGCTCAAATGACTATTCAAGAATTT
GCTTCTTACATTCGTTACAAGATCCCAATGAACATTCTTTTGTTCAACAACAATGGTTACA
CTATTGAAAGAGCTATTCTAGGACCAACCAGAAGTTATAATGATATTGCACCAGTTAAGT
GGACCGCTTTGCTTAACGCTTTTGGTGACTTTGAAAACAAATTCAGTGAAACTGTTACTGT
TTCCAAGAATAAGGAAATCATTGAAGTTTTGAACGAATGGAAAAAGGAGAAGGTTCCTT
CTAAGATCAAGTTGGCAGAAGTTATGTTACCAGTTATGGATATTCCTTCTGAGTTGGATG
CTATGTTGGTAAATGGTAAGCCAAAATAA

codon-optimized *rosePAR*
ATGTCTAACAAGGTTGTTTGTGTTACCGGTGCATCTGGTTACATAGCAAGTTGGCTTGTTA AGCTTCTTCTTCAACGAGGTTACACCGTTAAGGCATCTGTGCGAAACCCAAACGATCCAA CCAAGACCGAACACTTGCTTGCACTTGATGGTGCAAAGGAGAGACTTCAACTTTTCAAGG CAGATCTTTTGGAAGAAGGTTCTTTCGATTCTGCAGTTGAAGGTTGTGAAGGTGTTTTCCA CACCGCATCTCCATTCTACCACGATGTTACCGATCCAAAGGCAGAACTTCTTGATCCAGC AGTTAAGGGTACCCTTAACGTTCTTAACTCTTGTTCTAAGTCTCCATCTATAAAGCGAGTT GTTTTGACCTCTTCTATAGCAGCAGTTGCATACAACGGTAAGCCACGAACCCCAGATGTC GTGGTTGATGAAACCTGGTTCACCGATCCAGATGTTTGTAAGGAATCTAAGCTTTGGTAC GTTCTTTCTAAGACCTTGGCAGAAGATGCAGCATGGAAGTTCGTTAAGGAAAAGGGTATA GATATGGTTACCATAAACCCAGCAATGGTTATAGGTCCACTTTTGCAACCAACCCTTAAC ACCTCTGCTGCAGCAATACTTAACATAATAAAGGGTGCACGAACCTACCCAAACGCATCT TTCGGTTGGATAAACGTTAAGGATGTTGCAAACGCACACGTTCAAGCATTCGAAATACCA TCTGCATCTGGTAGATACTGTCTTGTTGAACGAGTTGCACACTTCACCGAAGTTCTTCAAA TAATACACGAACTTTACCCAGATTTGCAACTTCCAGAAAAGTGTTCTGATGATAAGCCAT TCGTTCCAACCTACCAAGTTTCTAAGGAAAAGGCAAAGTCTTTGGGTATAGAATTCATAC CACTTGATATATCTCTTAAGGAAACCATAGAATCTTTGAAGGAAAAGTCTATAGTTTCTTT CTGA

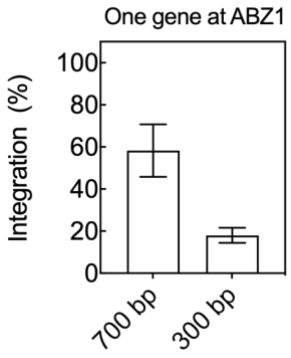

Figure S2.1. Homology donor length effects on the CRISPR-mediated gene integration in *K. marxianus* CBS 6556 *ura3Δ his3Δ*. All experiments were performed in biological triplicates. Bars represent the arithmetic mean, and error bars represent the standard deviation.

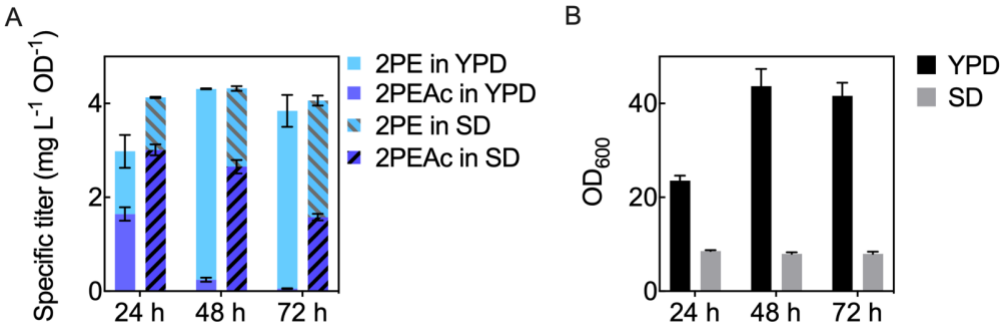

Figure S2.2. Medium effects on 2-PE acetylation and biomass accumulation of *K. marxianus* CBS 6556 *ura3Δ his3Δ*. (A) Specific titer of 2-PE and 2-PEAc after 24, 48, and 72 h. (B) OD_{600} at each time point for extracellular 2-PE and 2-PEAc measurement. The strain was cultivated in 25 mL YPD or synthetic defined medium with complete supplement mixture (SD) with 20 g/L D-glucose for biological triplicates at 30 °C from an initial OD_{600} of 0.05. All experiments were performed in biological triplicates. Bars and dots represent the arithmetic mean, while error bars represent the standard deviation.

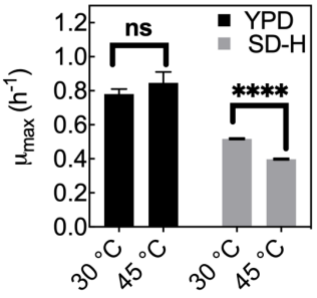

Figure S2.3. Undesirable effects of plasmid reliance on cell growth at elevated temperature for *K. marxianus* CBS 6556 *ura3Δ his3Δ*. The strain was inoculated in 25 mL YPD or SD-H with 20 g/L D-glucose at an initial OD_{600} of 0.05 and was cultivated at 30 °C shaker till the end of exponential phase. Statistically significant difference indicated with "ns" for P value > 0.12, and "****" for P value < 0.0001 in t tests. All experiments were performed in biological triplicates. Bars represent the arithmetic mean, and error bars represent the standard deviation.

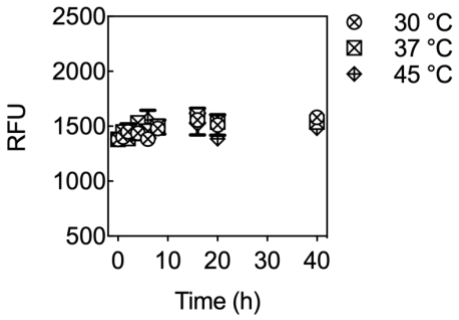

Figure S2.4. EGFP stability over 40 h incubation at 30, 37, and 45 °C. The *K. marxianus* strain harboring *EGFP* expression plasmid (P_{KmTEF3}-*EGFP*-*ScCYC1*t) was cultured in 25 mL SD-H medium at 30 °C from an initial OD_{600} of 0.05. After 14 h, cells were harvested by centrifugation at 5000 g for 10 min at 4 °C. Supernatants were discarded and cell pellets were washed and resuspended with 100 ml PBS. The cell resuspension was lysed by sonication for 90 s. The sonication process was carried out on ice, and there was a 7 s pause for every 3 s sonication. After centrifugation at 11,000 g for 30 min at 4 °C, the lysate supernatant was collected and incubated. The 20× diluent was then measured at 0, 1, 2, 4, 6, 8, 16, 20, 40 hours for relative green fluorescence intensity by BioTek®Synergy™ Neo2 multi-mode microplate Reader (Ex/Em 488/511). PBS was used as the blank of measurement, and all data was background subtracted.

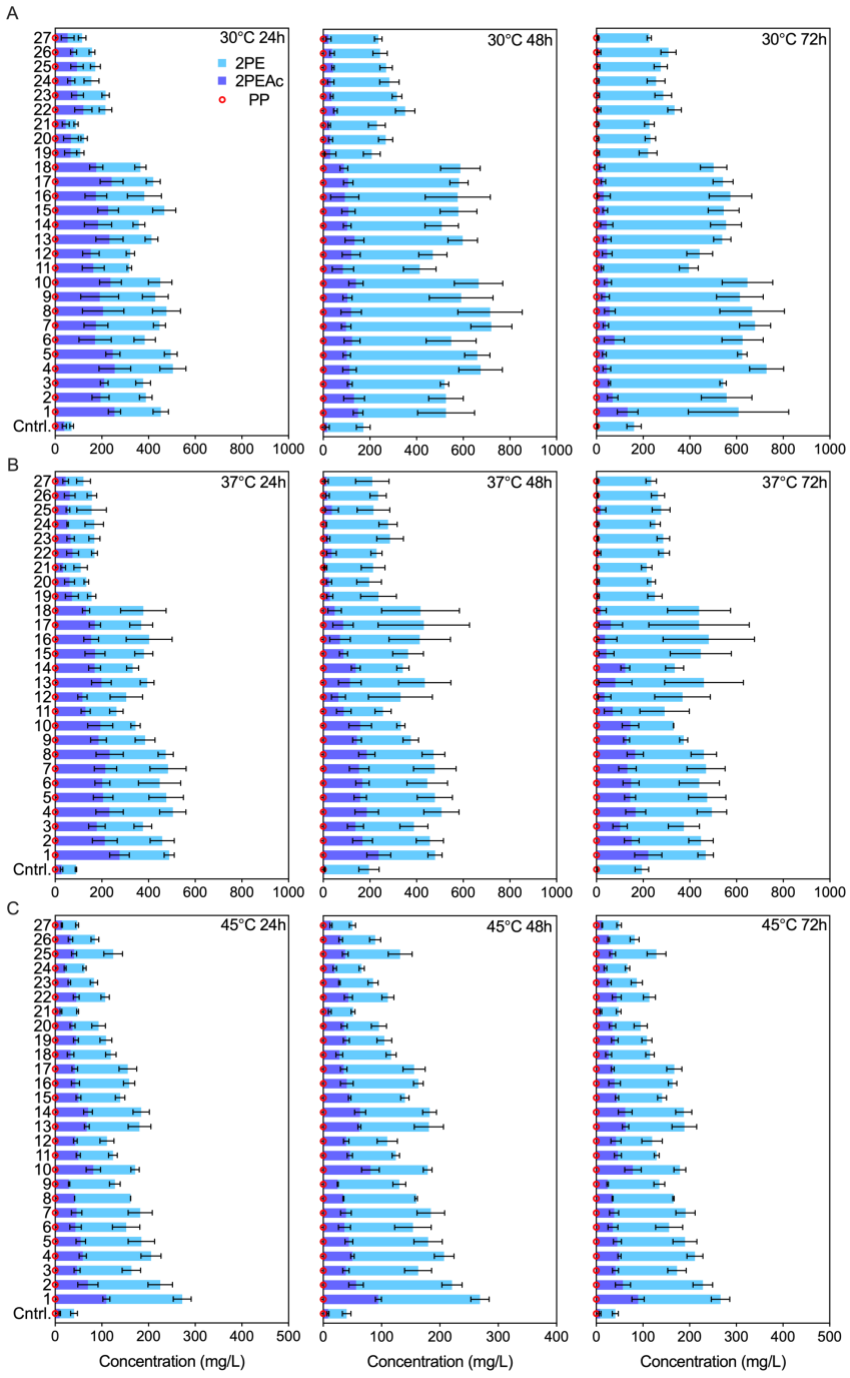

Figure S2.5. Extracellular formation of 2-phenylethanol (2-PE), 2-phenylethyl acetate (2-PEAc), and phenylpyruvate (PP) after 24, 48, and 72 h cultivation at (A) 30, (B) 37, and (C) 45 °C. Cell cultures were initiated from an OD_{600} of 0.05 in 25 mL YPD medium with 20 g/L glucose. The strain Cntrl. was *K. marxianus* CBS 6556 *ura3Δ his3Δ* as the negative control. Numbers 1 to 27 denoted corresponding gene overexpression combinations integrated into the genome of Cntrl. All experiments are performed at least in biological triplicates. Bars and dots represent arithmetic means, and error bars represent standard deviations.

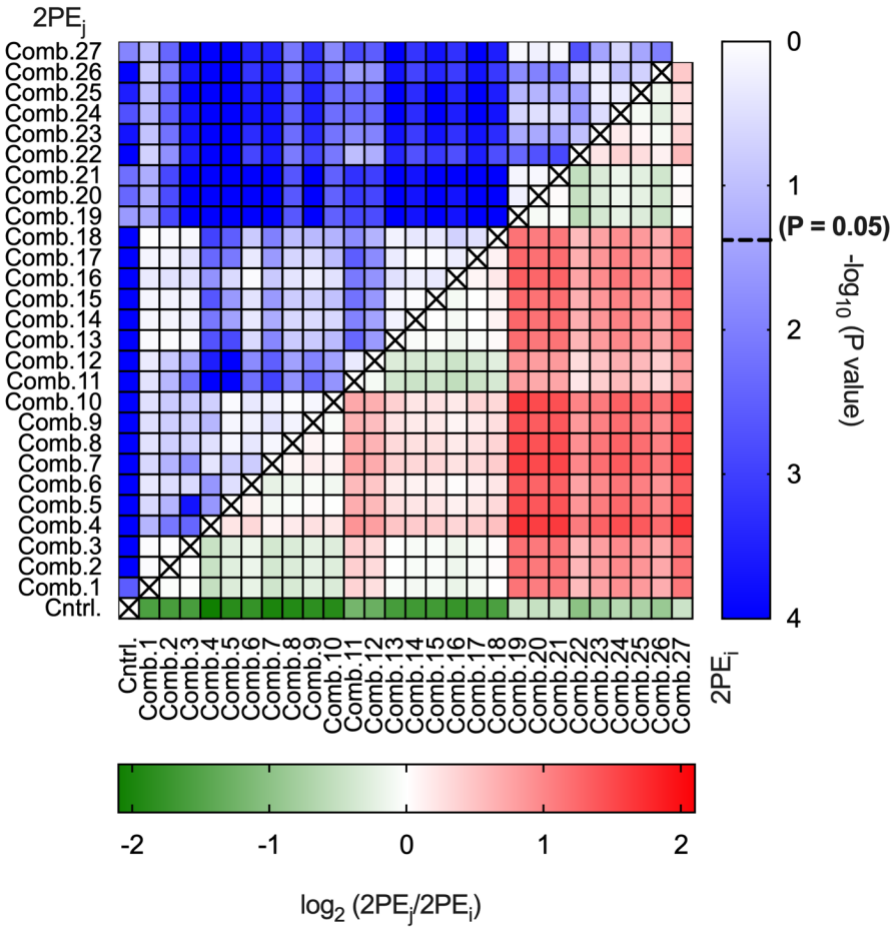

Figure S2.6. Effects of varied expression combinations of *KmARO4*^{K221L}, *KmARO7*^{G141S}, and *KmPHA2* on 2-PE accumulation after 72 h cultivation at 30 °C. Cell cultures were initiated from an OD_{600} of 0.05 in 25 mL YPD medium with 20 g/L glucose. The strain Cntrl. was *K. marxianus* CBS 6556 *ura3Δ his3Δ* as the negative control. Comb. 1 to 27 denoted the 3^3-factorial gene overexpression combination library integrated into the genome of Cntrl. The fold change (denoted by $\log_2(2PE_j/2PE_i)$) and the significance of difference (denoted by $-\log_{10}$(P value)) were calculated by the arithmetic means of at least triplicate biological samples.

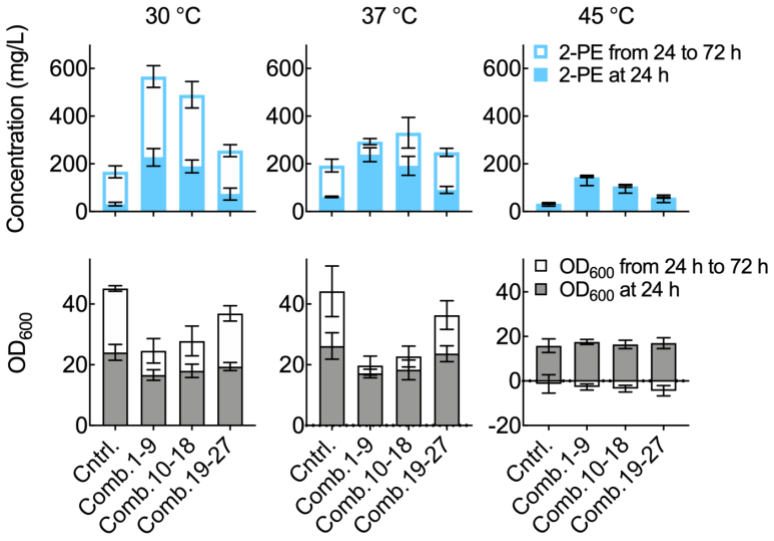

Figure S2.7. Temperature effects on 2-PE formation and biomass accumulation. The 2-PE concentration and biomass denoted by OD_{600} after 24 h cultivation, and their increase from 24 to 72 h were depicted at 30, 37, and 45 °C. Strains were cultivated in 25 mL YPD medium with 20 g/L D-glucose from an initial OD_{600} of 0.05. The 27 engineered strains were grouped into three subsets based on the overexpression level of *KmARO4*^{K221L}. The strain Cntrl. was *K. marxianus* CBS 6556 *ura3Δ his3Δ* as the negative control. Bars are arithmetic means of at least triplicate biological samples, and error bars are standard deviations.

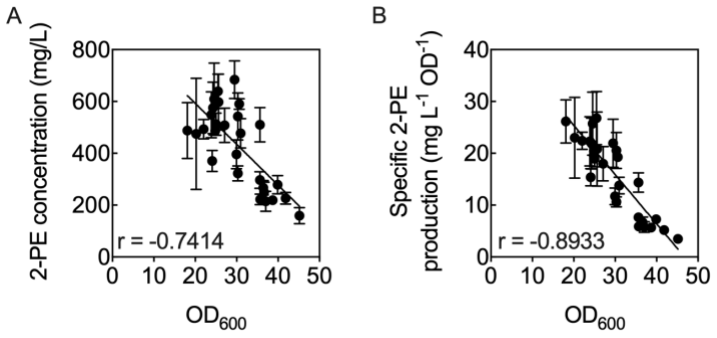

Figure S2.8. Correlation of the (A) 2-PE concentration and (B) specific 2-PE production with biomass accumulation (shown by OD_{600}) after 72 h cultivation at 30 °C. Cell cultures were initiated from an OD_{600} of 0.05 in 25 mL YPD medium with 20 g/L D-glucose. The linear correlation of the two variables in both (A) and (B) are measured by the Pearson correlation coefficient, r.

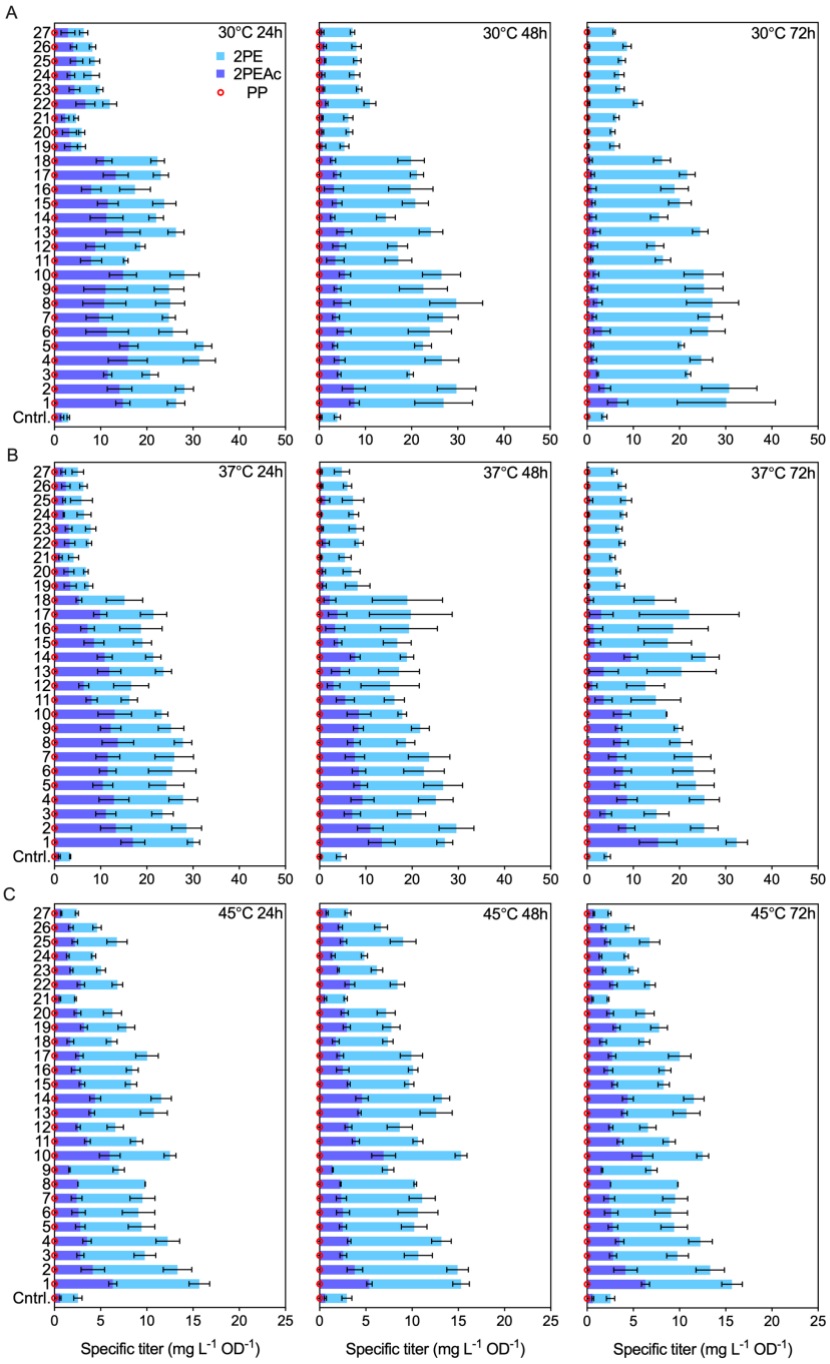

Figure S2.9. Specific production (mg L^{-1} OD^{-1}) of 2-PE, 2-PEAc, and PP after 24, 48 and 72 h cultivation at (A) 30, (B) 37, and (C) 45 °C. Cell cultures were initiated from an OD_{600} of 0.05 in 25 mL YPD medium with 20 g/L glucose. The strain Cntrl. was *K. marxianus* CBS 6556 *ura3Δ his3Δ* as the negative control. Numbers 1 to 27 denoted corresponding gene overexpression combinations integrated into the genome of Cntrl. All experiments are performed at least in biological triplicates. Bars and dots represent arithmetic means, and error bars represent standard deviations.

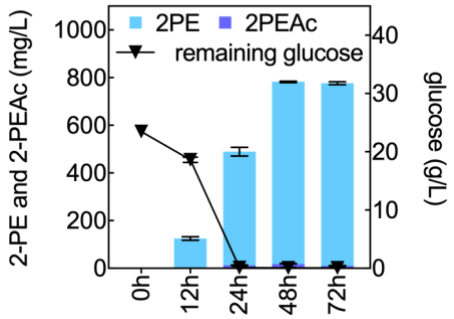

Figure S2.10. Extracellular 2-PE and 2-PEAc formation with glucose consumption in *K. marxianus* CBS 6556 *his3Δ eat1Δ ura3Δ*::(P_{KmTEF3})*KmARO10*, *abz1*::(P_{KmTEF3})*KmARO4*^{K221L}-(P_{KmPGK})*KmARO7*^{G141S}-(P_{KmTDH3})-*KmPHA2*. The strain was cultivated at 30 °C for 72 h in 25 mL YPD with 20 g/L D-glucose from an initial OD_{600} of 0.05. Bars and dots represent arithmetic means, and error bars represent standard deviations.

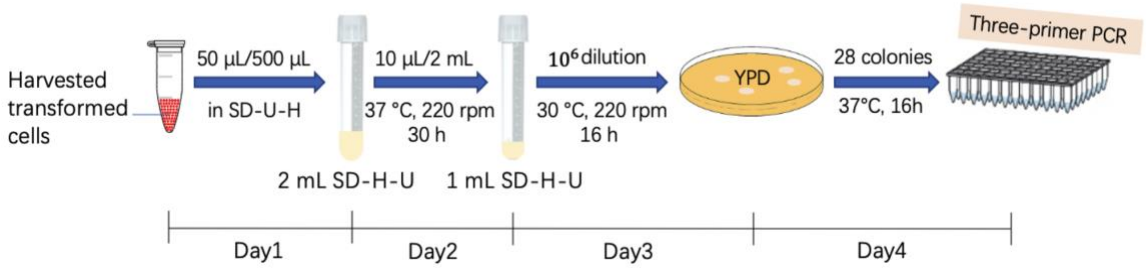

Figure S2.11. Schematic representation of the standard protocol for CRISPR-Cas9 mediated one-step integration up to three genes through yeast transformation to integration confirmation.

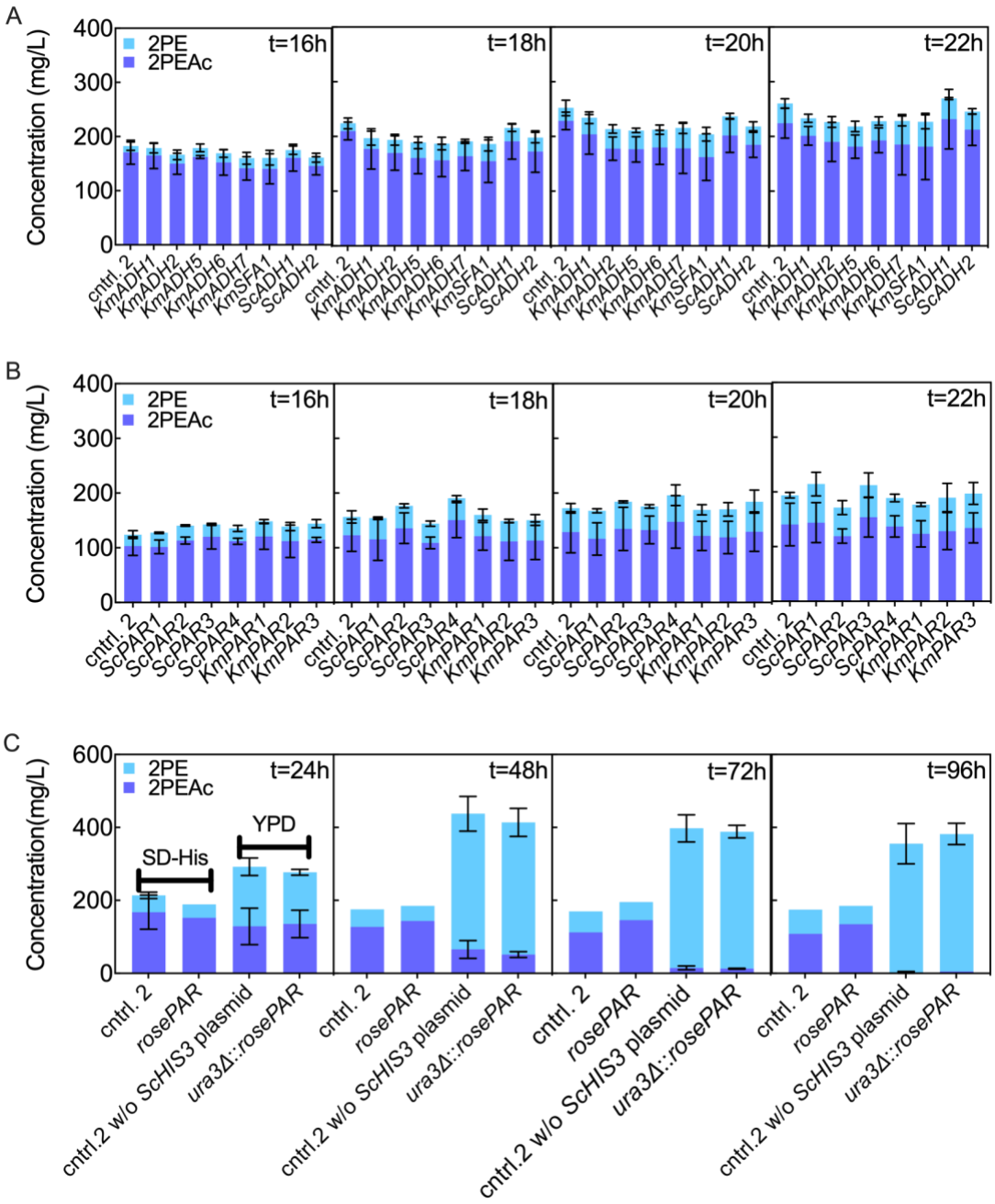

Figure S2.12. Effects of plasmid-based or chromosomal overexpression of homologous and heterologous genes encoding alcohol dehydrogenases (Adh) or phenylacetaldehyde reductases (Par) on 2-PE and 2-PEAc biosynthesis. The strain cntrl. 2 was *K. marxianus* CBS 6556 *ura3Δ his3Δ abz1*::(P_{ScTDH3})*KmARO4*^{K221L}-(P_{ScTDH3})*KmARO10* harboring a blank vector expressing *ScHIS3*. Both selective medium SD-H and rich medium YPD contained 20 g/L D-glucose in the beginning of yeast cultivation with an OD_{600} of 0.05. Bars represent arithmetic means, and error bars represent standard deviation.

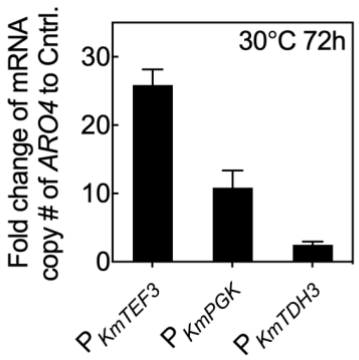

Figure S2.13. Relative transcriptional activity of the 700-bp natively derived *KmTEF3*, *KmPGK*, and *KmTDH3* promoter for *KmARO4* (including wild-type *KmARO4* and *KmARO4*^{K221L}) expression. The strain Cntrl. was *K. marxianus* CBS 6556 *ura3Δ his3Δ*. mRNA was isolated after 72 h cultivation in 25 mL YPD medium with 20 g/L D-glucose at 30 °C.

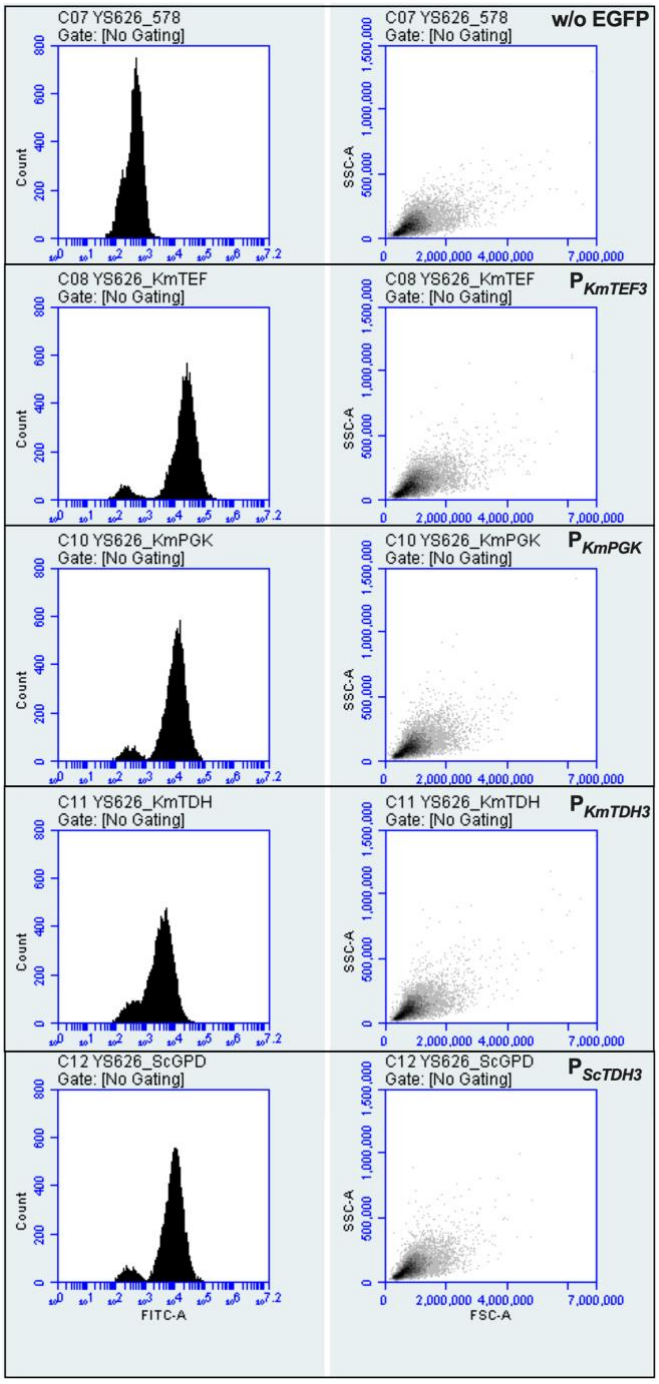

Figure S2.14. Histogram events of EGFP expression driven by P_{KmTEF3}, P_{KmPGK}, P_{KmTDH3}, and P_{ScTDH3} on plasmid at 30 °C, and corresponding cell size distribution.

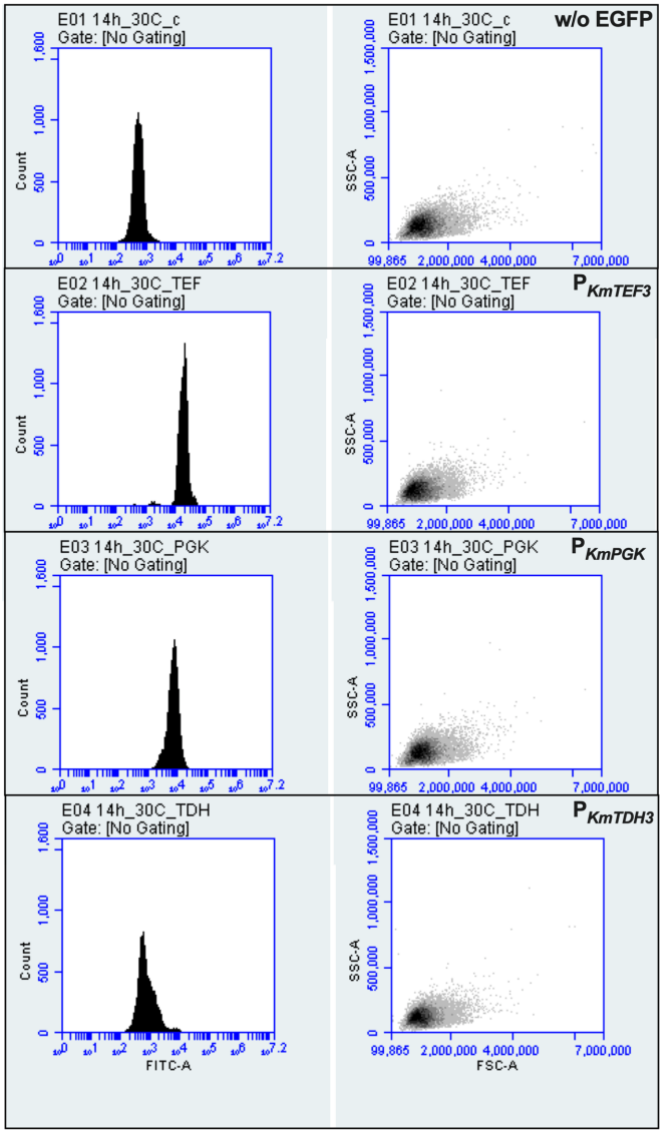

Figure S2.15. Histogram events of integrated EGFP expression driven by P_{KmTEF3}, P_{KmPGK}, and P_{KmTDH3} at 30 °C, and corresponding cell size distribution.

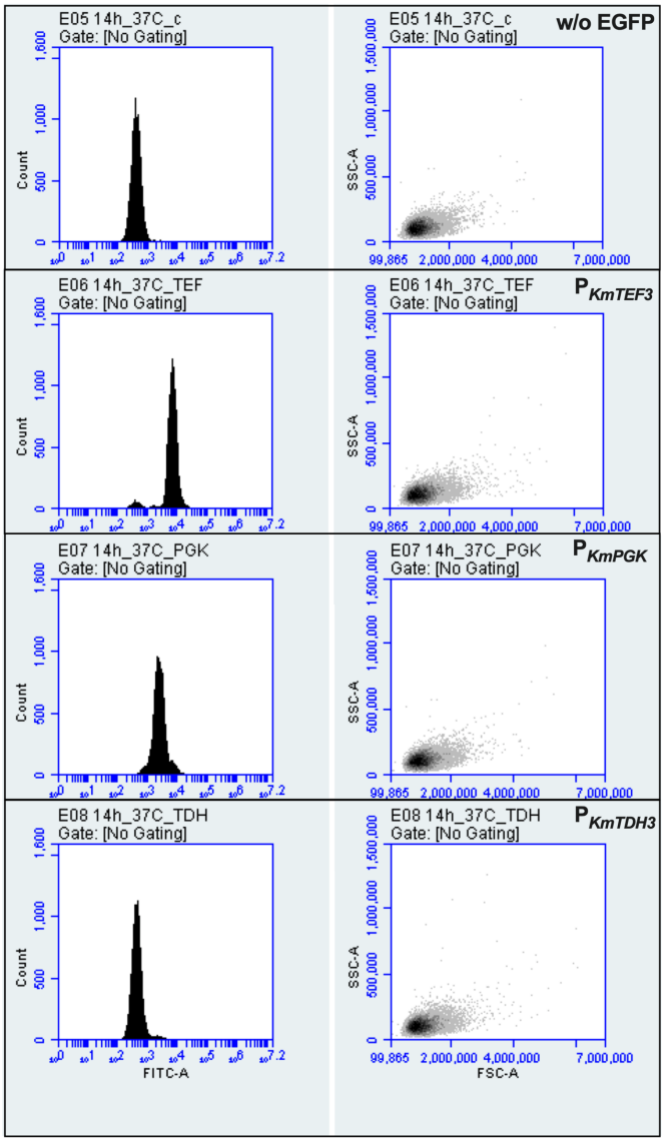

Figure S2.16. Histogram events of integrated EGFP expression driven by P_{KmTEF3}, P_{KmPGK}, and P_{KmTDH3} at 37 °C, and corresponding cell size distribution.

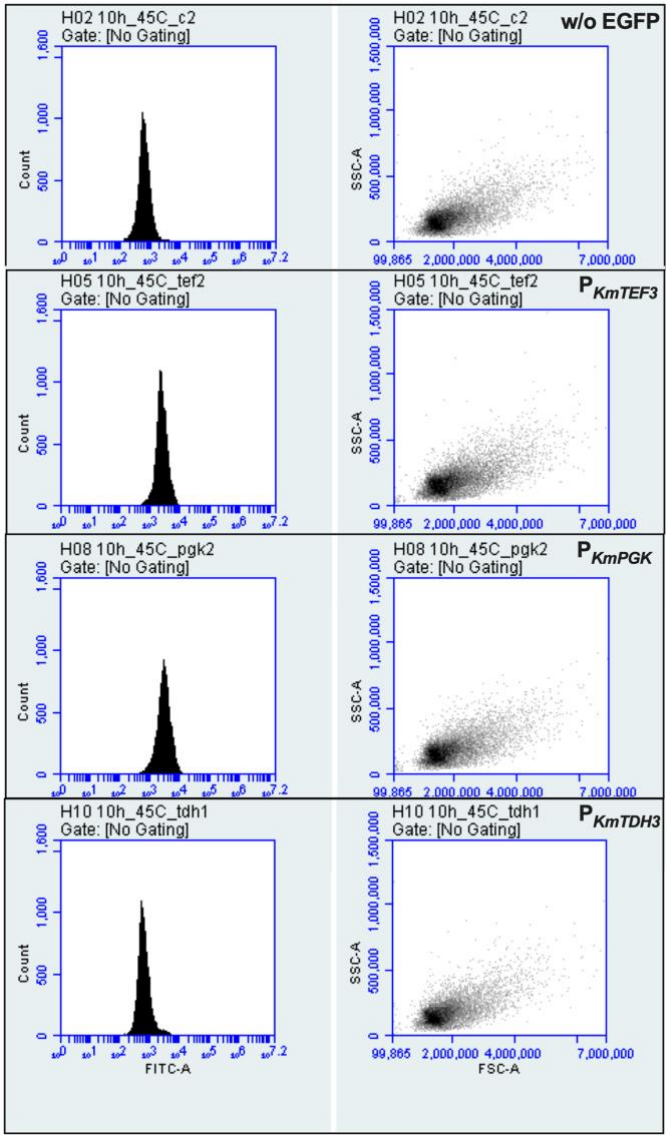

Figure S2.17. Histogram events of integrated EGFP expression driven by P_{KmTEF3}, P_{KmPGK}, and P_{KmTDH3} at 45 °C, and corresponding cell size distribution.

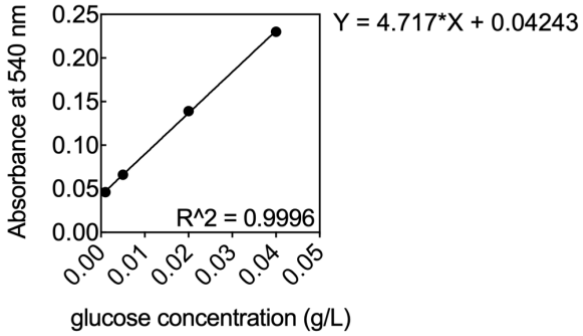

Figure S2.18. Standard curve of glucose concentration (g/L) to absorbance at 540 nm. 0.005 to 0.04 g/L of glucose solution is in the linear correlation region with the absorbance.

2.7 Reference

[1] Fabre CE, et al. 2-Phenylethyl alcohol: an aroma profile. Perfum Flavor. 1998;23:43-45.

[2] Keasling JD and Chou H. Metabolic engineering delivers next-generation biofuels. Nat Biotechnol. 2008;26(3):298-299.

[3] van Summeren-Wesenhagen PV and Marienhagen J. Putting bugs to the blush: metabolic engineering for phenylpropanoid-derived products in microorganisms. Bioengineered. 2013;4(6):355-362.

[4] Etschmann M, et al. Biotechnological production of 2-phenylethanol. Appl Microbiol and Biotechnol. 2002;59(1):1-8.

[5] Krings U and Berger RG. Biotechnological production of flavours and fragrances. Appl Microbiol and Biotechnol. 1998;49(1):1-8.

[6] Cao M, et al. Building microbial factories for the production of aromatic amino acid pathway derivatives: From commodity chemicals to plant-sourced natural products. Metab Eng. 2020;58:94-132.

[7] Löbs AK, et al. Highly Multiplexed CRISPRi Repression of Respiratory Functions Enhances Mitochondrial Localized Ethyl Acetate Biosynthesis in *Kluyveromyces marxianus*. ACS Synth Biol. 2018;7(11):2647-2655.

[8] Lin JL and Wheeldon I. Dual N- and C-Terminal Helices Are Required for Endoplasmic Reticulum and Lipid Droplet Association of Alcohol Acetyltransferases in *Saccharomyces cerevisiae*. PLOS ONE. 2014;9(8):e104141.

[9] Luttik MAH, et al. Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: Quantification of metabolic impact. Metab Eng. 2008;10(3):141-153.

[10] Shen L, et al. Overexpressing enzymes of the Ehrlich pathway and deleting genes of the competing pathway in *Saccharomyces cerevisiae* for increasing 2-phenylethanol production from glucose. J Biosci Bioeng. 2016;122(1):34-39.

[11] Guo D, et al. Metabolic Engineering of *Escherichia coli* for Production of 2-Phenylethanol and 2-Phenylethyl Acetate from Glucose. J Agric Food Chem. 2018;66(23):5886-5891.

[12] Chen XM, et al. Functional characterization of rose phenylacetaldehyde reductase (PAR), an enzyme involved in the biosynthesis of the scent compound 2-phenylethanol. J Plant Physiol. 2011;168(2):88-95.

[13] Gao F and Daugulis AJ. Bioproduction of the aroma compound 2-Phenylethanol in a solid–liquid two-phase partitioning bioreactor system by *Kluyveromyces marxianus*. Biotechnol Bioeng. 2009;104(2):332-339.

[14] Wittmann C, et al. Metabolic physiology of aroma-producing *Kluyveromyces marxianus*. Yeast. 2002;19(15): 1351-1363.

[15] Löbs AK, et al. Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. Synth Syst Biotechnol. 2017; 2(3):198-207.

[16] Schwartz C, et al. Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast *Yarrowia lipolytica*. Metab Eng. 2019;55:102-110.

[17] Schwartz C, et al. Multiplexed CRISPR Activation of Cryptic Sugar Metabolism Enables *Yarrowia Lipolytica* Growth on Cellobiose. Biotechnol J. 2018;13(9):1700584.

[18] Cao M, et al. A genetic toolbox for metabolic engineering of *Issatchenkia orientalis*. Metab Eng. 2020;59:87-97.

[19] Cao M, et al. CRISPR–Mediated Genome Editing and Gene Repression in *Scheffersomyces stipitis*. Biotechnol J. 2018;13(9):1700598.

[20] Dalvie NC, et al. Host-Informed Expression of CRISPR Guide RNA for Genomic Engineering in *Komagataella phaffii*. ACS Synth Biol. 2020;9(1):26-35.

[21] Young EM, et al. Iterative algorithm-guided design of massive strain libraries, applied to itaconic acid production in yeast. Metab Eng. 2018;48:33-43.

[22] Babaei M, et al. Metabolic Engineering of *Saccharomyces cerevisiae* for Rosmarinic Acid Production. ACS Synth Biol. 2020;9(8):1978-1988.

[23] Löbs AK, et al. CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol Biofuels. 2017;10(1):164.

[24] Nambu-Nishida Y, et al. Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. Sci Rep. 2017;7(1):8993.

[25] Rajkumar AS, et al. Biological Parts for *Kluyveromyces marxianus* Synthetic Biology. Front Bioeng and Biotechnol. 2019;7(97).

[26] Lang X, et al. Developing a broad-range promoter set for metabolic engineering in the thermotolerant yeast *Kluyveromyces marxianus*. Metab Eng Commun. 2020;11:e00145.

[27] Schwartz C, et al. Standardized Markerless Gene Integration for Pathway Engineering in *Yarrowia lipolytica*. ACS Synth Biol. 2017;6(3):402-409.

[28] Nonklang S, et al. High-Temperature Ethanol Fermentation and Transformation with Linear DNA in the Thermotolerant Yeast *Kluyveromyces marxianus* DMKU3-1042. Appl and Environ Microbiol. 2008;74(24):7514.

[29] Fu X, et al. Understanding the stress responses of *Kluyveromyces marxianus* after an arrest during high-temperature ethanol fermentation based on integration of RNA-Seq and metabolite data. Appl Microbiol Biotechnol. 2019;103(6):2715-2729.

[30] Wu XL, et al. Genome-wide landscape of position effects on heterogeneous gene expression in *Saccharomyces cerevisiae*. Biotechnol Biofuels. 2017;10(1):189.

[31] Lee KS, et al. Characterization of *Saccharomyces cerevisiae* promoters for heterologous gene expression in *Kluyveromyces marxianus*. Appl Microbiol Biotechnol. 2013;97(5):2029-2041.

[32] Romagnoli G, et al. Deletion of the *Saccharomyces cerevisiae* ARO8 gene, encoding an aromatic amino acid transaminase, enhances phenylethanol production from glucose. Yeast. 2015;32(1):29-45.

[33] Thorwall S, et al. Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. Nat Chem Biol. 2020;16(2):113-121.

[34] Pecota DC, et al. Sequential gene integration for the engineering of *Kluyveromyces marxianus.* J Biotechnol. 2007;127(3):408-416.

[35] Heo P, et al. Simultaneous integration of multiple genes into the *Kluyveromyces marxianus* chromosome. J Biotechnol. 2013;167(3):323-325.

[36] Hassing EJ, et al. Connecting central carbon and aromatic amino acid metabolisms to improve de novo 2-phenylethanol production in *Saccharomyces cerevisiae*. Metab Eng. 2019;56:165-180.

[37] Kim TY, et al. Biosynthesis of 2-phenylethanol from glucose with genetically engineered *Kluyveromyces marxianus*. Enzyme Microb Technol. 2014;61-62:44-47.

[38] Gu Y, et al. Engineering *Yarrowia lipolytica* as a chassis for de novo synthesis of five aromatic-derived natural products and chemicals. ACS Synth Biol. 2020;9(8):2096-2106

[39] Sekar BS, et al. Production of Natural 2-Phenylethanol from Glucose or Glycerol with Coupled *Escherichia coli* Strains Expressing l-Phenylalanine Biosynthesis Pathway and Artificial Biocascades. ACS Sustainable Chem Eng. 2019;7(14):12231-12239.

[40] Hartmann M, et al. Evolution of feedback-inhibited β/α barrel isoenzymes by gene duplication and a single mutation. Proc Natl Acad Sci USA. 2003;100(3):862.

[41] Schnappauf G, et al. Tyrosine and Tryptophan Act through the Same Binding Site at the Dimer Interface of Yeast Chorismate Mutase. J Biol Chem. 1998;273(27):17012-17017.

[42] Krappmann S, et al. Coevolution of transcriptional and allosteric regulation at the chorismate metabolic branch point of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 2000;97(25);13585-13590.

[43] Kruis AJ, et al. Ethyl acetate production by the elusive alcohol acetyltransferase from yeast. Metab Eng. 2017;41:92-101.

[44] McTaggart TL, et al. Synthesis of polyketides from low cost substrates by the thermotolerant yeast *Kluyveromyces marxianus*. Biotechnol Bioeng. 2019;116(7):1721-1730.

[45] Lertwattanasakul N, et al. Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. Biotechnol Biofuels. 2015;8(1):47.

[46] Signori L, et al. Effect of oxygenation and temperature on glucose-xylose fermentation in *Kluyveromyces marxianus* CBS712 strain. Microb Cell Fact. 2014;13(1):51.

[47] Fabre CE, et al. Production of 2-phenylethyl alcohol by *Kluyveromyces marxianus*. Biotechnol Prog. 1998;14(2):270-274.

[48] Zhu J, et al. Microbial host selection affects intracellular localization and activity of alcohol-O-acetyltransferase. Microb. Cell Factories, 2015;14(1):35.

[49] Lin JL, et al. Synthetic Protein Scaffolds for Biosynthetic Pathway Colocalization on Lipid Droplet Membranes. ACS Synth Biol. 2017;6(8):1534-1544.

[50] Shao Y, et al. Creating a functional single-chromosome yeast. Nature. 2018;560(7718), 331-335.

Chapter 3: Optimized Transformation Methods Fulfill Genome-wide Screening in Kluyveromyces marxianus

3.1 Abstract

Highly efficient transformation facilitates functional genomic studies by enabling adequate representation of a complex screening library with less host cells and transformed DNA. Our optimized chemical and heat shock induced method achieves 6×10^5 transformants per reaction on average, using 1.04 pmol DNA (equal to 4 μg for 6,275-bp plasmids) and 5×10^8 of cells. Electroporation gives an even higher yield of transformants than the chemical method with using much less DNA (0.26 pmol) and similar number of cells, obtaining 1.5×10^6 transformants from a single reaction. This is also the current highest transformation efficiency reported for *K. marxianus* in literature.

3.2 Introduction

The lack of high-efficiency transformation methods hampers functional genome-wide deletion of the non-conventional yeast, *Kluyveromyces marxianus*. The current transformation protocol [1] is not efficient enough so that it requires pooling a considerable number of transformations together to achieve acceptable coverage of a genome-wide library, which results in a too high initial density of cells without DNA successfully transferred into. Those unsuccessful transformed cells still try to compete for limited nutrients in the medium after inoculation but are no longer viable because of the auxotrophic selection pressure. They do harm to the overall cell outgrowth. The OD_{600}/mL of the culture at cell confluence is significantly lower than that of a typical

K.marxianus growth in the same medium. For example, starting from a culture of OD_{600}/mL = 4, it only reached 8 at cell confluence in the selective minimal medium without uracil, while a typical OD_{600}/mL for *K.marxianus* cultures at stationary phase in the same medium achieves 12 to 15.

To satisfy the 100× coverage requirement for a whole genome CRISPR plasmid library whose complexity is 33,894 as well as to minimize the number of inoculating cells, it is necessary to further improve the current heat shock and chemical induced transformation in *K. marxianus* [1]. In comparison with the chemical/heat shock method, electrical transformation, though sensitive to salt so that requiring thorough cell washing by high-resistance media (>600 Ω) and concentrated DNA (for as little disturbance to the high-resistance sample environment as possible), has the potential to yield even higher transformation rates by orders of magnitude [2]. Hence, electroporation is amenable to highly complex DNA samples such as genome-wide CRISPR guide libraries, and in the meanwhile, with lower quantities of DNA needed.

3.3 Results and Discussion

A couple of transformation parameters of the heat shock and chemical induced method were evaluated for their effects on plasmid transformation in *K. marxianus* CBS 6556. Overall speaking, adding carrier DNA does no harm in transformation [3,4], so in this study, 10 µL of 10 mg/mL carrier DNA (R&D Systems™ salmon sperm DNA) in a 500 µL transformation mix was a fixed parameter during the protocol optimization. The adequate number/concentration of cells for transformation, the heat shock strokes

inducing the most efficient foreign DNA passage through cell membrane, and the optimal concentration of polyethylene glycol 3350 (PEG 3350), lithium acetate (LiAc), and dithiothreitol (DTT) to enhance cell competence, were critical to achieve an efficient transformation in *K. marxianus* [4]. It is also crucial to harvest cells at the appropriate growth phase to maximize their capability to become highly competent after pretreatment. The optimum growth phase can be various among different yeast species. As for *Saccharomyces cerevisiae* and *Pichia pastoris*, the cell growth is stopped at the early exponential phase (OD_{600}/mL = 1; MicroPulser™ Instruction Manual and Applications Guide, Catalog no. 1652100, Bio-Rad, USA) for the best transformation practice, whereas *Yarrowia lipolytica* [5] and *K. marxianus* prefer stationary phase.

Therefore, the transformation method was optimized one step at a time from the following five aspects accordingly: solution for cell washing (Table 3.1), temperature and duration of heat shock (Table 3.2), number of cells and transformed DNA (Table 3.3), concentration of pretreatment chemical reagents (Table 3.4), and cell growth phase (Table 3.5). This new optimum transformation by heat shock and chemical reagent induction makes it possible to obtain 6×10^5 transformants per reaction, which is seven times more efficient than our current optimal [1] for *K. marxianus*.

When washing cells with buffer containing 0.1 M LiAc and 1× Tris EDTA (TE, pH 7.5) to get rid of the leftover culturing medium, transformation efficiency was 10% slightly higher than using ddH_2O (Table 3.1).

Table 3.1. Heat shock and chemical induced transformation of *K. marxianus* CBS 6556 ura3Δ with various solutions for cell washing (averaged by at least three biological replicates)

Cell washing solution	Average transformants no. (cfu)	Average fold-enhancement
ddH_2O	8.0×10^4	1.0
0.1 M LiAc, 1x TE	9.1×10^4	1.1

One of the most important characteristics of *K. marxianus* is its ability to grow at much higher temperatures than other yeasts. Although it is thermotolerant, there is still an upper limit of heat shock for *K. marxianus* that induced the most efficient DNA uptake while a significant number of cells survive. Cells were incubated at 43, 45, 47 and 49 °C for two different durations (5 and 15 min) to assess the effect of heat shock temperature and duration on transformation efficiency (Table 3.2). At the optimal temperature 47 °C, only 5 min were required to obtain the highest number of transformants, while with longer duration (i.e., 15 min), 30% less transformants remained. A slight 2-°C increase of temperature to 49 °C also caused a 30% decrease in the number of successful transformants. At temperatures lower than 47 °C (i.e., 45 and 43 °C), the number of transformants declined drastically to only 50% (at 45 °C) or even less than 25% (at 43 °C) of that obtained at 47 °C. Moreover, longer incubation could not compensate for the loss of transformation efficiency caused by lower temperatures, though small improvement was observed. These results indicated that shortening the time and elevating the temperature of heat shock enhances *K. marxianus* transformation.

Table 3.2. Heat shock and chemical induced transformation of *K. marxianus* CBS 6556 ura3Δ with various heat shock temperatures and durations (averaged by at least three biological replicates)

Temperature, Duration	Average transformants no. (cfu)	Average fold-enhancement
49 °C, 5 min	6.0×10^4	0.7
47 °C, 15 min	6.4×10^4	0.7
47 °C, 5 min	9.1×10^4	1.0
45 °C, 15 min	5.5×10^4	0.6
45 °C, 5 min	4.8×10^4	0.5
43 °C, 15 min	2.7×10^4	0.3
43 °C, 5 min	2.0×10^4	0.2

The number/concentration of cells and DNA molecules in the transformation mix also affect transformation rates (Table 3.3). As for *K. marxianus* CBS 6556, more DNA resulted in more transformants, though not exactly proportional. In contrast to *K. marxianus* DMKU3-1042 which achieved the highest transformation efficiency by using maximum number of cells (i.e., 7×10^9) [4], the number of transformants did not go further higher in *K. marxianus* CBS 6556 as increasing the number of cells from 7×10^8 to 9.8×10^8, with 1.04 pmol transformed DNA (equal to 4 µg of 6,275-bp plasmids). Notably, the transformation efficiency maintained at around 1.5×10^5 cfu/µg DNA even after supplying the 0.26 pmol DNA (equal to 1 µg of 6,275-bp plasmids) with an increasing number of cells from 3.5, 4.9, to 7×10^8. The number of cells was calculated by measuring optical density on Nanodrop 2000c UV–Vis spectrometer (Thermo Scientific) at 600 nm (OD_{600}), converting OD to cell density with the correlation function 1 OD/mL = 1.4×10^7 cells/mL (Figure S3.1), and then multiplying the cell density by the volume of used cell resuspension. The highest number of transformants, 4.5 x 10^5, was obtained when providing 7×10^8 of cells with 1.04 pmol (4 µg) plasmids.

Table 3.3. Heat shock and chemical induced transformation of *K. marxianus* CBS 6556 ura3Δ with various number of cells and transformed DNA (averaged by at least three biological replicates)

Cell no. ($\times 10^7$), DNA (pmol)	Average transformants no. (cfu)	Average fold-enhancement
35, 0.26	1.4×10^5	1.5
49, 0.13	9.1×10^4	1.0
49, 0.26	1.6×10^5	1.8
70, 0.13	9.7×10^4	1.1
70, 0.26	1.3×10^5	1.4
70, 0.52	2.8×10^5	3.1
70, 0.78	3.0×10^5	3.3
70, 1.04	4.5×10^5	4.9
98, 0.13	3.5×10^4	0.4
98, 1.04	1.8×10^5	2.0

The effects of PEG, LiAc and DTT concentration on *K. marxianus* transformation efficiency were also examined (Table 3.4). No transformant was obtained on plate under a 4,000-fold dilution as increasing the PEG concentration from 40 to 56%. When PEG concentration dropped to 30%, the transformation efficiency decreased by a magnitude. As the concentration of LiAc went up from 0.1 to 0.2 M, 50% more transformants were gained, while if further increased to 0.5 M, the transformation efficiency declined to only half of that by using 0.1 M LiAc. All mentioned above suggested that the optimum concentration of PEG is 40%, and 0.2 M for LiAc. The optimal DTT concentration, 30 nM, at which 3.5×10^5 transformants were obtained, was then identified by increasing DTT from 10 to 90 mM in 10-mM increments. The chemical concentration optimization was done using 0.26 pmol (1 μg) transformed DNA for each reaction.

Table 3.4. Heat shock and chemical induced transformation of K. marxianus CBS 6556 ura3Δ with various chemical concentrations (averaged by at least three biological replicates)

Pretreatment reagents	Avg. transformants (cfu)	Avg. fold-enhancement
30% PEG, 0.1M LiAc, 10mM DTT	1.2×10^4	0.1
40% PEG, 0.1M LiAc, 10mM DTT	1.3 × 105	1.0
56% PEG, 0.1M LiAc, 10mM DTT	No colony by 4000× dilution	N/A
40% PEG, 0.2M LiAc, 10mM DTT	2.0×10^5	1.5
40% PEG, 0.5M LiAc, 10mM DTT	6.5×10^4	0.5
40% PEG, 0.2M LiAc, 20mM DTT	1.7×10^5	1.3
40% PEG, 0.2M LiAc, 30mM DTT	3.5×10^5	2.7
40% PEG, 0.2M LiAc, 40mM DTT	3.3×10^5	2.5
40% PEG, 0.2M LiAc, 50mM DTT	2.7×10^5	2.1
40% PEG, 0.2M LiAc, 60mM DTT	2.3×10^5	1.8
40% PEG, 0.2M LiAc, 70mM DTT	2.8×10^5	2.2
40% PEG, 0.2M LiAc, 80mM DTT	2.7×10^5	2.1
40% PEG, 0.2M LiAc, 90mM DTT	2.5×10^5	1.9

Highly complex DNA libraries require pooling multiple transformations together. To minimize the transformation efficiency variation among different reactions, taking aliquots from one big culture is preferable to harvesting cells for each reaction individually from several smaller cultures. Therefore, a 25-ml cell culture system starting from an initial OD_{600} of 0.05 at 30 °C, 200 rpm in 250 mL baffled shake flask was chosen to identify the optimal growth time to harvest cells for transformation. The cell inoculum into a 25 mL medium was obtained from a single colony outgrowth in 1 ml liquid YPD medium at 30 °C, 200 rpm for 8 h. Five time points representing different cell growth phases were tested (Figure 3.1 and Table 3.5; 7 and 9 h for early exponential phase, 14 h for late exponential phase, 24 h for stationary phase, and 48 h for late stationary phase). All other transformation variables discussed before were set at their optimal values (harvested cells washed by 0.1 M LiAc and 1x TE buffer, 47 °C and 5 min for heat shock, 7×10^8 of cells, 1.04 pmol of DNA, 40% PEG, 0.2 M LiAc, 30 mM DTT).

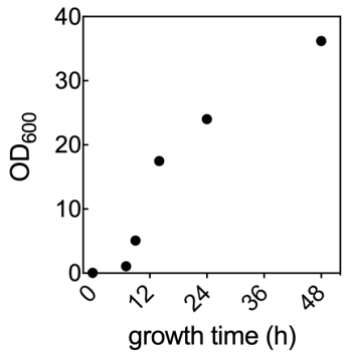

Figure 3.1. Growth of *K. marxianus* CBS 6556 *ura3Δ* with 25 mL YPD medium and an initial OD_{600} of 0.05 in 250 mL baffled shake flasks at 30 °C, 200 rpm.

The number of viable transformants greatly dropped by two orders of magnitude using *K. marxianus* cells harvested at early exponential phase (7 and 9 h of cultivation) compared with using cells reaching the late exponential phase (14 h), even though the same number of cells were used. As for cells growing further longer to stationary phase (24 h), the transformation efficiency decreased to half of the maximum. When coming to the late stationary phase (48 h), again only 1% of the maximum number of transformants remained. It was apparent that the growth phase when cells were harvested was critical for transformation efficiency, and the optimal timepoint to harvest cells was the late exponential phase for transformation in *K. marxianus*.

Table 3.5. Heat shock and chemical induced transformation of *K. marxianus* CBS 6556 ura3Δ with yeast cells harvested at various growth stages (averaged by at least three biological replicates)

Growth stage	Avg. transformants (cfu)	Avg. fold-enhancement
Single colony in 2mL medium, 16h	4.5×10^5	1.0
Inoculum in 25 mL medium, 7h	6.0×10^3	0.01
Inoculum in 25 mL medium, 9h	4.0×10^3	0.01
Inoculum in 25 mL medium, 14h	6.0×10^5	1.3
Inoculum in 25 mL medium, 24h	3.1×10^5	0.7
Inoculum in 25 mL medium, 48h	5.0×10^3	0.01

The optimized plasmid DNA transformation mediated by chemical reagents and heat shock achieved an over seven-times higher efficiency compared with the previous, for 6 x 10^5 transformants per reaction in average, using 1.04 pmol DNA (equal to 4 μg for 6,275-bp plasmids) and 5×10^8 of cells. Although this is already a feasible transformation method to meet the requirement of 100-fold coverage of genome-wide CRISPR libraries in *K. marxianus*, it still needs a great amount of DNA and at least 7 transformations pooled together, resulting in no less than 250 mL of medium to guarantee the cell outgrowth from an initial OD_{600}/mL of no more than 1.5, the starting OD_{600}/mL confirmed by experiments which does not hinder *K. marxianus* outgrowth in liquid cultures.

To further facilitate CRISPR library transformation for genome-wide screens in *K. marxianus*, electroporation, which has the potential to yield even more transformants compared with the heat shock and chemical induced method, was also elected for the transformation optimization.

Appropriate electroporation parameters, proper buffer conditions, care of cells from the very beginning of fresh single colony inoculation till the post-electroporation

cell recovery, and balanced quantities of cells with transformed DNA, all dependent on the target cells, thus, require strain-specific optimization.

As for applying exponential decay waves, the field strength ($= \frac{\text{Voltage (kV)}}{\text{Gap size (cm)}}$) is the most important parameter to provide a right electroporation condition which enables sufficient permeabilization with affordable cell membrane disturbance [6]. Our field strength optimization began with two default settings for *S. cerevisiae* on the Bio-Rad MicroPulser Electroporator ("Sc2": 1.5 kV, 0.2 cm; and "Sc4": 3.0 kV, 0.2 cm). Table 3.6 shows that *K. marxianus* accommodates higher field strength than *S. cerevisiae* to achieve its best electroporation performance when treating a similar number of cells.

Table 3.6. Electroporation transformation of *K. marxianus* CBS 6556 ura3Δ with different field strength (kV/cm; averaged by at least three biological replicates)

Voltage (kV), Gap size (cm)	Average transformants no. (cfu)	Average fold-enhancement
1.5 kV, 0.2 cm	2.5×10^5	0.4
3.0 kV, 0.2 cm	6×10^5	1.0
1.5 kV, 0.1 cm	6.8×10^5	1.1
1.8 kV, 0.1 cm	1.4×10^6	2.3
2.0 kV, 0.1 cm	8×10^5	1.3

Thorough cell washing with ice-cold ddH_2O and 1M sorbitol is widely used for yeasts. Previous studies also showed that pretreatment of cells by lithium acetate (LiAc) and dithiothreitol (DTT) greatly enhanced cell competence [7,8,9]. Hence, we washed *K.marxianus* cells using ice-cold ddH_2O and 1M sorbitol, pipetted gently, always kept cell pellets and resuspensions on ice, and adopted the optimal concentration of LiAc and DTT from the heat shock and chemical induced method for pretreatment to make electrocompetent cells. It is especially crucial to discard all culturing media and the

pretreatment buffer during cell washing even at a sacrifice of some cells, since the electroporation efficiency is very sensitive to salt concentration. The higher the salt concentration is, the lower the sample resistance is, the more significantly the voltage applied to the sample decreases from the setting. The pretreatment duration and temperature are also pivotal factors to increase cell competence. The best combination is 25 °C for 30 min with gentle shaking (100 rpm). Electroporation efficiency declined at higher temperatures or shorter pretreatment duration (data not shown). To maximize the pretreatment efficiency and reproducibility, supernatant was required to be completely discarded even with losing some cells before resuspending cells in the pretreatment buffer.

The optimum growth phase to harvest cells elected for the chemical and heat shock method was adopted to conduct electroporation. A 25 mL culture in 250 ml baffled shake flask starting from an initial OD_{600}/mL of 0.05 at 30 °C, 200 rpm to the end of exponential phase (~16 h to achieve an ultimate OD_{600}/mL of around 20 for *K. marxianus* CBS 6556 *ura3Δ*) was found to guarantee cell health for sufficient electrocompetency. As for post-electroporation cell recovery, unlike recommended ice-cold medium for *S. cerevisiae*, recovery medium pre-warmed at 37 °C enabled more transformant survivals for *K. marxianus* (Table 3.7).

Table 3.7. Electroporation transformation of *K. marxianus* CBS 6556 ura3Δ with different recovery conditions (averaged by at least three biological replicates)

Recovery conditions	Avg. transformants (cfu)	Avg. fold-enhancement
No recovery (3 kV, 0.2 cm)	1×10^5	0.2
37 °C pre-warmed (3 kV, 0.2 cm; 2 h)	6×10^5	1.0
37 °C pre-warmed (1.8 kV, 0.1 cm; 2 h)	1.4×10^6	2.3
Ice-cold (1.8 kV, 0.1 cm; 2 h)	5.8×10^5	1.0
37 °C pre-warmed (1.8 kV, 0.1 cm; 3 h)	1.5×10^6	2.5
37 °C pre-warmed (1.8 kV, 0.1 cm; 4 h)	1.7×10^6	2.8

Using a similar number of host cells, electroporation reaches an over 2-fold higher transformation efficiency than the chemical and heat shock method, with much less DNA (Table 3.5 and 3.8). This makes electroporation a more amenable alternative to the heat shock transformation of chemically competent cells for genome-wide screening experiments which require highly complex DNA libraries.

Table 3.8. Electroporation transformation of *K. marxianus* CBS 6556 ura3Δ with varied number of cells and transformed DNA (averaged by at least three biological replicates)

Cell no. ($\times 10^7$), DNA (pmol)	Average transformants no. (cfu)	Average fold-enhancement
42, 0.26	8.0×10^5	0.6
42, 0.13	3.9×10^5	0.3
56, 0.26	7.0×10^5	0.5
70, 0.26	1.4×10^6	1.0
70, 0.13	9.1×10^5	0.7
84, 0.26	1.5×10^6	1.1
105, 0.26	1.2×10^6	0.9

3.4 Materials and Methods

3.4.1 Molecular Cloning and Reagents

All molecular cloning reagents and enzymes were purchased from NEB. Plasmids created and used in this work are listed in Table S3.1. The library backbone pIW1213 was constructed by linearizing plasmid pIW601 (Addgene ID: 98907) with primers

ML_249 and ML_240 using Q5® High-Fidelity DNA polymerase and assembling a 51-bp insert which contains 5-bp random nucleotides flanked by 23-bp upstream and downstream homology into the linearized pIW601 by NEBuilder® HiFi DNA Assembly Master Mix (Gibson assembly) to get rid of the constitutive Cas9 expression cassette. All primers (Table S3.2) were purchased from IDT™. All molecular cloning was accomplished in TOP 10 competent *E. coli* (Thermo Fisher Scientific).

Other reagents used in this work were purchased from Fisher Scientific, Sigma-Aldrich or as otherwise noted: BD Difco™ Yeast Nitrogen Base without Amino Acids (Fisher Scientific), CSM-Ura powder (Sunrise Science Products), D-glucose (Fisher Scientific), Gibco™ Bacto™ Yeast Extract (Fisher Scientific), Gibco™ Bacto™ Peptone (Fisher Scientific), Miller's LB powder (Sigma-Aldrich), Ampicillin sodium salt (Fisher Scientific), Agar (Fisher Scientific), Pfaltz & Bauer Polyethylene glycol 3350 (Fisher Scientific), Lithium acetate dihydrate (Sigma-Aldrich), 10× TE buffer (pH 7.4; Fisher Scientific), Dithiothreitol (Fisher Scientific), R&D Systems™ Salmon Sperm DNA (Fisher Scientific), and Sorbitol (Fisher Scientific).

3.4.2 *K. marxianus* Strains, Media, and Cultivation

Kluyveromyces marxianus CBS 6556 *ura3Δ* was used for all transformation experiments described in this chapter. All liquid yeast cultures were conducted in an INFORS HT Multitron incubation shaker with the speed and temperature set at 200 rpm and 30 °C, respectively. To minimize the transformation efficiency variation among individual reactions for highly complex DNA libraries that require pooling multiple transformations, a 25 ml cell culture system starting from an initial OD_{600} of 0.05 in 250

mL baffled shake flask was elected to achieve the appropriate growth phase for competent cell pretreatment. The cell inoculum into the 25 mL medium was obtained from a single colony on YPD agar plate growing in 1 ml liquid YPD medium for 8 h (YPD: 10 g/L Gibco™ Bacto™ Yeast Extract, 20 g/L Gibco™ Bacto™ Peptone, 20 g/L D-glucose, adding 23 g/L agar as needed). As for heat shock and chemical induced transformations, cells from heat shock were resuspended in 1 mL synthetic defined (SD) minimal glucose media without uracil (SD-U: 6.7 g/L BD Difco™ Yeast Nitrogen Base without amino acids, 0.77 g/L CSM-Ura, and 20 g/L D-glucose), and then plated on SD-U agar plates with 1,000- to 4,000-fold dilution for viable transformant counting. For each electroporation reaction, after applying the voltage pulse, 1 mL of YPD containing 1 M sorbitol pre-warmed at 37 °C was immediately added into the cuvette, and then the cell resuspension was transferred into 14 mL Corning™ Falcon™ culturing tube for a 2-h recovery at 30 °C, 200 rpm.

3.4.3. Correlation Between *K. marxianus* Cell Density and OD_{600}/mL

The number of cells were quantified by flow cytometry. Overnight 1 mL of *K. marxianus* cultures in YPD were washed by equal volume of 1× PBS, and then diluted appropriately into 1 mL cell suspension (10-mm pathlength) for OD_{600} measurement on Nanodrop 2000c UV–Vis spectrometer (linear measurement range of OD_{600}/mL from 0.1 to 0.7; Thermo Scientific). This cell suspension was further diluted by 2, 4, 8, and 16 folds (dilution ratio and the number of dilutions welcomed to be altered accordingly) for samples to be counted on the BD Accuri™ C6 Plus flow cytometer at a rate of no more than 2,500 events per second with the medium fluidics rate of 35 µL/min. The counting

events was set as 10,000. The corresponding volume of each cell suspension was used to calculate the cell density, and then to plot the cell density versus each corresponding OD_{600} (Figure S3.1).

3.4.4 *K. marxianus* transformation

The optimized heat shock and chemical induced method was accomplished based on our previous protocol [1]. Briefly, the desired starting *K. marxianus* strain was cultivated in 25 mL YPD medium per flask to reach the end of exponential phase (OD_{600}/mL about 20). Cell culture was then harvested by centrifugation at 3,000 g for 5 min in a sterile 50 mL centrifuge tube. After washing twice with 25 mL of the sterile 1× TE and 0.1 M LiAc buffer, cells were resuspended in the appropriate volume of the same buffer for a cell resuspension of OD_{600}/mL = 50 (i.e., 7×10^8 of *K. marxianus* cells per mL), and then 1 mL of the cell resuspension was harvested by centrifugation at 5,000 g for 1 min in a sterile 1.5 mL Eppendorf tube, mixed with 10 µL of 10 mg/mL carrier DNA (R&D Systems™ salmon sperm DNA) by gently vortexing. 1.04 pmol transformed DNA (equal to 4 µg of 6,275-bp plasmids) were added and then 500 µL of the same-day prepared room-temperature transformation buffer (40% w/w autoclave sterilized polyethylene glycol 3350 (PEG 3350), 0.2-µm filtered 0.2 M LiAc, 1× TE (pH 7.4), and 30 mM DTT). All components in the tube were mixed thoroughly by pipetting and were incubated at room temperature for 15 min. Subsequently, the mixture was heat shocked at 47 °C for 5 min using a solid heat block containing diH_2O in each well. Transformants were then directly harvested by centrifugation at 5,000 g for 30 s. Supernatant was removed and cell pellets were resuspended in 1 mL SD-U media. The resuspension was

diluted accordingly, spread on SD-U plates, and then incubated at 37 °C until they were visible for counting (usually after 24 h for *K. marxianus* CBS 6556 *ura3Δ*).

The electroporation protocol was built upon the current available methods for *K. marxianus* and *S. cerevisiae*. In brief, after the 25 mL YPD cultures reaching to the desired OD_{600}/mL about 20, every flask was immediately immersed in ice bath to the level of the cell culture, with additional ice added to the ice bath for the flask fully packed in ice up to its neck, but in the meanwhile, water from melting ice would not wick up under the mouth of the flask. The culture chilling lasted for at least 1 h to stop growth, with flask swirled every 30 min for even cooling. The fully chilled cell cultures were then poured into 50 mL prechilled-on-ice centrifuge tube, and cells were harvested by centrifugation at 3,000 g for 5 min in a prechilled-to-4 °C centrifuge. Cell pellets were washed once with 25 mL ice-cold ddH_2O and once with equal volume of ice-cold 1 M sterile sorbitol. Then the cell pellets were resuspended in 10 mL same-day freshly prepared room-temperature pretreatment buffer (0.2-μm filtered 0.2 M LiAc, 1× TE (pH 7.4), and 30 mM DTT). The cell resuspension was gently shaken at 100 rpm, 25 °C for 30 min. The tube containing cell resuspension in the pretreatment buffer was then immersed in ice bath for 20 min with tube inverted for three time after the first 10 min for even chilling. The fully chilled cell resuspension was centrifuged at 4 °C, washed twice by 25 mL ice-cold water, and then twice by 25 mL ice-cold 1 M sorbitol. In the last round of cell wash in 1 M sorbitol, the OD_{600}/mL was measured to calculate the volume of the final resuspension to obtain a cell density of 8.4×10^8 per 80 μL. Half of the calculated volume was then used for adding ice-cold 1 M sorbitol to obtain the electrocompetent cell

suspension because cell pellets themselves also account for volumes, additional ice-cold 1 M sorbitol could be needed to achieve the calculated cell suspension volume. The 50 mL centrifuge tube containing cell pellets or resuspension was always kept immersed in the ice bath. 1.5 mL pre-chilled-in-minus 20 °C sterile Eppendorf tubes (whose number dependent on the desired number of transformations) were put on ice, and 0.26 pmol DNA (equal to 4 µg of 6,275-bp plasmids) were added into each tube. 80 µL of the electrocompetent cell suspension were mixed with the DNA in each Eppendorf tube, and incubated on ice for 10 min. During the 10-min waiting time, same number of pre-chilled-in-minus 20 °C 0.1-cm-gap electroporation cuvettes were put on ice. Cell-DNA mixtures were transferred into cuvettes and pulsed at 1.8kV with the Bio-Rad MicroPulser Electroporator (the shocking chamber slide kept in 4 °C fridge until needed for performing electroporation). Immediately, 1 mL pre-warmed-to-37 °C YPD with 1 M sorbitol was added and all contents in the cuvette were gently mixed right in the cuvette. In our case, the time constant (TC) of 3.5 to 4.5 ms indicated an electroporation success obtaining an average of 1.5×10^6 transformants. Now all cuvettes can be set at room temperature, and mixtures from each cuvette were transferred into a 14 ml culture tube by 1 mL sterile syringe with needle, respectively. Transformants in YPD with 1 M sorbitol were recovered at 30 °C, 200 rpm for 2 h. After recovery, cell cultures were diluted by 10,000 folds, spread on SD-U plates, and incubated at 37 °C until they were visible for counting (usually after 24 h for *K. marxianus* CBS 6556 *ura3Δ*).

3.5 Supporting Information

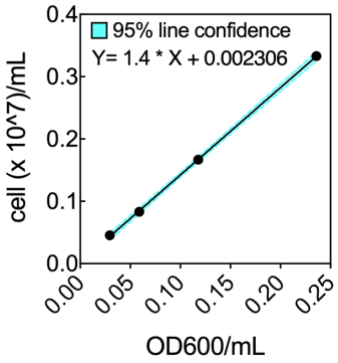

Figure S3.1. Cell density versus OD_{600} of *K. marxianus* CBS 6556.

Table S3.1. Plasmids used in this study

Plasmids	Description	Reference (Addgene#)
pIW601	P_{ScTEF1}-KmCas9-SV40-ScCYC1t, P_{KmPRP1}-$tRNA^{Gly}$-PspXI recognition site-SUP4	Löbs et al. [10] (98907)
pIW1213	Removing P_{ScTEF1}-KmCas9-SV40-ScCYC1t in pIW601	This chapter

Table S3.2. Primers used for cloning in this study

Primers	Sequence (5' to 3', 20-nt spacers or priming parts are underlined)	Use
ML_249	CCTAGGGCGTATACTCCAACTTG	pIW601_linearize_f
ML_240	GAGCTCCAATTCGCCCTATAGTG	pIW601_linearize_r
ML_251	CACTATAGGGCGAATTGGAGCTCCACTTCCTA GGGCGTATACTCCAACTTG	pIW601_reassemble_f
ML_242	CAAGTTGGAGTATACGCCCTAGGAAGTGGAGC TCCAATTCGCCCTATAGTG	pIW601_reassemble_r

Table S3.3. Strains used in this study

Strains	Description	Reference
E. coli TOP 10	F- mcrA Δ (mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ (araleu)7697 galU galK rpsL (StrR) endA1 nupG	Thermo Fisher Scientific
YS402	*K. marxianus* CBS 6556 ura3Δ	Löbs et al. [10]

3.6 Reference

[1] Li, M., et al. CRISPR-mediated multigene integration enables Shikimate pathway refactoring for enhanced 2-phenylethanol biosynthesis in *Kluyveromyces marxianus*. Biotechnol Biofuels. 2021;14(1): 3.

[2] Aune TEV and Aachmann FL. Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. Appl Microbiol Biotechnol. 2010;85(5):1301-1313.

[3] Gietz RD and Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2007;2(1):31-34.

[4] Abdel-Banat BMA, et al. Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. Yeast. 2010;27(1):29-39.

[5] Yamane T, et al. Dissection of centromeric DNA from yeast *Yarrowia lipolytica* and identification of protein-binding site required for plasmid transmission. J Biosci Bioeng. 2008;105(6):571-578.

[6] Jordan ET, et al. Optimizing electroporation conditions in primary and other difficult-to-transfect cells. J Biomol Tech. 2008;19(5):328-334.

[7] Meilhoc E, et al. High Efficiency Transformation of Intact Yeast Cells by Electric Field Pulses. Nat. Biotechnol. 1990;8(3):223-227.

[8] Iborra F. High efficiency transformation of *Kluyveromyces marxianus* by a replicative plasmid. Curr Genet. 1993:24(1):181-183.

[9] Thompson JR, et al. An improved protocol for the preparation of yeast cells for transformation by electroporation. Yeast. 1998;14(6):565-571.

[10] Löbs AK, et al. CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol Biofuels. 2017;10(1):164.

Chapter 4: Differentially CRISPR/Cas9 Mediated Whole Genome Editing Efficiencies in Glucose, Lactose, and Xylose Elucidate Glucose Repression in *K. marxianus*

4.1 Abstract

The flexible, efficient, and multiplexed functional type II CRISPR system revolutionizes genomic screening. However, because of the high sequence specificity as well as appropriate epigenetic environment particularly required for functional CRISPR/Cas9 editing, it is necessary to design and validate an exclusive genome-wide library for every new given species of interest. We characterized a four-fold single-guide RNA (sgRNA) library targeting both promoter sequences and coding regions in *K. marxianus* CBS 6556. As assisted by culturing in three carbon sources, glucose, lactose, and xylose, the differentially induced editing efficiencies with the same sgRNA can be used to identify gene regulations involved in glucose repression.

4.2 Introduction

A great challenge of widely applying the CRISPR/Cas9 system for genomic studies in non-model species is brought by its high sequence specificity. Unlike transformation methods and expression cassettes which can be easily adopted from a close species, genetic diversity necessitates genome-wide CRISPR libraries exclusively designed for each given species of interest. Efficient CRISPR-mediated functional genome deletions are heavily dependent on the high cutting activity of sgRNA targeting each CDS. However unfortunately, even the knock-out rate of the same gene can be

enormously varied because of choosing different sgRNA sequences. Therefore, validation of a redundantly designed library is a must to guarantee sufficient gene coverage by efficient CRISPR-mediated mutation for reliable whole genome associated screens. Our previous CRISPR screening works in another non-conventional yeast *Yarrowia lipolytica* showed that around half the number of sgRNAs in an unbiasedly designed library were of acceptable gene editing success that enabled genome-wide functional studies [1,2]. To elucidate the landscape of sgRNA activities along the *K. marxianus* functional genome, we analyzed the cutting performance of an unbiased sgRNA library designed without using any on-target cutting efficiency prediction algorithm in this chapter.

Beside the 20-nt spacer sequence, an accessible genomic region is also a necessity for efficient editing by CRISPR/Cas9. Genomic accessibility perturbation results from transcriptional regulation. Therefore, the cutting activity variation of the same guide sequence can help identify differentially expressed genes (DEGs) and narrow down the repressor binding region. In this chapter, we discussed one of the most universal environmental signaling, glucose repression, which accompanies epigenetic effects on sgRNA cutting efficiency.

4.3 Results and discussion

4,588 genes were annotated in *K. marxianus* CBS 6556 by aligning the 4,952 annotated CDSs of *K. marxianus* DMKU3-1042 to the CBS 6566 genome. Four guides were randomly selected for each gene both at the coding region and the upstream

promoter of the gene, respectively. crRNA secondary structure and sequence uniqueness were then used as the only exclusion parameters to filter out unqualified sgRNA candidates. Any 20-nt crRNA sequence that forms no less than four homopolymers, has a predicted melting temperature greater than 30 °C, or is redundant along the genome within the 14 nucleotides closest to the protospacer adjacent motif (PAM), will be removed. Finally, a library of 33,243 qualified crRNA sequences (7.3 per gene on average) without selection favoritism in their on-targeting cutting rates, together with 651 non-targeting guides (2% of the library) as the negative control was synthesized and subsequently cloned into the vector with sgRNA expression driven by a synthetic hybrid RNA polymerase III (Pol III) promoter (Table S4.1). Sequencing by MiSeq 1×75 revealed that the original library design was well represented in the final plasmid library (Figure S4.1).

The library was transformed into *K. marxianus* CBS 6556 *ura3Δ* by 100-fold coverage at the minimum. (Figure 4.1A) Successful transformants underwent outgrowth over three subculture cycles, where cultures reached cell confluence and got refreshed every 24 h. The gene *KU70* was functionally knocked out in both control (cntrl) and experimental (expmt) strains. A strain with malfunctional *KU70* is no longer able to repair the CRISPR/Cas9-induced double-stranded break (DSB) by non-homologous end joint (NHEJ) pathway. Therefore, the more efficiently an sgRNA generates a genomic DNA DSB, the more likely it gets depleted in the library harvested from the new cell population at confluence, since cells expressing this sgRNA were not viable anymore because of the unrepaired DNA damage. To quantify the on-targeting cutting activity of

each sgRNA, Cutting Score (CS) was used, defined as the log2 value of the ratio of normalized relative abundance of an sgRNA in the cntrl strain over the expmt strain. The higher the CS value of a library sgRNA than the average of the non-targeting negative controls, the more active the sgRNA was. The farther the CS value of both library and non-targeting negative controls away from 0, the more likely the library was biased to active sgRNAs.

The evolution of the genome-wide library over three-day subculture cycles at two temperatures (30 and 37 °C) and with three different carbon sources (2% glucose, lactose, and xylose) were shown in Figure 4.1B. All crRNAs in the library design were detected in the cntrl group from every culturing condition by exactly matching the trimmed sequencing results to the designed crRNA sequences. After two growth cycles (day 2), significant changes in CS distribution of the 33,243 library sgRNAs were observed, and this new distribution remained in the following 3rd subculture. Active guides able to induce DSBs after two-day subculture were much less represented in the expmnt strain compared with day 1. However overall, CS values of both day 2 and day 3 indicated that most of the guides could not efficiently induce DSBs. It is apparent from the violin plot (Figure 4.1B) that the CS median of the targeting library was not significantly higher than that of the non-targeting control, and their CS medians were both close to 0. The non-targeting control population behaved as expected, whose CS distribution violin shape maintained during the library evolution.

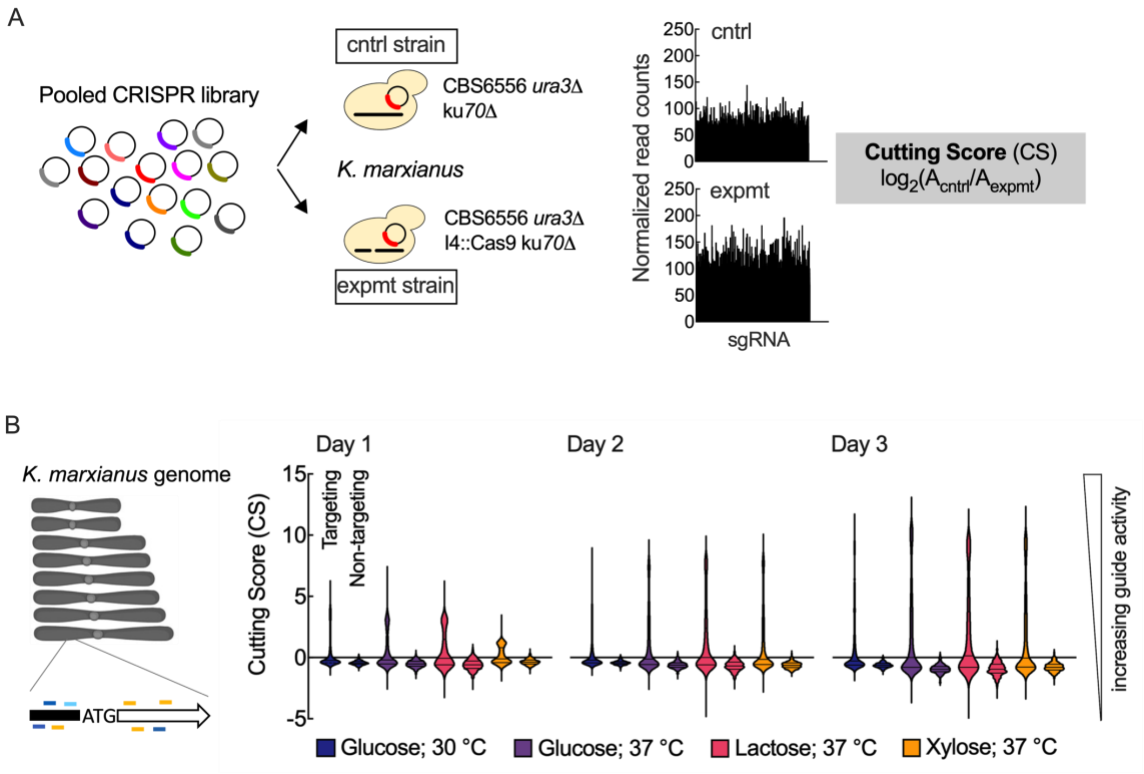

Figure 4.1. CRISPR-Cas9 enabled genome-wide screening and sgRNA library validation in *K. marxianus*. (A) Schematic representation of genome-wide screening workflow for library activity validation. (B) Genome coverage and the overall library activity under two temperatures (30 and 37 °C) and three carbon sources (2% glucose, lactose, and xylose in defined minimal media) after one, two, and three days of growth to cell confluence. Transformations were done at least in biological duplicates with over 100-fold coverage of the library in each replicate. CS values shown are calculated by combining raw reads from at least two replicates.

Based on the observation that the cutting fitness distribution of the library did not significantly change since the 2nd growth cycle, we selected CSs of day 2 for all subsequent studies according to different culturing temperatures and carbon sources.

Figure 4.2A again showed that a vast number of sgRNAs of the library were statistically the same as the non-targeting negative controls in respect of CSs. However, histograms also indicated that at higher temperature (i.e., 37 °C) in particular, there occurred a small but significant cluster of CS values at the other end of the CS spectrum,

whose median CS in average was about 10 greater than that of the other major cluster overlapping with non-targeting controls. Therefore, we elected the number of clusters as two in the Fuzzy C-means (FCM) clustering analysis. Membership grades were assigned to each CS value for being out of the lower cluster. Corresponding CS cutoffs for the memberships of 75%, 50%, 25%, and 10% under each culturing condition were summarized in Table 4.1.

Table 4.1. CS cutoffs based on membership grades indicating the degree to which library crRNAs belong to the cluster distinct from the non-targeting one in four culturing conditions

Membership grades	30 °C, glucose	37 °C, glucose	37 °C, lactose	37 °C, xylose
75%	3.772	4.466	4.840	4.702
50%	2.936	3.485	3.787	3.675
25%	2.102	2.505	2.736	2.649
10%	1.379	1.657	1.824	1.760
Inactive (*)	0.8933	1.491	1.917	1.433

(*) The CS cutoff to differentiate active sgRNAs from inactive was the minimum CS value at which the CS of a library crRNA was significantly out from the CS population of 651 non-targeting 20-nt sequences.

As for all 33,243 library sgRNAs, 16,186 of them targeted gene coding sequences (~3.5-fold coverage), and 17,057 at promoter regions (~3.7-fold coverage). For sgRNAs whose CS values were statistically out of the non-targeting population in t-test, they were defined as “active”; otherwise, they were “inactive”. The CS cutoff identified by multiple t-tests to separate the active from inactive was 0.8933 and 1.491 for 2% glucose at 30 and 37 °C, 1.917 and 1.433 for 2% lactose and xylose at 37 °C, respectively. Figure 4.2B (left) showed that approximately a quarter of the sgRNAs targeting CDSs were active, while more than a third designed for promoters were active. The consistency of the number of active and inactive guides characterized under four different culturing

conditions indicated that guides' on-target cutting efficiencies were maintained in general. Promoter regions had more active guides identified than CDS in this unbiasedly designed library, implying that the gene regulation regions of the *K. marxianus* CBS 6556 genome were overall more accessible to DNA binding than the gene coding sequences.

Within the active sgRNA spectrum (Figure 4.2B (right)), a membership grade of 75% to the higher end cluster whose mean CS value was 6.053 and 7.143 for 2% glucose at 30 and 37 °C, 7.714 and 7.504 for 2% lactose and xylose at 37 °C, respectively, was set as the lower limit of sgRNAs considered to be high active. The low active sgRNAs were those whose CSs were categorized below 25% memberships. The intermediate category consisting of memberships greater than 25% but lower than 75% was defined as medium active. At least half of the active guides were also considered to be high active. Taking the active sgRNA subset share of the full library into account, about one eighth of the unbiasedly designed sgRNAs targeting CDS were of high cutting activity, and for promoter, it is one fifth.

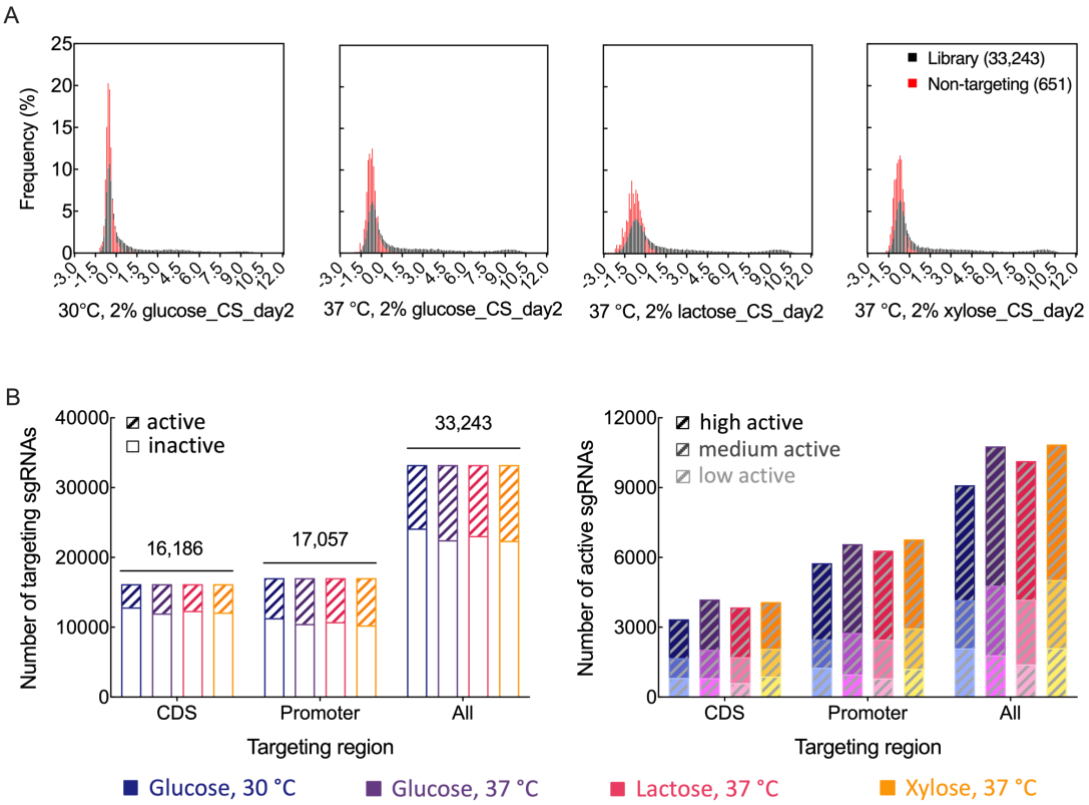

Figure 4.2. Distribution of CRISPR-Cas9 cutting efficiency with varied cultivation temperatures and carbon sources. (A) Histogram of the full library and the non-targeting sgRNA as negative control. (B) Library coverage at the coding sequence (CDS) and the upstream (promoter) of a gene by sgRNA activity classification: active (CS out of the non-targeting population with statistical significance; box with pattern fill), and inactive (CS belonging to the non-targeting population with statistical significance; box without fill). The active subpopulation consisted of high (membership > 75% to the cluster farthest away from the non-targeting group by Fuzzy C-means analysis; dark color), medium (25% < membership < 75%; medium color), and low active (membership < 25%; light color). CS values were determined after two growth cycles (2 days) from at least duplicate biological samples.

Histograms of CS distribution at varied cultivation temperatures (Figure 4.2A) showed that the overall library was more active at 37 °C, since a more significant aggregation of sgRNAs appeared at the higher value end of CS spectrum even when there was a universal CS spreading out indicated by that the frequency peaks of the non-

targeting controls at 37 °C were only half in the height of the same subpopulation at 30 °C.

Venn diagrams (Figure 4.3A and C) of high active and inactive sgRNAs for 2% glucose at 30 and 37 °C further confirmed that the higher temperature enhanced the sgRNA on-target editing efficiency in general. 4,940 sgRNAs of the library were identified as high active for glucose at 30 °C, and this number increased to 5,987 at 37 °C (Figure 4.3A). 85% of the high active guides identified in glucose at 30 °C were also recognized in the high active subset at 37 °C, while only 70% was found vice versa. Contrary to the high active subgroup, less inactive sgRNAs were identified at 37 °C compared with 30 °C (Figure 4.3C).

The CS distribution of the full library was generally consistent among different carbon sources (Figure 4.2). For both subsets of high active and inactive sgRNAs, the majority (>75%) identified for one carbon source also agreed with the other two (Figure 4.3B and D), suggesting that the 20-nt guide sequence is the most critical factor for cutting activity because it determined the functional expression and stability of the RNA molecule, as well as the targeting genome region to some extent. Therefore, the outliers from the inactive sgRNA intersection can probe putative epigenetic effects caused by carbon source repression and derepression.

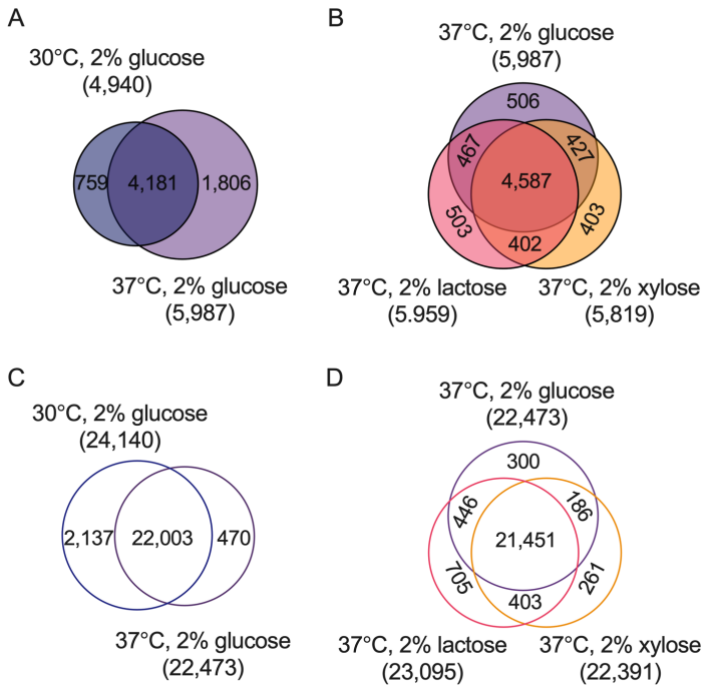

Figure 4.3. Temperature and carbon source effects on apparent cutting efficiency. (A) Venn diagram of high active sgRNAs for glucose and their overlap at different temperatures. (B) Venn diagram of high active sgRNAs identified at 37 °C and their overlap according to different carbon sources as the sole culturing variable. (C) Venn diagram of inactive sgRNAs for glucose and their overlap at different temperatures. (D) Venn diagram of inactive sgRNAs identified at 37 °C and their overlap according to different carbon sources as the sole culturing variable.

Our previous study showed that *K. marxianus* CBS 6556 preferred utilizing glucose to lactose or xylose [3]. To study the glucose repression and derepression implied by varied CS of the same sgRNA only because of feeding with different sugars, crRNAs whose CS values were less than 10% of likelihood out of the non-targeting CS population (below the membership threshold of 10% in Table 4.1) were considered having very low cutting efficiency, approaching inactive. 22,804 were found in this subcategory in 2% glucose at 37 °C, and 22,542 out of them also in agreement with that at 30 °C. The 22,542 in consensus were elected for subsequent outlier analysis to increase the

confidence to call a crRNA as low active as to approach inactive in glucose. The crRNA outliers were defined as unions as well as intersections whose CSs categorized as above 25 or 50% membership in 2% lactose or/and 2% xylose. Their targeting CDSs were identified as conditionally repressed in glucose. Figure 4.4A showed that 318 (targeting 276 CDSs in total) of the 22,542 extremely low active crRNAs in glucose were also found in the group of at least lower medium active in lactose or xylose (>25% in union). This number decreased to 72 (targeting 59 CDSs) and 124 (targeting 109 CDSs) when the outlier selection cutoff was more rigidly enforced as better than lower medium active in both two other sugars (>25 in intersec) and at least upper medium active in no less than one other sugar (>50% in union), respectively. The harshest outlier cutoff was at least as upper medium active in both lactose and xylose (>50% in intersec). Only 27 qualified crRNAs (targeting 20 CDSs) were found.

KEGG pathway analysis was used to functionally validate the glucose repressed genes identified by sgRNA deactivation in glucose. Glycolysis/Gluconeogenesis (KEGG ID: kmx00010; p = 0.028), Metabolic pathways (KEGG ID: kmx01100; p = 0.028), and Biosynthesis of secondary metabolites (KEGG ID: kmx01110; p = 0.033) were enriched based on the GO terms included in the category of ">25% in union". It was expected that gluconeogenesis was repressed in the presence of glucose. Galactose metabolism (KEGG ID: kmx00052; p = 0.014) and beta-Alanine metabolism (KEGG ID: kmx00410; p = 0.014) were identified to be repressed in the category of ">25% in intersec". When the preferrable glucose was present in the medium, it was reasonable that the utilization of other carbon sources (e.g., galactose) was hindered. The de novo biosynthesis of beta-

alanine was three enzymatic reactions from pyruvate, one of the critical intermediates from glycolysis [4]. There was a trade-off between the pyruvate fed into TCA cycle to generate reducing equivalents required by respiration and that converted to beta-alanine. It was not surprising that during the fast aerobic growth especially when adding L-alanine as nutrient supplement in the medium, carbon flux from pyruvate was in prior into TCA cycle. Galactose metabolism (KEGG ID: kmx00052; p = 0.004) and Glycolysis/Gluconeogenesis (KEGG ID: kmx00010; p = 0.028) were again enriched as repressed pathways for the category ">50% in union". Glycolysis/Gluconeogenesis (KEGG ID: kmx00010; p = 0.010), Galactose metabolism (KEGG ID: kmx00052; p = 0.011), Glycerolipid metabolism (KEGG ID: kmx00561; p = 0.030), and Other glycan degradation (KEGG ID: 00511; p = 0.030) were repressed in glucose as shown in the smallest category, ">50% in intersec". Glycerolipid was involved in membrane protein anchoring and molecule signaling [5]. The higher utilization efficiency of glucose compared with lactose and xylose could explain that less structural glycerolipid was needed by cells when fed with glucose. Polymeric glycans form the skeleton of the yeast cell wall. Less net degradation rate of cell walls referred to a less stressful overall cell growth in the culture, which was consistent with the fact of growing with preferrable glucose.

MIG1 encodes one of the most important and universal regulation factors of glucose repression in yeasts. A transcriptomic study in *K. marxianus* DMKU3-1042 identified 689 CDSs down regulated by *MIG1* expression [6]. Figure 4.4B showed that a considerable number of putative glucose repressed CDSs identified by differentially

CRISPR/Cas9 mediated cutting efficiencies of the same targeting crRNA in glucose and other carbon sources agree with those negatively regulated by the Mig1 transcription factor. With the 25% cutoff to differentiate the glucose repression from derepression, around 24% of identified CDSs in the ">25% in union" by CS screens in three carbon sources were included in the differentially expressed CDSs with and without down-regulation of Mig1 repressor. The consensus percentage rose to 41% for the intersection category. By increasing the cutoff to 50%, the percentage of Mig1p-regulated CDSs identified in the union went up from 24% to 34%, and for the intersection subset, it increased to 60%. The union category where calls for glucose effects were of lower confidence compared with the intersection category since it was possible that a crRNA also had comparably low activity in lactose or xylose with that in glucose. The lower cutoff of CS in non-glucose sugars was used to differentiate sgRNA activities among different carbon sources, less confident to call glucose derepression.

Glucose repression is a combined result of redundant regulatory paths [7]. *PCK1*, encoding a critical enzyme for gluconeogenesis, was not recognized to be downregulated by *MIG1* expression [6]. However, the variation of CSs of the same guide targeting *PCK1* promoter due to different carbon sources together with the congregation of CSs for crRNAs at the CDS, indicated that *PCK1* was not only a transcriptionally regulated gene in response to glucose, but also a redundantly regulated gene whose promoter contains other DNA motifs to bind repressor recruiters beside Mig1. A Mig1 binding motif variant GTGGGG(A/G/T)(A/T)G [7] was partially found (GGATG) as the PAM and the first three bases closest to the PAM of the promoter sequence targeting by crRNA "PCK1_-

393:-395_R". Another uncharacterized motif GGGAGG which was found significantly overrepresented among genes redundantly repressed by Mig1 and Mig2 in literature [7] was the PAM sequence and the three PAM-proximal bases of the promoter region targeting by crRNA "PCK1_-243:-245_R". Because of the NGG sequence requirement as well as the crucial crRNA-DNA binding within the seed region (PAM-proximal 10 to 12 bp) for functional Cas9 editing, the differentially CRISPR/Cas9 enabled gene editing efficiencies can be used as a screening tool to elucidate transcriptional regulation mediated by zinc finger proteins [8].

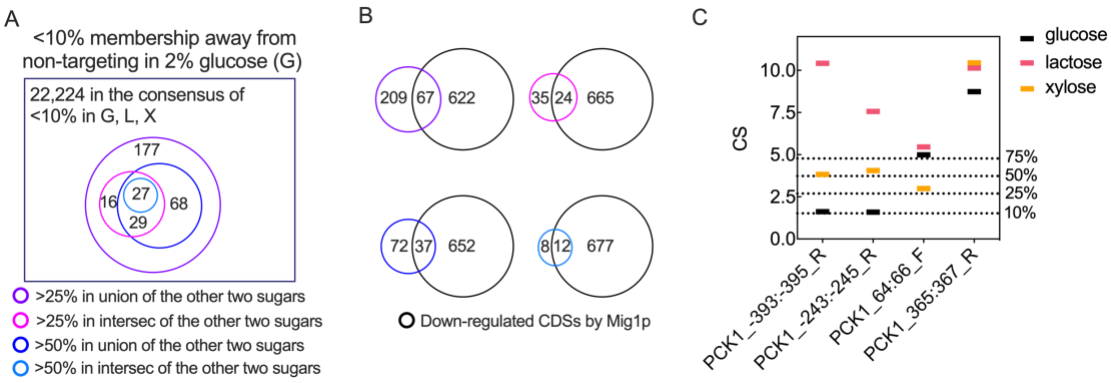

Figure 4.4. Glucose repression and de-repression indicated by differentially CRISPR/Cas9 mediated gene editing rates in the presence and absence of glucose. (A) crRNAs with significantly lower cutting efficiency in glucose (G) compared with lactose (L) and/or xylose (L). (B) Comparison of glucose repressed CDSs identified by differentially induced DNA cutting rates with and without glucose and those identified by differential gene expressions with and without functional Mig1 protein. (C) A glucose repressed gene, PCK1, which was not recognized from transcriptomics analysis by only looking into the Mig1 repressor, however identified by differentially CRISPR/Cas9 enabled gene editing efficiencies fed with and without glucose.

4.4 Materials and Methods

4.4.1 sgRNA Library Cloning and Reagents

All molecular cloning reagents and enzymes were purchased from NEB. Plasmids created and used in this work are listed in Table S4.1. The transformation quality control plasmid pIW1250 was made by assembling PspXI-digested pIW1213 and a 60-bp insert containing 20-bp upstream and downstream homology as well as the 20-bp NonTargeting_1 sequence. pIW1213 together with a list of 33,894 20-nt crRNA sequences were sent to JGI for library construction. All primers (Table S4.2) were purchased from IDT™. All molecular cloning was accomplished in TOP 10 competent *E. coli* (Thermo Fisher Scientific).

Other reagents used in this work were purchased from Fisher Scientific, Sigma-Aldrich or as otherwise noted: BD Difco™ Yeast Nitrogen Base without Amino Acids (Fisher Scientific), CSM-Ura powder (Sunrise Science Products), D-glucose (Fisher Scientific), D-Lactose monohydrate (Fisher Scientific), D(+)-Xylose (Thermo Scientific), Gibco™ Bacto™ Yeast Extract (Fisher Scientific), Gibco™ Bacto™ Peptone (Fisher Scientific), 5-FOA (Fisher Scientific), Miller's LB powder (Sigma-Aldrich), Ampicillin sodium salt (Fisher Scientific), Agar (Fisher Scientific), DNase and RNase free, 0.2-micron filtered water (Fisher Scientific), Lithium acetate dihydrate (Sigma-Aldrich), 10× TE buffer (pH 7.4; Fisher Scientific), Dithiothreitol (Fisher Scientific), and Sorbitol (Fisher Scientific).

4.4.2 *K. marxianus* Strains, Media, and Cultivation

All constructed strains are listed in Table S4.3. The NHEJ-deficient *K marxianus* strain CBS 6556 *ura3Δ ku70Δ* was constructed by KU70-URAblaster integration into the *KU70* locus to knock it out, and then *URA3* marker excision in YPD with 1 g/L 5-FOA [9]. The experimental strain was subsequently obtained by integrating the constitutive Cas 9 expression cassette cloned from plasmid pIW601 into a non-coding site I4 [10], located near to a cluster of actively transcribed non-essential genes via CRISPR-mediated homologous recombination.

All liquid yeast cultures were conducted in an INFORS HT Multitron incubation shaker with the speed of 200 rpm and temperatures set as desired (30 or 37 °C). To minimize the transformation efficiency variation among multiple reactions required to achieve 100-fold sgRNA library coverage, a 25 mL of cell culture system starting from an initial OD_{600} of 0.05 in 250 mL baffled shake flask was elected to harvest host cells for transformation. The cell inoculum into the 25 mL medium was obtained from a single colony on YPD agar plate growing in 1 mL liquid YPD medium for 8 h (YPD: 10 g/L Gibco™ Bacto™ Yeast Extract, 20 g/L Gibco™ Bacto™ Peptone, 20 g/L D-glucose, adding 23 g/L agar for plates). Washed transformants were pooled in 500 mL synthetic defined (SD) minimal glucose media without uracil (SD-U: 6.7 g/L BD Difco™ Yeast Nitrogen Base without amino acids, 0.77 g/L CSM-Ura, and 20 g/L D-glucose or D-lactose or D-xylose as desired) for the first subculture (day 1; 2-L baffled flask), and then into 50 mL SD-U for the 2nd (day 2; 250 mL baffled flask) and 3rd (day 3; 250 mL baffled

flask) by inoculating 1 mL (containing over 1.7×10^8 transformants for an over 5000-fold library coverage) from the previous subculture that reached cell confluence.

4.4.3 Library Transformation Achieving a 200-fold Coverage

The ratio of the least represented plasmid to the most is about 1:190. The optimized electroporation described in the previous section 3.4.4 was used to achieve an over 200-fold sgRNA library coverage in both CBS 6556 *ura3Δ ku70Δ* and CBS 6556 *ura3Δ ku70Δ I4::Cas9*, in order to statistically guarantee that every single plasmid from the original library got represented at least for once in the yeast transformant starter. Fourteen electroporating reactions were harvested, resuspended in equal volume of SD-U, and pooled into 500 mL SD-U for outgrowth. For library transformation into each strain, duplicated transformations using the quality control plasmid pIW1250 were included. Every single transformation (14 + 2 in total for each strain) was checked for its apparent successful rate by plating on SD-U by 10,000-fold dilution for cell counting. As for CS control strain, CBS 6556 *ura3Δ ku70Δ*, the total number of transformants was calculated by adding up all 14 reactions; while for the experimental strain, CBS 6556 *ura3Δ ku70Δ I4::Cas9*, the transformant average of biological duplication using the quality control plasmid pIW1250 timed by 14 indicated the real number of successful transformants with library plasmids.

4.4.4 Library Harvesting, Isolation, and Barcoding for High Throughput Sequencing

At the end of every subculturing, 1 mL of cultures was treated by DNase I at 30 °C for 2 h (adding 40 µL of 10× DNase I Reaction Buffer, 6 µL of 2,000 units/ml

enzymes for day 1 cultures and 4 µL for day 2 and day 3 cultures) to get rid of plasmids remaining in the medium. Yeast cell pellets were harvested by centrifugation for 1 min at 3,000 g to isolate plasmid using Zymoprep Yeast Plasmid Miniprep II kit. If plasmid isolation is not going to be done right after cell harvesting, cell pellets were stored at -80 °C. Before plasmid isolation, cell pellets were washed in 1 mL ddH_2O, resuspended in 1 mL Solution 1 in the kit, and mixed with 30 µL Zymolyase™. The amount of enzyme was doubled for our cells harvested from stationary phase cultures as recommended by the protocol to ensure efficient cell lysis. Considering the DNA binding capacity of 5 µg per column, two columns were used to capture all isolated yeast plasmids from every 1 mL cultures (two reactions on one column, and the rest three on the other). 30 µL of DNase, RNase-free and 0.2-micron filtered water was used to firstly elute DNA captured on one column, then the elution to elute DNA from the other. All DNA solutions were kept in -20 °C.

To avoid bubble products brought from the PCR in barcoding sequencing amplicons and to minimize sequence bias because of library amplification, the amount of DNA template as well as the amplification cycles were both limited. As for the 50-µL Q5® High-Fidelity DNA polymerase reaction, we found 0.1 ng of the 6,275-bp plasmid template and 16 cycles were optimum settings for library amplification. The amount of library plasmid in the isolated DNA samples was quantified by bulk SYBR Green (SsoAdvanced™ Universal SYBR® Green Supermix, Bio-Rad) quantitative PCR (qPCR) on Bio-Rad CFX Connect™ using primers qPCR_1f and qPCR_1r (Table S4.2). The amplified and barcoded libraries were cleaned up and size selected by Agencourt®

AMPure®SPRI paramagnetic beads (0.9× for 250-bp amplicons). Libraries were pooled in equimolar amount for a conservative expectation to obtain around 450 reads per crRNA on a NextSeq High Output 75-bp Single End run at UCR IIGB Genomics Core Facility. After demultiplexed by commercial Illumina barcodes, sequencing results were further post-demultiplexed based on designed internal barcodes in the forward primer, trimmed to only 20-nt left, and mapped to each designed crRNA with no mismatching tolerance. Pairwise comparisons between two biological replicates for raw reads and normalized reads were done respectively to confirm the consistency of library representation by yeast cultures.

4.4.5 Cutting score (CS) calculation and statistical analysis

Since both raw reads and normalized reads from each biological replicate agreed well with others, the combined normalized reads per million from all replicates were used to calculate CS. As for crRNA which had zero read, a pseudo count of 1 was added for a valid log2 calculation. Active and inactive crRNAs from the total of 33,243 were identified by two-sided multiple t-test implemented in Graphpad Prism 8 to compare the CS of each library guide to the null distribution consisting of 651 non-targeting guides. The cutoff for statistical significance was 0.05 corrected by the recommended Holm-Sidak method. Fuzzy C-means clustering (FCM) algorithm was used to grade all active guides by the membership assigned to each CS, which quantified the probability for them to be in the cluster distinct from the non-targeting population.

4.5 Supporting Information

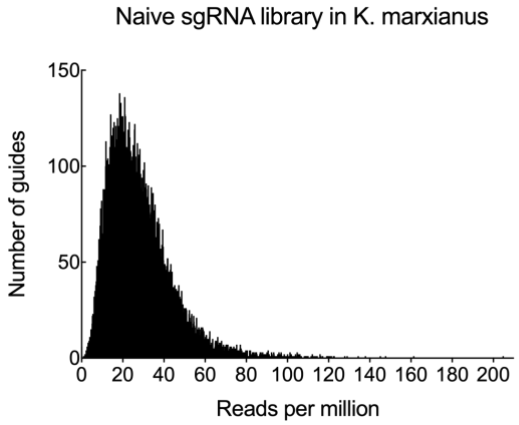

Figure S4.1. Validation of sgRNA library construction. A pooled triplicates of library amplicons were sent for sequencing by MiSeq 1×75. Only 6 out of the 33,984 were not detected.

Table S4.1. Plasmids used in this study

Plasmids	Description	Reference
pIW1213	Removing PScTEF1-KmCas9-SV40-ScCYC1t in pIW601	Chapter 3
pIW1250	pIW1213 with 20-bp NonTargeting_1 sequence	This chapter
KmLibrary_1	pIW1213 with 33,243 of 20-bp spacer sequences and 651 of 20-bp non-targeting sequences	This study

Table S4.2. Primers used for cloning in this study

Primers	Sequence (5' to 3', 20-nt spacers or priming parts are underlined)	Use
ML_249	CCTAGGGCGTATACTCCAACTTG	pIW601_linearize_f
ML_240	GAGCTCCAATTCGCCCTATAGTG	pIW601_linearize_r
ML_251	CACTATAGGGCGAATTGGAGCTCCACTTCCT AGGGCGTATACTCCAACTTG	pIW601_reassemble_f
ML_242	CAAGTTGGAGTATACGCCCTAGGAAGTGGAG CTCCAATTCGCCCTATAGTG	pIW601_reassemble_r
CasML_NT1_f	GAATCCCGTCAGTGTCAACCTATAAGGAGGT ACCACCGGCGGTTTTAGAGCTAGAAATAG	NonTargeting1_insert_f
CasML_NT1_r	CTATTTCTAGCTCTAAAACCGCCGGTGGTACC TCCTTATAGGTTGACACTGACGGGATTC	NonTargeting1_insert_r
OligoPool_1	GTGACACGGGTTCGAATCCCGTCAGTGTCAN NNNNNNNNNNNNNNNNNNTTTTAGAGCTA GAAATAGCAAGTTAAAATA	Insert to assemble KmLibrary_1

Primers	Sequence (5’ to 3’, 20-nt spacers or priming parts are underlined)	Use
qPCR_1f	TAACAAGGCGAGGGAGTG	Library_quantification_f
qPCR_1r	TTCAAGTTGATAACGGACTAGC	Library_quantification_r
L5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACACGGGTTCGAATCCCGT	Amplicon_Illumina_f
L6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTACGTGACACGGGTTCGAATCCC	Amplicon_Illumina_f
L7’	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGATACGTGACACGGGTTCGAATCCC	Amplicon_Illumina_f
L8’	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCGTTACTTGACACGGGTTCGAATCCCGT	Amplicon_Illumina_f
Cr_1669	CAAGCAGAAGACGGCATACGAGATTCGCCTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1670	CAAGCAGAAGACGGCATACGAGATATAGCGTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1671	CAAGCAGAAGACGGCATACGAGATGAAGAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1672	CAAGCAGAAGACGGCATACGAGATATTCTAGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1673	CAAGCAGAAGACGGCATACGAGATCGTTACCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1709	CAAGCAGAAGACGGCATACGAGATGTCTGATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1710	CAAGCAGAAGACGGCATACGAGATTTACGCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1711	CAAGCAGAAGACGGCATACGAGATTTGAATAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r

Table S4.3. Strains used in this study

Strains	Description	Reference
E. coli TOP 10	F- mcrA Δ (mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ (araleu)7697 galU galK rpsL (StrR) endA1 nupG	Thermo Fisher Scientific
YS402	*K. marxianus* CBS 6556 ura3	Löbs et al. [11]
CS cntrl strain	YS402 ku70Δ	This chapter
CS expmt strain	CS cntrl strain I4::P_{ScTEF1}-KmCas9-SV40-ScCYC1t	This chapter

4.6 Reference

[1] Baisya D, et al. Genome-wide functional screens enable the prediction of high activity CRISPR-Cas9 and -Cas12a guides in *Yarrowia lipolytica*. Nat Commun. 2022;13(1): 922.

[2] Ramesh A, et al. acCRISPR: An activity-correction method for improving the accuracy of CRISPR screens. bioRxiv. 2022;499789.

[3] Löbs AK, et al. High throughput, colorimetric screening of microbial ester biosynthesis reveals high ethyl acetate production from *Kluyveromyces marxianus* on C5, C6, and C12 carbon sources. Biotechnol J. 2016;11(10):1274-1281.

[4] Ji RY, et al. Metabolic Engineering of Yeast for the Production of 3-Hydroxypropionic Acid. Front Microbiol. 2018;9:2185.

[5] Henry SA, et al. Metabolism and regulation of glycerolipids in the yeast *Saccharomyces cerevisiae*. Genetics. 2012;190(2):317-349.

[6] Nurcholis M, et al. MIG1 as a positive regulator for the histidine biosynthesis pathway and as a global regulator in thermotolerant yeast *Kluyveromyces marxianus*. Sci Rep. 2019;9(1):9926.

[7] Westholm JO, et al. Combinatorial control of gene expression by the three yeast repressors Mig1, Mig2 and Mig3. BMC Genom. 2008;9(1):601.

[8] Lundin M, et al. Importance of a flanking AT-rich region in target site recognition by the GC box-binding zinc finger protein MIG1. Mol Cell Biol. 1994;14(3):1979-1985.

[9] Bever-Sneary D, et al. RNA Polymerase II-Driven CRISPR-Cas9 System for Efficient Non-Growth-Biased Metabolic Engineering of *Kluyveromyces marxianus*. Metab. Eng. Commun. 2022. Currently under review.

[10] Rajkumar AS, et al. Biological Parts for *Kluyveromyces marxianus* Synthetic Biology. Front Bioeng Biotechnol. 2019;7:97.

[11] Löbs AK, et al. CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol Biofuels. 2017;10(1):164.

Chapter 5: Optimized Genome-wide crRNA Library for Cas9-mediated Functional Screening in *K.marxianus*

5.1 Abstract

Deactivating mating type switching makes homothallic *K. marxianus* a much more robust platform especially for high-throughput screens heavily dependent on differentially accumulated gene editing fitness in a population isolated at culture confluence. An efficient crRNA library with adequate gene coverage is another critical prerequisite to realize genome-wide functional screening. Given that most of the randomly selected crRNAs cannot induce efficient cutting with our current CRISPR/Cas9 configuration (Figure 4.1B), six crRNA activity prediction algorithms provided by CHOPCHOP v3 was used to generate an optimized library predicted to be biased to high on-target activities. In-house uniqueness assessment criteria were applied to minimize off-target effects, but in the meanwhile, save the most active to the maximum possible extent. This optimized library enabled us to identify the first set of essential genes for the *Kluyveromyces* genus.

5.2 Introduction

As shown from Chapter 4, generating crRNA sequences in a completely random way from the CBS 6556 genome results in a library greatly biased to low active guides. Inefficient CRISPR-mediated editing obscures gene fitness screening results [1]. Therefore, a sgRNA library promising adequate gene coverage by sufficient cutting

efficiency is a critical prerequisite to elucidate functional genomics of *K. marxianus* CBS 6556. To increase the possibility of electing highly active guides, six on-target efficiency prediction algorithms embedded in CHOPCHOP v3 were used to design the library with an even higher crRNA redundancy (10 per CDS).

The false discovery rate (FDR) of identifying essential genes or any other gene subgroups related to a phenotype of interest, is heavily dependent on how much confidence we can put on the targeting fidelity of each crRNA in the library. We enhanced the default uniqueness assessment in CHOPCHOP v3 by enforcing at least one mismatch occurrence within the PAM-proximal 12 nucleotides, if any. In the meanwhile, crRNAs with the top highest predicted on-target efficiencies were kept in the final library to the maximum extent possible. As for those that did not meet the default uniqueness requirement but were found only within repeated elements of the same CDSs or because of gene redundancy, conditional exemption was applied to reassess their sequence uniqueness.

In addition to a crRNA library of high on-target activity and low off-target leakage, a stable strain platform is another important factor to increase both the sensitivity and accuracy to recognize editing effects. As a homothallic yeast species [2], wildtype *K. marxianus* is not an intrinsically ideal choice for a high-throughput screening heavily dependent on differentially accumulated gene editing fitness in the population at culture confluence. Theoretically speaking, as long as the editing efficiency does not reach 100%, even a lethal loss of function can be accumulated to some extent because of

the stress-induced mating type switch that allows mating with another cell with the intact gene function (Figure S5.1).

Therefore, in this chapter we constructed stable *a* haploids for a more robust screening platform, annotated the CBS6556 genome by ab initio gene predictor, validated an optimized genome-wide sgRNA library, and identified the first set of essential genes of the *Kluyveromyces* genus.

5.3 Results and Discussion

Two transposes Alpha3 and Kat1 were found responsible for homothallic sexual reproduction in *K. marxianus* [2,3,4]. Gene *ALPHA3* and *KAT1* were functionally knocked out by CRISPR/Cas9 to obtain stable haploids. Figure 5.1A showed a colony PCR (cPCR) design for mating type validation. With the dual knockouts of *ALPHA3* and *KAT1*, stable *a* haploids to study functional genomics were obtained (Figure 5.1B).

A stable haploid is not supposed to switch its original mating type to the other anymore even under mating induction. Therefore, our engineered *a* haploids (no. 1 to 4 in Figure 5.1B) constructed via functionally knocking out gene *ALPHA3* and *KAT1* were confirmed to be stable since only the cPCR testing for *MATa* locus gave a band of 1,874 bp, and no band was obtained in the cPCR for *MATα*. In contrast, the wildtype *a* haploids (no. 5 to 8 in Figure.5.1B) obtained after mating induction on 2% glucose plate turned to be *α/a* diploids. The two cPCRs for them both gave a single band, 2,659 bp for *MATα*, and 1,874 bp for *MATa*, respectively.

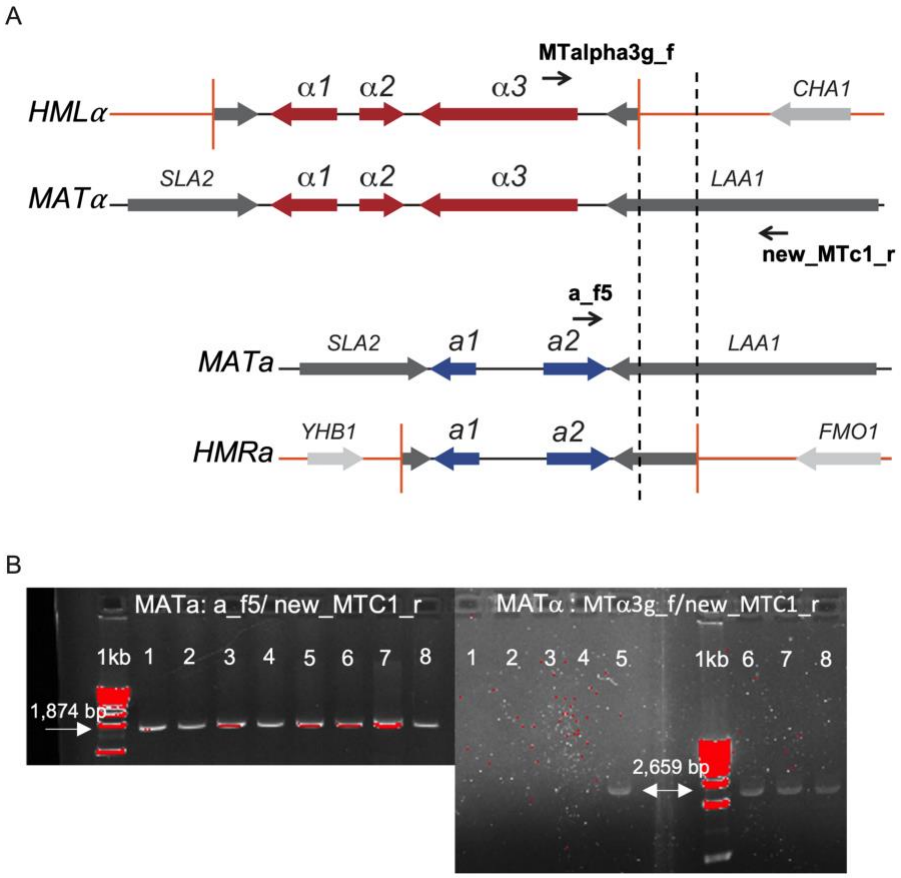

Figure 5.1. Confirmation of the mating type stability of engineered *K. marxianus a* haploids. (A) Adjusted schematic diagram [2] of mating type loci *MATα/a*, and silenced *HMLα* for *a* haploids, *HMRa* for *α* haploids. Primers designed for mating type validation are shown. (B) Colony PCR results for engineered stable *a* haploids (1 to 4) and wildtype *a* or *α* haploids (5 to 8) after induced for self-mating on 2% glucose plate for 54 h at 30 °C. Colony #1 is *K. marxianus* CBS 6556 *ura3Δ α3Δ kat1Δ a* haploid. Colony #2 is *K. marxianus* CBS 6556 *ura3Δ α3Δ kat1Δ I4::Cas9 a* haploid. Colony #3 is *K. marxianus* CBS 6556 *ura3Δ ku70Δ α3Δ kat1Δ a* haploid. Colony #4 is *K. marxianus* CBS 6556 *ura3Δ ku70Δ α3Δ kat1Δ I4::Cas9 a* haploid. Colony #5 is *K. marxianus* CBS 6556 *ura3Δ a* haploid. Colony #6 is *K. marxianus* CBS 6556 *ura3Δ I4::Cas9 a* haploid. Colony #7 is K. marxianus CBS 6556 *ura3Δ ku70Δ a* haploid. Colony #8 is *K. marxianus* CBS 6556 *ura3Δ ku70Δ I4::Cas9 α* haploid.

Figure 5.2 depicts the workflow to generate CHOPCHOP prediction for sgRNAs of *Kluyveromyces marxianus* CBS 6556. The CBS 6556 genome was annotated by an ab initio gene predictor AUGUSTUS trained independently by gene models of *K.marxianus*

DMKU3-1042 and strain FIM1, the two with the largest number of annotated coding sequences (CDSs) in *K. marixnuas* species published on NCBI. Based on each predicted annotation, the first 5%-65% of every CDS was located along the genome. The CBS 6556 genome FASTA file and the identified start and end position of the first 5%-65% of each CDS were both used as input to the command-line version of CHOPCHOP v3 [5]. The output was a list of 20-nt crRNAs targeting the first 5%-65% of all CDSs with a multidimensional evaluation of each guide, considering possible secondary structure formation (self-complementarity score), sequence uniqueness on the genome (MM0, MM1, MM2 and MM3), and a comprehensive naïve cutting score averaged by six independently predicted on-target efficiencies after normalization (Designer v1, Designer v2, Spacer Scoring for CRISPR (SSC), CRISPRscan, CRISPRspec, and uCRISPR) [6,7,8,9,10,11].

Self-complementarity of a crRNA might hinder sufficient targeting, thus inhibiting on-site cutting. The secondary structure of each 20-nt crRNA sequence is predicted by free energy minimization using RNAfold in ViennaRNA Package [12]. The number of self-complementary stem loops no shorter than 4-nt formed within the crRNA was defined as self-complementary score.

The uniqueness of crRNA was well considered in the library design to avoid off-target effects. We enhanced the uniqueness assessment originally embedded in CHOPCHOP v3 by introducing a new score seed_MM0, which was the number of exact matching of seed sequence, the PAM-proximal 12 nucleotides located at the 3′ end of a crRNA [13], to the rest of the genome. The crRNAs confirmed to repeat only within the

same CDS but not anywhere else in the genome were exempted from the uniqueness assessment using MM0 = seed_MM0 = MM1 = MM2 = 0 and MM3 = 0, 1, 2, 3 to be considered as unique. Beside repetitive elements occurring in one CDS, the *K. marxianus* CBS 6556 genome has gene duplication. As for crRNAs which have off-targeted alignments only within the other identified repetitive CDSs because of gene redundancy, they were also considered as unique. Therefore, the gene essentiality in our study was not affected by gene duplication. If a gene's function plays an essential role to maintain cell viability, this gene is going to be identified as essential from CRISPR-mediated fitness screens even if it has multiple copies along the genome.

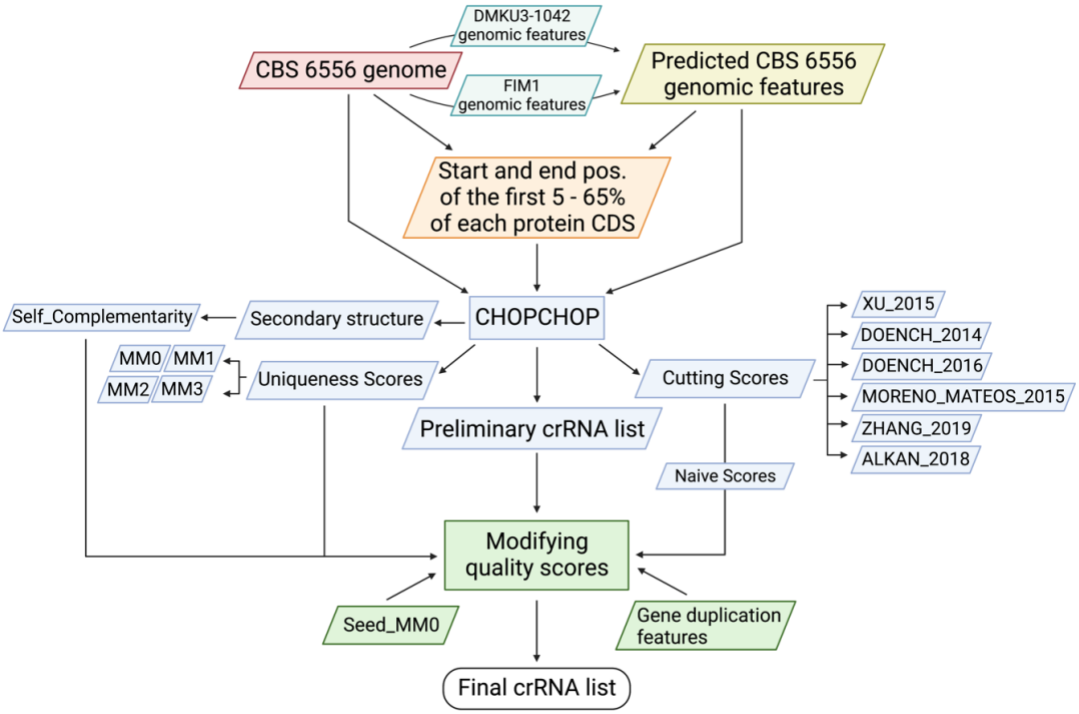

Figure 5.2. Workflow diagram of the genome-wide sgRNA library design by CHOPCHOP on-target efficiency prediction and customized uniqueness enforcement.

There were 4,943 genes with 407,130 crRNAs in total identified by the gene model of *K. marxianus* DMKU3-1042, and 5,000 with 407,165 by *K. marxianus* FIM1, enabling an over 80-fold gene coverage, respectively. Considering the transformation efficiency and culture size, a 10-fold CRISPR crRNA library is feasible while still maintaining a good gene coverage.

About 96% of the annotated genes were unique in the CBS 6556 genome (4,825 identified by the DMKU3-1042 gene model, and 4,870 by the FIM1 model). Less than 2% of these unique genes have internal repetitive elements no shorter than 20 bp (74 out of 4,825, and 83 out of 4,870). For most unique genes, which did not have long enough internal sequence repetitiveness, only crRNAs having no stable secondary structure formed and no more than three other alignments in the genome with no less than three mismatches and at least one of it in the seed sequence were considered unique (self-complementarity score = seed_MM0 = MM0 = MM1 = MM2 = 0, and MM3 = 0, 1, 2, or 3). 20 predicted crRNAs with the top highest naïve cutting scores of each CDS were elected to apply the uniqueness assessment criterion in case any guide of the top 10 highest predicted cutting activity was removed. As for the left 2% unique CDSs which has repeated spacer sequences, none of them was able to have a close number of qualified guides to 10. We then substituted seed_MM0 = MM0 = 0 by seed_MM0 + 1 = MM0 + 1 = crRNA_occurrence (the number of times a crRNA occurring within the CDS of interest) for reasonable and necessary uniqueness exemption to bring back some falsely discarded crRNAs which were in the most active 10 predicted by CHOPCHOP.

Putative gene duplication in the *K. marxianus* CBS 6556 genome was identified by aligning redundant genes in *K. marxianus* DMKU3-1042 to predicted CDSs of CBS 6556. 118 genes were found having at least one duplicated copy by the gene model DMKU3-1042, and 130 by FIM1. Even with the same protein function, the DNA sequences can be different. Therefore, like the strategy used to elect qualified guides for unique sequences with longer-than-20-bp internal repetitive elements, the harshest uniqueness threshold (self-complementarity score = seed_MM0 = MM0 = MM1 = MM2 = 0, and MM3 = 0, 1, 2, or 3) was firstly applied, and then limitations on seed_MM0 and MM0 were relaxed accordingly. As for genes duplicated in the genome, seed_MM0 = MM0 & MM0 + 1 ≤ gRNA_occurrence (the number of times a crRNA repeats within redundant genes) & MM0 + 1 ≤ the number of CDSs of the redundant genes was used as the new uniqueness cutoff to regain qualified crRNAs. It guaranteed that this special uniqueness exemption was only valid for crRNA repetitiveness because of corresponding gene duplication.

Given all above, there were 50,427 crRNAs elected from the gene model of *K. marxianus* DMKU3-1042, and 50,368 from the FIM1 model to generate a 10-fold genome-wide CRISPR library targeting CDSs. For both gene models, over 98% of the crRNAs in the library only targeted at one location along the genome without forming stable secondary structure (self-complementarity score = 0). The two gene models agreed well with each other in general, not only in the total number of crRNAs predicted, but also given that 49,476 (over 98% out of both) were shared. Taking the advantage of using multiple reference gene models to predict putative CDSs in the CBS 6556 genome, we

elected to use the union of 51,319 crRNA targeting all predicted CDSs, which consisted of 49,476 in the consensus, 951 exclusives from the model DMKU3-1042, and 892 only by the model FIM1. Notably, around 25% (1,084 out of the 4,179) of the validated highly active crRNAs from our previous screens were also recognized by CHOPCHOP prediction as in the group of top 10 most active guides (or top 20 for the worst). The rest 3,095 were also located in the genome, so they were added to the library for cutting efficiency control among different libraries and screens. 465 out of them were not within the first 5 – 65% of CDS, so they were excluded from the functional analysis. 1,200 randomly generated 20-nt ATGC sequences (accounting for 2% of the library) whose first 11 nt of the 3' end did not exactly match to the PAM-proximal 11 nucleotides located at the 3′ end of any spacer sequence on the CBS 6556 gnome, were added as negative control. 55,614 20-nt sequences were ultimately included in the 10-fold CRISPR/Cas9 library for genome-wide functional screening. (Figure S5.2 and 5.3)

The same method as described in Chapter 4 was used to evaluate on-target cutting efficiencies by CS for this new library which was intentionally designed to be biased towards high cutting activities (Figure 5.3A). The fitness of each crRNA targeting the first 5 – 65% of putative CDSs was quantified by a fitness score (FS) defined as the log 2 value of its normalized relative abundance in the experimental (expmt) strain over that in the control (cntrl) strain (Figure 5.2B). Not like CS screens, strains used for fitness experiments have intact NHEJ pathway. crRNAs only got depleted in the outgrowth when indels caused by DNA repairs to sufficient Cas9-induced DSBs led to the loss of function of a gene that was somehow essential for cell viability. The higher a gene FS

averaged by its targeted crRNA FSs from the high throughput screen, the more harmless the loss of function of this gene, the less essential the gene to maintain normal cell viability. The two biological replicates for both CS and FS screens agreed well with each other (Figure S5.3) in the number of raw reads of each crRNA from separate sequencing runs, so we combined reads of the two replicates as to calculate CS and FS.

Figure 5.3A showed that CRISPR cutting was happening as soon as in the first outgrowth of library transformants (day 1), indicated by that the median of CSs of non-targeting guides shifted to -3, while the median CS of the targeting guides was around 0. It referred to that the relative abundance of non-targeting guides increased by 8 folds even though this subgroup only accounted for 2% of the total library. The cutting efficiency of each library guide still could not differentiate the more active from the less at day 1, since now the relative guide abundance remained statistically the same between the cntrl and expmt strains (CS median close to zero).

As the day 1 culture got refreshed into the 2nd (day 2) and then the 3rd (day 3) subculture, it was clear that crRNAs with higher cutting efficiencies started to be recognized from others since day 2 (Figure 5.3A and C). At day 3, two distinct CS clusters formed (Figure 5.3A). The one consisting of CSs above 6.5 where the deviation of CSs was much smaller than the other with CSs between 0 to 6. It implied that CS is positively correlated to cutting efficiency. As for crRNAs efficient enough to enable gene editing in every single cell when allowing the CRISPR/Cas9 expressed with adequate amount of time (after three cycles of subculture in our case), their CS values turned to be aggregated. Generally speaking, a good number of crRNAs in this CHOPCHOP elected

library are fairly efficient, as the CS distribution of the 51,319 library guides was highly similar with the 4,179 validated good ones from previous high throughput screens, and was also comparable to that of the 34 positive controls which were confirmed by individual CRISPR experiments to have an over 40% knock-out rate, not to mention that its median was 7 CS away from that of the non-targeting guide CS distribution (equal to 128-fold lower normalized guide representation in the expmt strain).

The FS distribution (Figure 5.3B) was expected by CS, since a sufficient cut was the prerequisite for indel occurrence leading to reading frame shift and multiple premature termination codons (PTCs) which resulted in a functional gene knock-out. FS of the non-targeting guides indicated that the negative effects of constitutively expressing Cas9 and sgRNA was negligible comparing with making a DSB in the genome, since as early as before the detrimental crRNA got depleted and those neutral or even beneficial gene edits accumulated, the FS median of non-targeting guides was almost 8-fold higher than the targeting guides (FS at day 1, Figure 5.3B). The number of harmful gene edits and the corresponding crRNAs was significantly reduced in the cell culture at confluence since day 2 (FS at day 2 and 3, Figure 5.3B).

The comparison of CS and FS confirmed that at least two cycles of subculturing were needed to recognize the difference of cutting efficiencies and gene edit effects. It was a general success to define a non-targeting 20-nt sequence meeting the criterion that its first 11 nt of the 3’ end had no exact match to the first 11-nt of the 3’ end of any identified crRNA sequences in the genome. Only 1% of the 1,200 non-targeting sequences had a positive CS since day 2. The negative proportional correlation of CS and

FS of the randomly generated and elected 20-nt ATGC sequences indicated that the more essential a gene is to maintain basic cell viability, the easier it was accessible to a Cas9-induced cleavage. A significant number of crRNAs clustered around FS of -9 at day 3 (Figure 5.3B), and interestingly shown in the Figure 5.3C for day 3, the CS of these crRNAs, in contrast, spread from -0.5 to 9, covering nearly a full range of cutting efficiencies. It implied that the more essential a gene is for cell viability, the less dependent on the gene editing efficiency to make the knocked-out effect detectable.

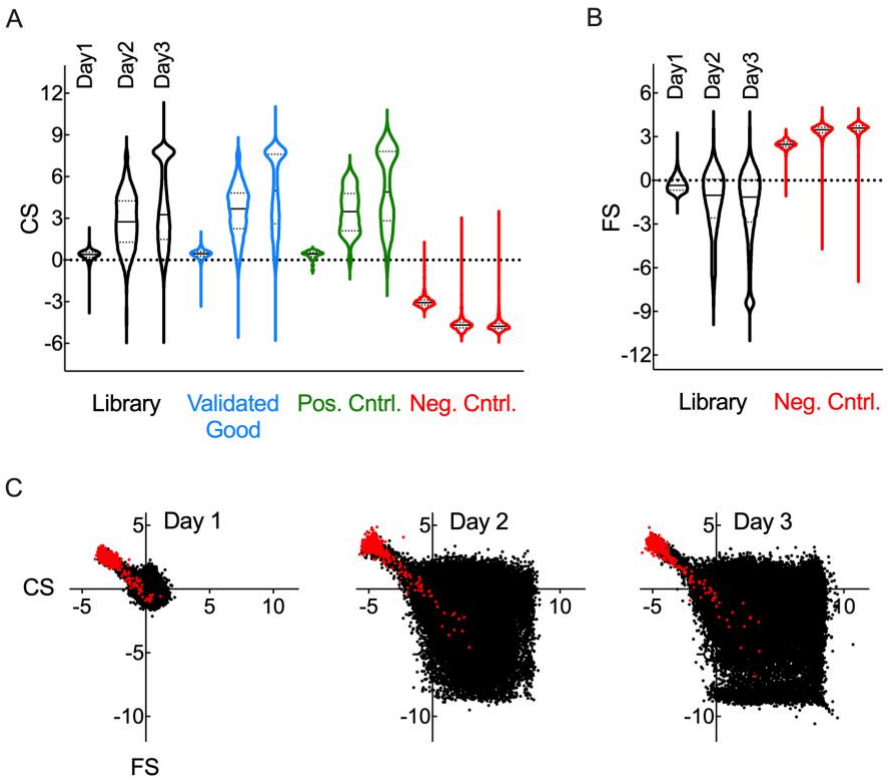

Figure 5.3. Validation of on-target cutting efficiency (CS) and fitness (FS) of CHOPCHOP predicted genome-wide sgRNA in *K. marxianus* CBS 6556. (A) Gene editing activity (CS) distribution of 51,319 library sgRNAs selected by CHOPCHOP as the top 10 guides for each CDS with respect to cutting efficiency, 4,179 validated good guides from previous genome-wide screening libraries, 34 positive controls characterized by individual CRISPR experiments, and 1,200 non-targeting guides as negative control. (B) sgRNA fitness (FS) distribution of the predicted library and negative controls. (C) Comparison of FS to CS for each sgRNA. Cutting score (CS) experiments used *K. marxianus* CBS 6556 *ura3Δ* with functionally disrupted non-homologous end joint (NHEJ) pathway by inactivation of *KU70* and with constitutive *Streptococcus pyogenes* Cas9 expression integrated at the non-coding site I4 (CBS 6556 *ura3Δ I4::Cas9 ku70Δ*) as the experimental (expmt) strain, and CBS 6556 *ura3Δ ku70Δ* as the control (cntrl) strain. Fitness score (FS) experiments used CBS 6556 *ura3Δ I4::Cas9* as the expmt strain and CBS 6556 *ura3Δ* as the control. Cntrl strains provided the relative abundance baseline of each guide after transformed into cells. Loss of the relative abundance of an sgRNA in the expmt strain of CS experiments indicated efficiency to induce a DNA break and get repaired by the NHEJ pathway. Loss of the relative abundance of an sgRNA in the expmt strain of FS experiments indicated how detrimental LOF indels were for cell fitness. Transformations were done in biological duplicates with over 100-fold coverage of the library in each replicate, and libraries were harvested at cell confluence for every subculture cycle (day1, 2 and 3). CS and FS values shown were calculated by combining raw reads from both two replicates.

Higher gene coverage and higher editing efficiency (CS) are both preferable to study functional genome with more confidence. However unfortunately, these two parameters cannot achieve their own best simultaneously. One of our recent works [14] introduced a quantitative metric, ac-coefficient, to resolve this dilemma by optimizing the tradeoff between guide editing efficiency (CS) and library coverage. The CS threshold and averaged gene coverage of the genome resulting in the highest value of ac-coefficient were elected using CS obtained at day 2 and day 3, respectively (Figure 5.4). The optimum CS threshold was 3.0 for day 2, with corresponding gene coverage of 5 crRNA per gene in average, while as for day 3, the CS cutoff increased to 6.5, and gene coverage declined to 3. These two optimized CS thresholds suggested by ac-coefficient were also expected based on the CS distribution of day 2 and 3 (Figure 5.3A). CS of day 2 formed a single but more spreading-out population whose median was around 3.0, while at day 3, though an almost identical CS population to that of day 2 maintained, the size of it shrank, and a new CS population emerged, with narrower value distribution where the lowest was about 6.5 (Figure 5.3A). Subculturing library transformants for three times was shown to better guarantee adequate amount of time than only doing two evolution cycles to differentiate CRISPR/Cas9-mediated gene editing efficiencies quantified by foldchange of the relative abundance of corresponding crRNA between the control with no Cas9 expression and the Cas9 expressing group at cell culture confluence, as well as to reaggregate close efficiencies indicated by CS. The comparison of putative essential genes identified by CS of day 2 and day 3 also implied that the day 3 data enabled

essential gene prediction of higher confidence, though over 80% of the predicted essential genes by day 2 data were in the consensus set.

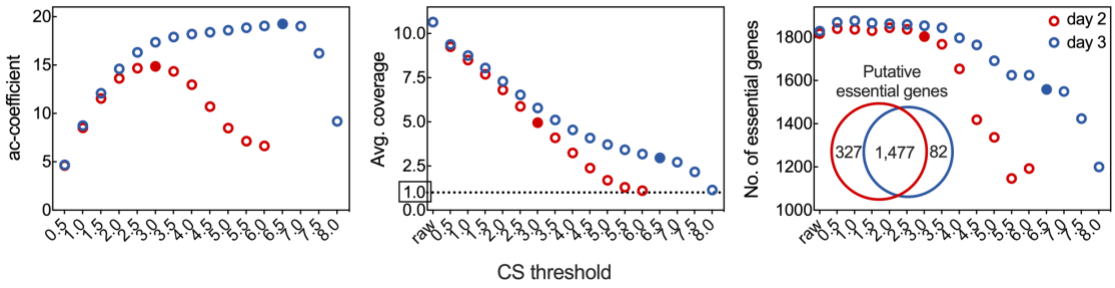

Figure 5.4. Essential genes identification for *K. marxianus* CBS 6556 optimized by ac-coefficient to achieve the theoretically best balance between editing efficiency and gene coverage.

5.4 Materials and Methods

5.4.1 Molecular Cloning and Reagents

All molecular cloning reagents and enzymes were purchased from NEB. Plasmids created and used in this work are listed in Table S5.1. CRISPR plasmids to functionally knock out gene *ALPHA3* and pseudo gene *KAT1* to prohibit mating type switching were constructed using pIW601 (Addgene ID 98907) linearized at PspXI and re-assembled with a 60 bp insert (listed in Table S5.2) containing 20 bp upstream and downstream homology as well as the 20 bp target sequence by Gibson assembly. Colony PCRs (cPCRs) to confirm strains' mating types were performed using Taq polymerase. All primers (Table S5.2) were purchased from IDT™. All molecular cloning was accomplished in TOP 10 competent *E. coli* (Thermo Fisher Scientific).

Other reagents used in this work were purchased from Fisher Scientific, Sigma-Aldrich or as otherwise noted: BD Difco™ Yeast Nitrogen Base without Amino Acids

(Fisher Scientific), CSM-Ura powder (Sunrise Science Products), D-glucose (Fisher Scientific), Gibco™ Bacto™ Yeast Extract (Fisher Scientific), Gibco™ Bacto™ Peptone (Fisher Scientific), 5-Fluoroorotic acid (5-FOA; Fisher Scientific), Miller's LB powder (Sigma-Aldrich), Ampicillin sodium salt (Fisher Scientific), Agar (Fisher Scientific), Tris Base (Fisher Scientific), Boric Acid (Fisher Scientific), EDTA disodium salt (Sigma-Aldrich), 12 M Hydrochloric Acid (Fisher Scientific), 1% Ethidium Bromide (Fisher Scientific), Agarose (Fisher Scientific), DNase and RNase free, 0.2-micron filtered water (Fisher Scientific), Lithium acetate dihydrate (Sigma-Aldrich), 10× TE buffer (pH 7.4; Fisher Scientific), Dithiothreitol (Fisher Scientific), and Sorbitol (Fisher Scientific).

5.4.2 Construction and Validation of Stable CBS 6556 *a* Haploids for CS and FS Screens

Stable *a* haploids for FS screening were first constructed, and then the gene *KU70* was functionally knocked out in both strains with and without Cas9 expression integrated, to obtain the pair of stable *a* haploids for the CS experiment. A simplified heat shock and chemical induced transformation method (Chapter 2) was used in these three individual CRISPR experiments to disrupt the gene *ALPHA3*, *KAT1*, and *KU70*, respectively. Primers MS_007/MS_008, ML_459/ML_444, and SeqML_040/SeqML_041 were used to confirm that CRISPR-mediated indels resulted in corresponding open reading frame shift and multiple premature termination codons for functional gene disruption by cPCR and Sanger sequencing of the amplicons. Plasmids were removed from selected colonies that gave desirable gene disruption results by culturing in 1 mL YPD containing 1 g/L 5-FOA at 30 °C for 16 h.

The *MATa* loci flanked by gene *SLA2* and *LAA1* and the silenced *HML*α loci flanked by partial *SLA2* and *LAA1* were identified in the CBS 6556 *a* haploid genome by using BLASTn to search homologs annotated in CBS 608 and DMKU3-1042 in CBS 6556. The confirmation method for the loss of homothallism in our genetic engineered stable *a* haploids was simplified from two previous studies [2,3]. A reverse primer, new_MTc1_r, was designed to bind outside the consensus area of the full and partial *LAA1*. Two forward primers, MTα3g_f and a_f5, were for the gene α*3* and *a2*, respectively. Single colonies of engineered stable and wildtype (serving as control for PCR success) *a* haploids were inoculated in 1 mL YPD at 30 °C overnight. Cell pellets were spun down at 5,000 g for 1 min and washed by equal volume of ddH_2O. Cell pellets were then resuspended in 20 µL ddH_2O. 4 µL of cell resuspension was dotted on 2% glucose plate. The second 4 µL was dotted right on top of the first drop after the first one was dry. Plates were incubated at 30 °C for 48 to 72 h for adequate mating induction. Each colony was then tested by two cPCRs with primer pairs MTα3g_f/ new_MTc1_r for *MAT*α and a_f5/ new_MTc1_r for *MATa*, respectively. Stable haploids could not switch its original mating type to the other even after mating induction. Therefore, our engineered *a* haploids that were made by functionally knocking out gene *ALPHA3* and *KAT1* was confirmed to be stable if only the cPCR testing for *MATa* gave a band of 1,874 bp, while no band was obtained in the cPCR for *MAT*α. In contrast, the wildtype *a* haploids obtained from mating induction should be α*/a* diploids. The two cPCRs for them were both expected to give a single band, 2,659 bp for *MAT*α, and 1,874 bp for *MATa*. To obtain enough genomic DNA as template for mating type validation, colonies

were first resuspended in 15 µL ddH2O, boiled on thermocycler for 10 min at 95 °C, and then spun down. 5 µL supernatant was used as the template for each 20 µL Taq reaction.

Glycerol stocks were made and stored in -80 °C for the confirmed stable *a* haploids CBS 6556 *ura3Δ ku70Δ α3Δ kat1Δ* with and without Cas9 expression integrated at the genomic site *I4* for CS screening, and stable *a* haploids CBS 6556 *ura3Δ α3Δ kat1Δ* with and without integrated Cas9 for FS screening.

5.4.3 Optimized Library Design

The CBS 6556 genome was annotated by an ab initio gene predictor AUGUSTUS trained independently by gene models of *K.marxianus* DMKU3-1042 and that of FIM1, the two with the largest number of annotated coding sequences (CDSs) of *K. marixnuas* species on NCBI. Based on each predicted annotation, the first 5% – 65% of each CDS was located along the genome. The CBS 6556 genome FASTA file and the identified start and end position of the first 5% – 65% of each CDS were both used as input to the command-line version of CHOPCHOP v3 [5]. The output was 20-nt crRNAs targeting the first 5% – 65% of all CDSs with a multidimensional evaluation of each guide, considering possible secondary structure formation (self-complementarity score), sequence uniqueness on the genome ((MM0, MM1, MM2 and MM3), and a comprehensive naïve cutting score averaged by six independently predicted on-targeting efficiencies after normalization (Designer v1, Designer v2, Spacer Scoring for CRISPR (SSC), CRISPRscan, CRISPRspec, and uCRISPR) [6,7,8,9,10,11].

Self-complementarity within the nucleotides of a crRNA might hinder sufficient targeting, thus inhibiting on-site cutting. The secondary structure of each 20-nt crRNA

sequence is predicted by free energy minimization using RNAfold in ViennaRNA Package [12]. The number of self-complementary stem loops no shorter than 4-nt formed within the crRNA was defined as self-complementary score.

The uniqueness assessment originally embedded in CHOPCHOP v3 was enhanced by introducing a new score seed_MM0 which was the number of exact matchings of the seed sequence, the PAM-proximal 12 nucleotides located in the 3′ end of a crRNA, to the rest of the genome. The harshest uniqueness cutoff, self-complementarity score = seed_MM0 = MM0 = MM1 = MM2 = 0, and MM3 = 0, 1, 2, or 3, was applied to the top 20 guides ranked according to the on-targeting efficiency predicted by CHOPCHOP v3. If a CDS had less than 10 guides to elect from after applying this uniqueness cutoff, its targeting crRNA was checked in the following two steps to see if it was valid for conditional uniqueness exemption. Firstly, if a crRNA was confirmed to repeat only within the same CDS but not anywhere else on the genome, it was also able to be classified as unique if seed_MM0 + 1 = MM0 + 1 = crRNA_occurrence (the number of times a crRNA occurring within the CDS of interest), self-complementarity score = MM1 = MM2 = 0, and MM3 = 0, 1, 2, or 3. Secondly, if a crRNA was found only among the multiple copies due to gene redundancy, seed_MM0 = MM0 & MM0 + 1 ≤ gRNA_occurrence (the number of times a crRNA is repeated within CDSs of the duplicated genes) & MM0 + 1 ≤ the number of CDSs of the duplicated genes, was used as the new uniqueness cutoff to regain putatively qualified crRNAs.

5.4.4 Library Construction

55,514 of 128-nt oligos concatenating upstream 5'-AGAAGTAACAAGGCGAGGGAGTGACACGGGTTCGAATCCCGTCAGTGTCA-3' homologous, the designed 20-nt spacer sequences, and the downstream 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA-3' were synthesized by Agilent. Primers linearize_f2 and linearize_r were used to linearize plasmid pIW1213 with Q5 polymerase to obtain the library plasmid backbone. Primers OLS_f and OLS_r were used to amplify the 10-pmol ssDNA library with Q5 polymerase. All PCR amplicons were cleaned up by Agencourt® AMPure®SPRI paramagnetic beads (0.8× for the linearized backbone, and 1.8× for the 128-bp inserts). SureVector CRISPR Library Cloning Kit (Part Number G7556A) manufactured by Agilent Technologies was chosen to assemble and amplify the library. About 1 to 1.5 mg of library plasmids were isolated by ZymoPURE II Plasmid Gigaprep Kit.

5.4.5 Library Transformation, Outgrowth, and Library Preparation for NGS

Strategies of library transformation, media selection, cell culturing, and library preparation for sequencing were all the same as described in the Materials and Method session of Chapter 4.

5.4.6 Cutting Score (CS) and Fitness Score (FS) Calculation

Since both raw reads and normalized reads from each biological replicate in CS and FS screens agreed well with others, combined normalized reads per million from all

replicates were used to calculate the scores. As for crRNA which had zero read, a pseudo count of 1 was added to obtain a valid log2 calculation. CS was defined as the log2 value of the normalized reads per million in CBS 6556 *ura3Δ ku70Δ α3Δ kat1Δ a* haploid over the normalized reads per million in CBS 6556 *ura3Δ ku70Δ α3Δ kat1Δ I4::Cas9 a* haploid. FS was defined as the log2 value of the normalized reads per million in CBS 6556 *ura3Δ α3Δ kat1Δ a* haploid over the normalized reads per million in CBS 6556 *ura3Δ α3Δ kat1Δ I4::Cas9 a* haploid.

5.4.7 Essential gene identification

acCRISPR [14] was used to maximize the performance of FS screen to identify putative essential genes by balancing the trade-off between confidence on guide activity and gene coverage.

5.5 Supporting information

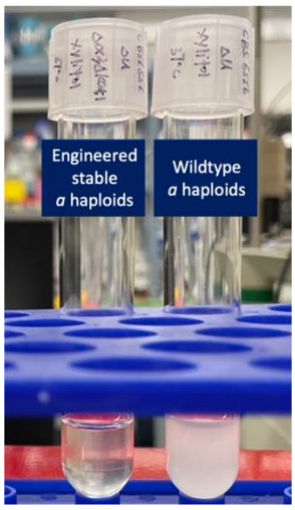

Figure S5.1. Transformants outgrowth of engineered stable *a* haploids (*K. marxianus* CBS 6556 *ura3Δ α3Δ kat1Δ*; left) and wildtype *a* haploids (*K. marxianus* CBS 6556 *ura3Δ*; right) transformed with the CRISPR/Cas9 plasmid pIW447 [15] in selective minimal medium containing 2% xylitol as the sole carbon source. pIW447 expresses an efficient sgRNA targeting *XYL2*, an essential gene for xylitol utilization in *K. marxianus* CBS 6556.

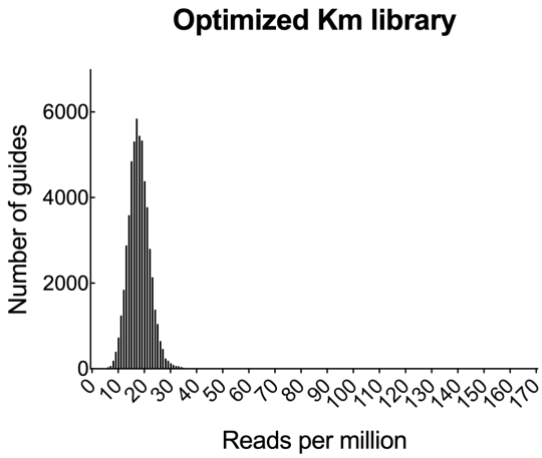

Figure S5.2. Validation of the optimized sgRNA library construction. A total number of 14,474,589 raw reads were obtained for the library of sequence complexity of 55,614. Only one non-targeting guide was not detected. Each sequence would have 18 reads per million total reads if all sequences were theoretically evenly represented. The library formed a normal distribution with the mean of 18.

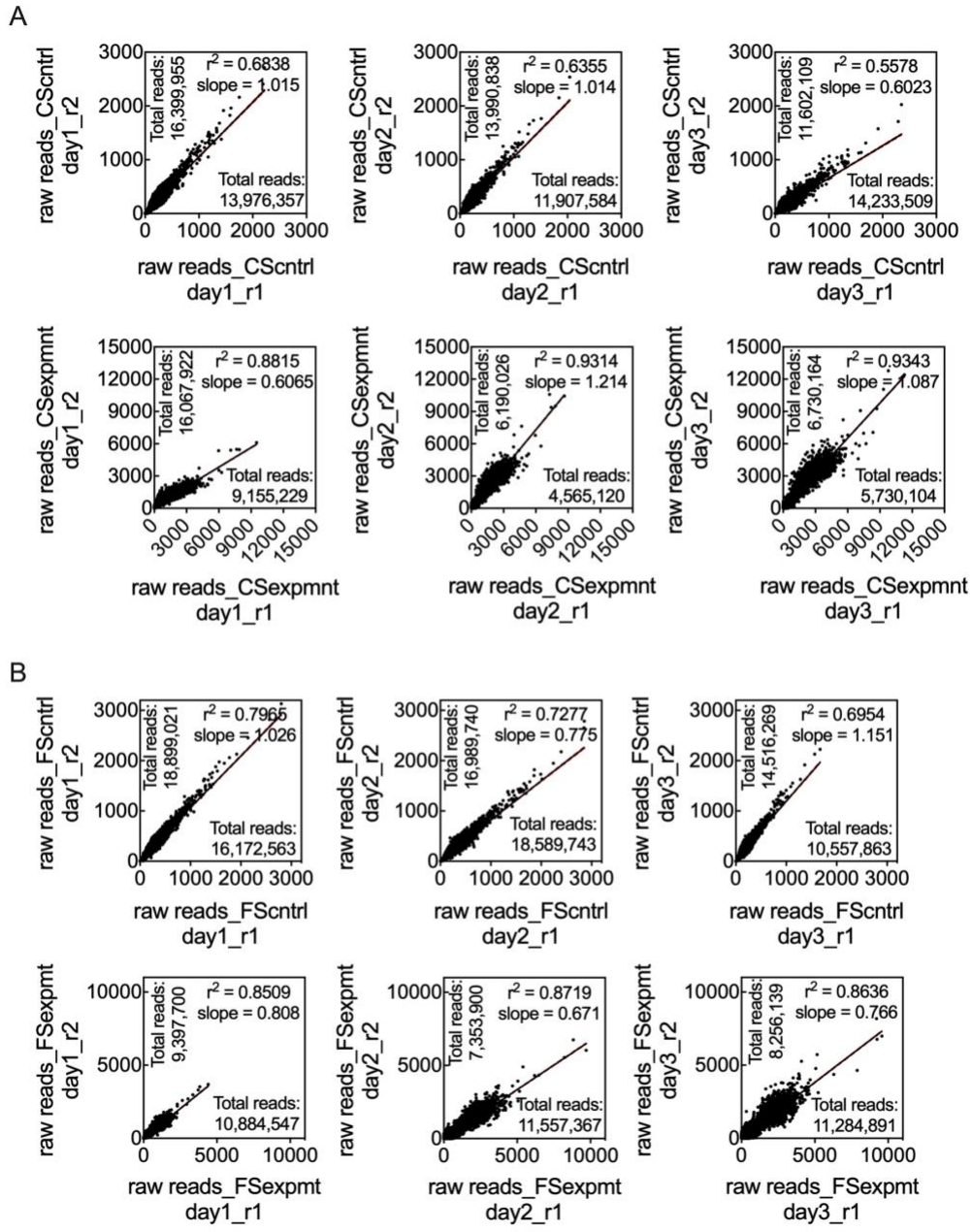

Figure S5.3. Raw reads consistency between screening duplicates. (A) Comparison of raw reads from the two CS experiment replicates. (B) Comparison of raw reads from the two FS experiment replicates.

Table S5.1. Plasmids used in this study

Plasmids	Description	Reference (Addgene#)
pIW601	P_{ScTEF1}-KmCas9-SV40-ScCYC1t, P_{KmPRP1}-tRNA^{Gly}-PspXI recognition site-SUP4	Löbs et al. [15] (98907)
pIW447	pIW601with XYL2 targeting sgRNA	Löbs et al. [15]
pCRISPR_ALPHA3	pIW601with ALPHA3 targeting sgRNA	This chapter
pCRISPR_KAT1	pIW601with KAT1 targeting sgRNA	This chapter
pCRISPR_KU70	pIW601with KU70 targeting sgRNA	This chapter
pIW1213	Removing PScTEF1-KmCas9-SV40-ScCYC1t in pIW601	Chapter 4
pIW1250	pIW1213 with 20-bp NonTargeting_1 sequence	Chapter 4
KmLibrary_3	pIW1213 with 51,319 of 20-bp spacer sequences and 1,200 of 20-bp non-targeting sequences	This chapter

Table S5.2. Primers used for cloning in this study

Primers	Sequence (5' to 3', 20-nt spacers or priming parts are underlined)	Use
CasMS_005	GAATCCCGTCAGTGTCAACCTAAAACGTTTTGAAGACACAGGTTTTAGAGCTAGAAATAG	pCRISPR_ALPHA3_insert_f
CasMS_006	CTATTTCTAGCTCTAAAACCTGTGTCTTCAAAACGTTTTAGGTTGACACTGACGGGATTC	pCRISPR_ALPHA3_insert_r
CasML_181	TTCGAATCCCGTCAGTGTCAGACATTGGGTTGTTAACTGGTTTTAGAGCTAGAAATAGCA	pCRISPR_KAT1_insert_f
CasML_182	TGCTATTTCTAGCTCTAAAACCAGTTAACAACCCAATGTCTGACACTGACGGGATTCGAA	pCRISPR_KAT1_insert_r
CasML_237	TTCGAATCCCGTCAGTGTCAGCTGGGAAGATACATCGATGTTTTAGAGCTAGAAATAGCA	pCRISPR_KU70_insert_f
CasML_238	TGCTATTTCTAGCTCTAAAACATCGATGTATCTTCCCAGCTGACACTGACGGGATTCGAA	pCRISPR_KU70_insert_r
MS_007	GATTTGGAAAGAGGTGTCGGAATCAC	ALPHA3_ko_f
MS_008	CACGTACTTCTTCCTTTATCTGAAG	ALPHA3_ko_r
ML_459	CAACGGAGCTTTGTTATAGGTATGC	KAT1_ko_f
ML_444	ACCCCGACAACATCTCAGAATC	KAT1_ko_r
SeqML_040	GAAACAACAAGTAGAGGAACAC	KU70_ko_f
SeqML_041	CGAATGTCGTGTCATCAAATG	KU70_ko_r
linearize_f2	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	pIW1213_linearize_f
linearize_r	TGACACTGACGGGATTCGAAC	pIW1213_linearize_r
OligoPool_3	AGAAGTAACAAGGCGAGGGAGTGACACGGGTTCGAATCCCGTCAGTGTCANNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA	Insert to assemble KmLibrary_3
OLS_f	AGAAGTAACAAGGCGAGGGA	Oligo_amplification_f
OLS_r	TTTTTCAAGTTGATAACGGACTAGC	Oligo_amplification_r

Primers	Sequence (5’ to 3’, 20-nt spacers or priming parts are underlined)	Use
qPCR_1f	TAACAAGGCGAGGGAGTG	Library_quantification_f
qPCR_1r	TTCAAGTTGATAACGGACTAGC	Library_quantification_r
L5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACACGGGTTCGAATCCCGT	Amplicon_Illumina_f
L6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTACGTGACACGGGTTCGAATCCC	Amplicon_Illumina_f
L7’	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGATACGTGACACGGGTTCGAATCCC	Amplicon_Illumina_f
L8’	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCGTTACTTGACACGGGTTCGAATCCCGT	Amplicon_Illumina_f
Cr_1669	CAAGCAGAAGACGGCATACGAGATTCGCCTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1670	CAAGCAGAAGACGGCATACGAGATATAGCGTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1671	CAAGCAGAAGACGGCATACGAGATGAAGAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1672	CAAGCAGAAGACGGCATACGAGATATTCTAGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1673	CAAGCAGAAGACGGCATACGAGATCGTTACCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1709	CAAGCAGAAGACGGCATACGAGATGTCTGATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1710	CAAGCAGAAGACGGCATACGAGATTTACGCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r
Cr_1711	CAAGCAGAAGACGGCATACGAGATTTGAATAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACTCGGTGCCACTTTTTCAAG	Amplicon_Illumina_r

Table S5.3. Strains used in this study

Strains	Description	Reference
E. coli TOP 10	F- mcrA Δ (mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ (araleu)7697 galU galK rpsL (StrR) endA1 nupG	Thermo Fisher Scientific
YS402	*K. marxianus* CBS 6556 ura3Δ	Löbs et al. [15]
FS cntrl stable *a* haploid	YS402 α3Δ kat1Δ	This study
FS expmt stable *a* haploid	FS cntrl stable *a* haploid I4::P_{ScTEF1}-KmCas9-SV40-ScCYC1t	This study
CS cntrl stable *a* haploid	FS cntrl stable *a* haploid ku70Δ	This study
CS expmt stable *a* haploid	CS cntrl stable *a* haploid I4::P_{ScTEF1}-KmCas9-SV40-ScCYC1t	This study

5.6 Reference

[1] Schwartz C, et al. Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast *Yarrowia lipolytica*. Metab. Eng. 2019;55: 102-110.

[2] Cernak P, et al. Engineering *Kluyveromyces marxianus* as a Robust Synthetic Biology Platform Host. 2018;mBio 9(5).

[3] Lane MM, et al. Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. Antonie Van Leeuwenhoek. 2011;100(4):507-519.

[4] Rajaei N, et al. Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. Proc Natl Acad Sci USA. 2014;111(43):15491-15496.

[5] Labun K, et al. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. 2019;47(W1):W171-W174.

[6] Doench JG, et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat Biotechnol. 2014;32(12):1262-1267.

[7] Doench JG, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34(2):184-191.

[8] Xu H, et al. Sequence determinants of improved CRISPR sgRNA design. Genome Res. 2015;25(8):1147-1157.

[9] Moreno-Mateos MA, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods. 2015;12(10):982-988.

[10] Alkan F, et al. CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. Genome Biol. 2018;19(1):177.

[11] Zhang D, et al. Unified energetics analysis unravels SpCas9 cleavage activity for optimal gRNA design." Proc Natl Acad Sci USA. 2019;116(18): 8693-8698.

[12] Lorenz R, et al. ViennaRNA Package 2.0. Algorithms Mol Bio. 2011;6(1): 26.

[13] Zhang XH, et al. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Mol Ther Nucleic Acids. 2015;4:e264.

[14] Ramesh A, et al. acCRISPR: An activity-correction method for improving the accuracy of CRISPR screens. bioRxiv. 2022;499789.

[15] Löbs AK, et al. CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol. Biofuels, 2017;10(164).

Chapter 6: Summary and Conclusion

The budding yeast *Kluyveromyces marxianus* is a good example of exploiting benefits of using non-conventional hosts for next generation bioproduction. It is thermotolerant to temperatures upward of 52 °C. A growth rate double that of *S. cerevisiae* at 30 °C is highly preferrable to produce growth or biomass formation associated products with high TCA cycle flux for complete carbon oxidation. As a Crabtree negative yeast species, it avoids toxic or unintended ethanol formation under aerobic condition even with high concentration of glucose. Its broad range of substrate utilization enables sufficient conversion of sugars and organic acids hydrolyzed from lignocellulosic biomass wastes to bioproducts of higher commercial values.

The increasing demand of natural flavors and fragrances added in food, beverages, and personal care products brings a unique opportunity of exploiting the potential of *K. marxianus* in related biotechnology applications. As a generally considered as safe (GRAS) microbe, *K. marxianus* is a great fit in producing food and pharmaceutical-grade molecules. The *Kluyveromyces* genus natively has strong metabolic pathways to synthesize aromatic compounds. Of all these biochemicals, 2-phenyl ethanol (2-PE), a compound with rose aroma, is the most important commercially, which is an attractive chemical candidate to work on with. Benefit from the inherent strong Shikimate and Ehrlich pathway in *K. marxianus*, engineering this thermotolerant yeast for overproduction of food-grade natural 2-PE has a bright future.

The first piece of work in this thesis (Chapter 2) enhances native 2-PE production in K. marxianus by developing a multigene integration tool that takes advantage of CRISPR/Cas9 induced double-stranded breaks on the genome as a selection to enable one-step integration of an insert that encodes one, two, or three gene expression cassettes. Integration of an up to 5-kbp insert containing three overexpression cassettes achieved an efficiency of 51 ± 9% at the *ABZ1* locus. A full factorial library of $KmARO4^{K221L}$, $KmARO7^{G141S}$, and *KmPHA2*, each driven by three different promoters that span a wide expression range, was constructed for Shikimate pathway refactoring. It revealed that high expression of the tyrosine-deregulated $KmARO4^{K221L}$ and native *KmPHA2*, with the medium expression of feedback insensitive $KmARO7^{G141S}$, results in the highest increase in 2-PE biosynthesis, producing 684 ± 73 mg/L. Ehrlich pathway engineering by overexpression of *KmARO10* and disruption of *KmEAT1* further increases 2-PE production to 766 ± 6 mg/L. The best strain achieves 1943 ± 63 mg/L 2-PE after 120 h of fed-batch operation in 25 mL shake flask cultures.

Advanced transformation methods, CRISPR/Cas9-mediated screening for gene regulation developed upon chromatin accessibility, stable *a* haploids of *K. marxianus* CBS 6556, and the highly efficient genome-wide sgRNA library for functional genomic deletions descried in Chapter 3, 4 and 5 are all useful tools to study functional genomics and metabolism in *K. marxianus* and other non-conventional yeasts.

Our optimized chemical and heat shock induced method achieves 6×10^5 transformants per reaction in average, using 1.04 pmol DNA (equal to 4 µg for 6,275-bp plasmids) and 5×10^8 of cells. Electroporation gives an even higher yield of

transformants than the chemical method using much less DNA (0.26 pmol) and similar number of cells. Obtaining 1.5×10^6 transformants from a single electroporation reaction is also currently the most efficient method reported for *K. marxianus* in literature. Highly efficient transformation facilitates functional genomic studies by enabling adequate representation of a complex screening library with less required host cells as well as transformed DNA.

With the type II CRISPR/Cas9 system and efficient DNA transformation, we characterized an 8-fold single-guide RNA (sgRNA) library targeting both on promoter sequences and coding regions of the CBS 6556 genome. By culturing library transformants in three carbon sources, glucose, lactose, and xylose, the differentially induced editing efficiencies with the same sgRNA makes the genome-wide CRISPR/Cas9 library a powerful tool to elucidate gene regulations involved in glucose repression.

Knocking out *ALPHA3* and *KAT1* significantly avoids single cell mating type switching in *K. marxianus*, which makes it a much more robust platform for high-throughput screens heavily dependent on differentially accumulated gene editing fitness from a population growing to cell confluence. An efficient sgRNA library with adequate gene coverage was obtained using six crRNA activity prediction algorithms provided by CHOPCHOP v3. In-house uniqueness assessment criteria were applied to minimize off-target effects, but in the meanwhile, save the most active to the maximum possible extent. This optimized library enabled us to identify the first set of essential genes for the *Kluyveromyces* genus.

Overall, this thesis developed stable *α/a* haploid strains for *K. marxianus* CBS 6556, advanced CRISPR-mediated genetic tools, and highly efficient transformation protocols. They all facilitate genetic engineering and system metabolic engineering in *K. marxianus*, as well as better understanding of its genetics and metabolism.