

RESEARCH

Open Access



CRISPR-mediated multigene integration enables Shikimate pathway refactoring for enhanced 2-phenylethanol biosynthesis in *Kluyveromyces marxianus*

Mengwan Li¹, Xuye Lang¹, Marcos Moran Cabrera¹, Sawyer De Keyser¹, Xiyun Sun¹, Nancy Da Silva² and Ian Wheeldon^{1,3*}

Abstract

Background: 2-phenylethanol (2-PE) is a rose-scented flavor and fragrance compound that is used in food, beverages, and personal care products. Compatibility with gasoline also makes it a potential biofuel or fuel additive. A biochemical process converting glucose or other fermentable sugars to 2-PE can potentially provide a more sustainable and economical production route than current methods that use chemical synthesis and/or isolation from plant material.

Results: We work toward this goal by engineering the Shikimate and Ehrlich pathways in the stress-tolerant yeast *Kluyveromyces marxianus*. First, 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 $51 \pm 9\%$ 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 fed-batch operation in shake flask cultures.

Conclusions: The CRISPR-mediated multigene integration system expands the genome-editing toolset for *K. marxianus*, a promising multi-stress tolerant host for the biosynthesis of 2-PE and other aromatic compounds derived from the Shikimate pathway.

Keywords: Metabolic engineering, Flavors and fragrances, Thermotolerance, Expression regulation

Background

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

*Correspondence: iwheeldon@enr.ucr.edu

¹ Department of Chemical and Environmental Engineering, University of California Riverside, Riverside, CA 92521, USA

Full list of author information is available at the end of the article



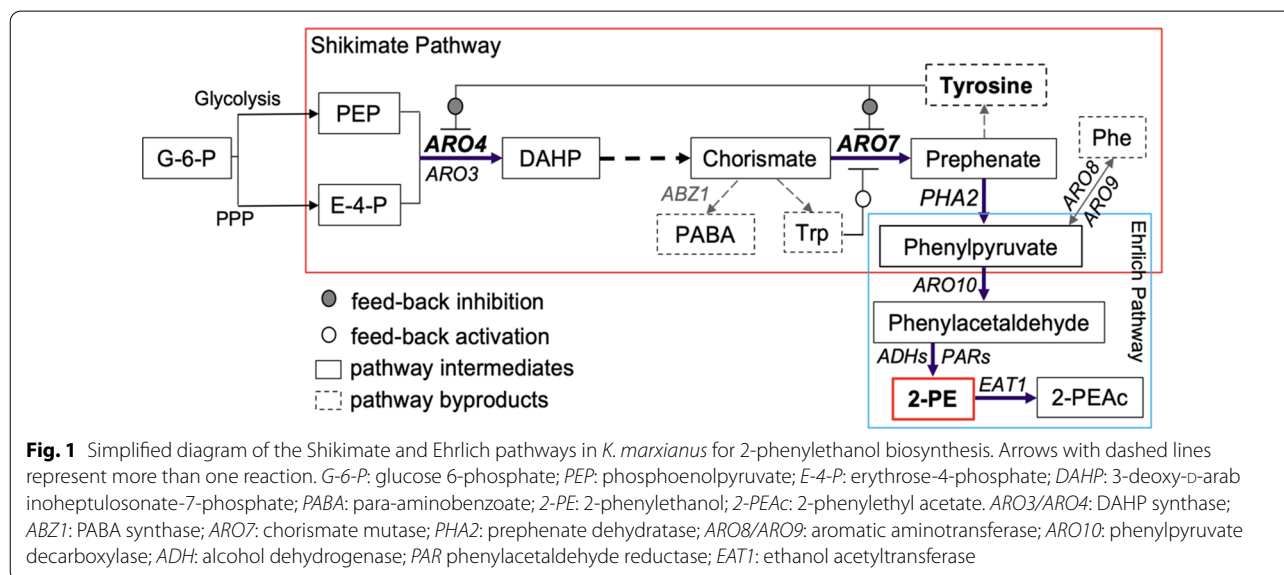
© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

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>) [3]. 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 [4]. 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 [5]. The chemical catalysis route also creates environmental and health challenges due to the reliance on petrochemical reagents [6]. 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; Fig. 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 (ATase or Atf) [7–9]. 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 [10]. 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 [11–13]. Finally, in situ product removal also has been shown to increase 2-PE titers beyond 2 g/L by alleviating product toxicity [14].

Many of the Shikimate and Ehrlich pathway metabolic engineering studies used the common yeast *Saccharomyces cerevisiae* as a production host [15–17], but the pathways have also been engineered in non-conventional yeasts such as *Yarrowia lipolytica*, *Scheffersomyces stipitis*, *Pichia pastoris*, and *Kluyveromyces marxianus* [18–21]. In our own work, we have been studying and engineering *K. marxianus* because wild type strains synthesize the short-chain, volatile ester ethyl acetate in grams per liter quantities, and medium-chain esters such as isoamyl acetate and 2-PEAc are produced in tens of milligram per liter quantities without metabolic engineering [8]. *K. marxianus* is classified as a generally regarded as safe (GRAS) organism, is thermotolerant to ~50 °C, and can metabolize various C5, C6, and C12 sugars [22–25]. Notably, *K. marxianus* is also one of the fastest-growing eukaryotes, with a maximum growth rate nearly twice that of *S. cerevisiae* at 30 °C under glucose limiting conditions [26, 27]. Finally, *K. marxianus* can use



some genetic parts, including promoters and terminators, from *S. cerevisiae*, both of which are part of the saccharomyces subgenera within hemiascomycetes [28, 29]. For all of these reasons, we elected to 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 [30]. Targeted gene disruption, integration, and regulation have been made possible by CRISPR-Cas9 and CRISPR activation/interference (CRISPRa/i) systems for various non-conventional yeasts [31–35]. CRISPR-based genome editing has also been demonstrated in *K. marxianus*, but efficient multigene integration for pathway refactoring [36, 37] has not yet been established [8, 27, 38, 39].

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 3³ 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.

Results

CRISPR-mediated one-step, multigene integration

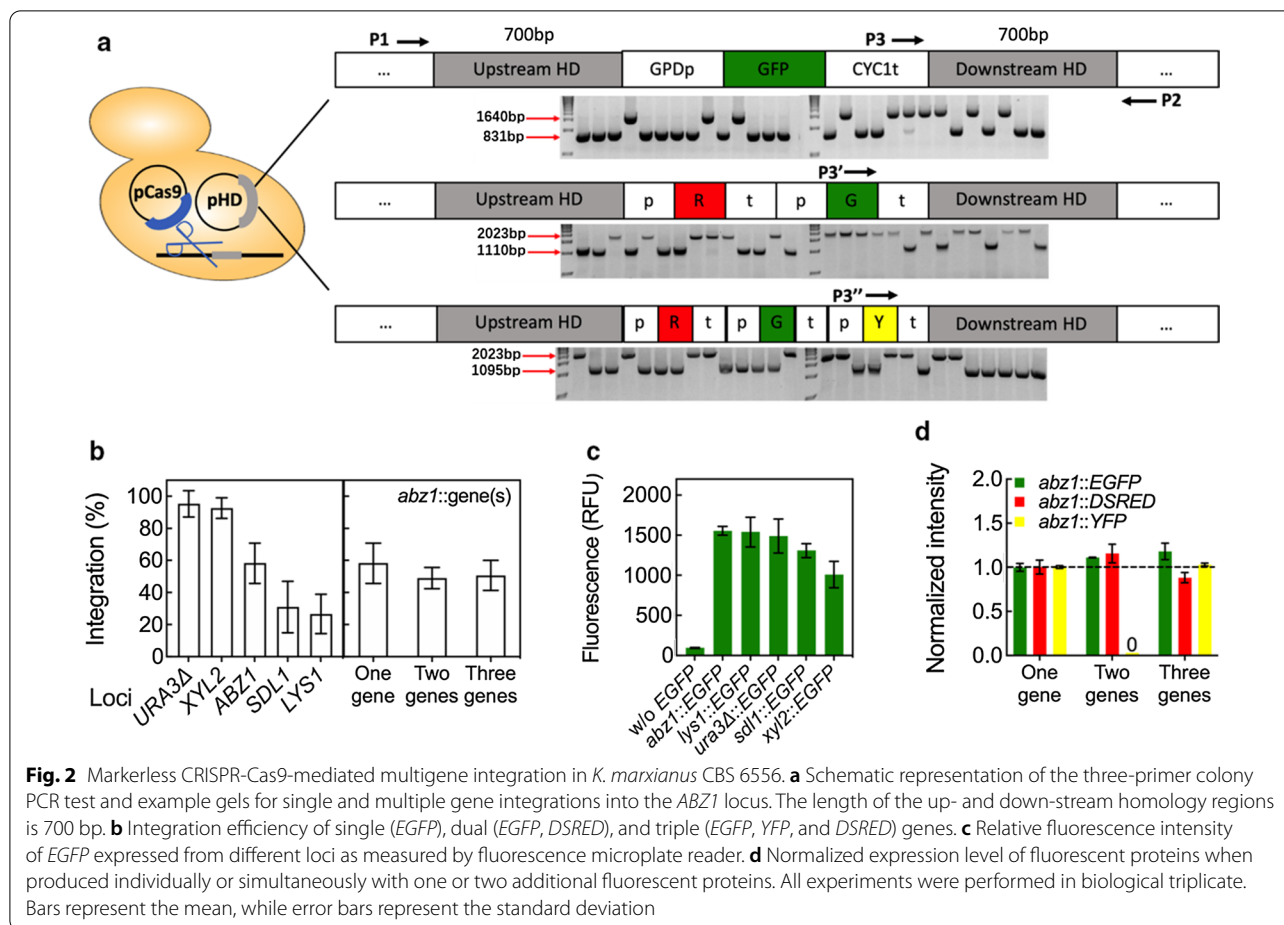
Given the importance of the Shikimate pathway to 2-PE production, we sought to increase flux to phenylpyruvate (PP), the precursor to the Ehrlich pathway, via Shikimate pathway refactoring. To accomplish this in *K. marxianus*, we first needed to identify a series of variable strength promoters, and secondly, create a multigene integration tool. We recently designed and characterized a set of *K. marxianus* 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 [40].

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* [41]. 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 Fig. 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 Additional file 1: Table S1. The *URA3* locus was selected because wild type function was previously eliminated when creating the base strain *K. marxianus* CBS 6556 *ura3Δ his3Δ* [27]. *ABZ1* encodes aminodeoxychorismate synthase, which activates a parasitic reaction diverting chorismate to para-aminobenzoate (PABA; Fig. 1). *LYS1* was selected because it was previously identified as a hotspot for linear DNA integration [24], while *SDL1* is a non-essential gene in *K. marxianus*, and the gene knockout is insensitive to growth at elevated temperature [42]. 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 [27].

Heterologous gene integration at each of the selected loci was tested using fluorescent reporters *EGFP*, *YFP*, and *DSRED* (Fig. 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

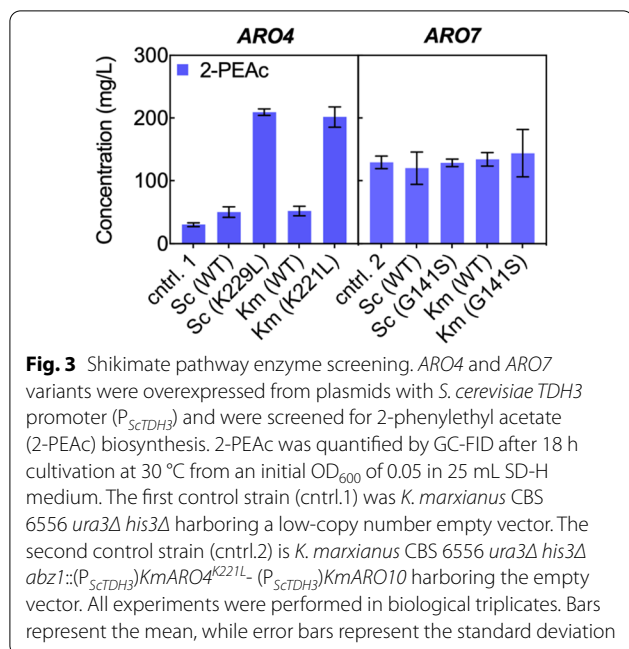


efficiency (Additional file 1: Figure S1); therefore, 700 bp up- and downstream homology was used for all subsequent gene integrations.

Previous reports suggested that heterologous gene expression can differ depending on the integration site within the yeast genome [41, 43]. Figure 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 (Fig. 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.

Alleviation of feedback inhibition in the Shikimate pathway

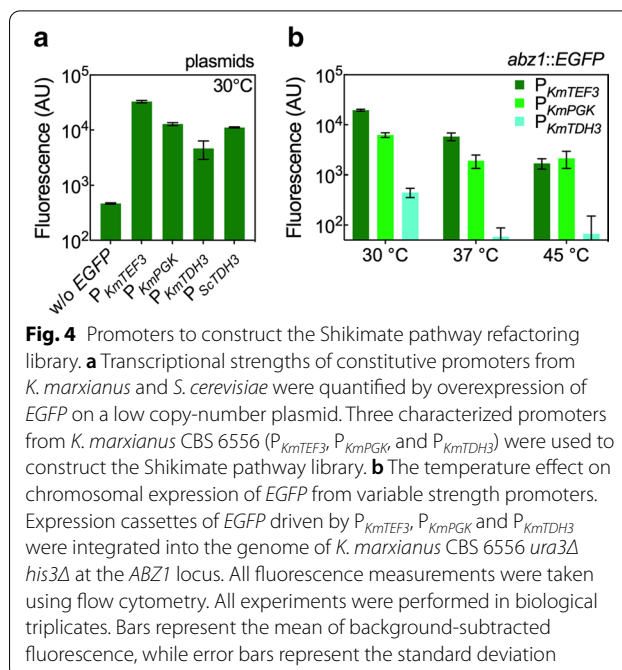
It has been shown that *S. cerevisiae* *ARO4* and *ARO7* are feedback inhibited by tyrosine, thus limiting Shikimate pathway flux. The inhibition effect is eliminated in the feedback insensitive variants *ScARO4*^{K229L} and *ScARO7*^{G141S} [10]. To assess potential feedback inhibition in *K. marxianus*, we overexpressed a series of wild type and mutant *ARO4* and *ARO7* genes from both *S. cerevisiae* BY4742 and *K. marxianus* CBS 6556. In a synthetic defined medium, *K. marxianus* CBS 6556 *ura3Δ his3Δ* accumulated more ester (2-PEAc) than alcohol (2-PE; Additional file 1: 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* (Fig. 3). The *ScARO4* mutation mapped to K221L in *KmARO4*, overexpression of which also produced an increased 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*.

Validation of pathway refactoring promoters

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 [40]. 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 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* [44].

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; Additional file 1: Figure S2). In addition, plasmid expression significantly reduced growth rate at elevated temperatures (Additional file 1: Figure S3). 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 (Fig. 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 (Additional file 1: Figure S4). 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.

Shikimate pathway refactoring in *K. marxianus*

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

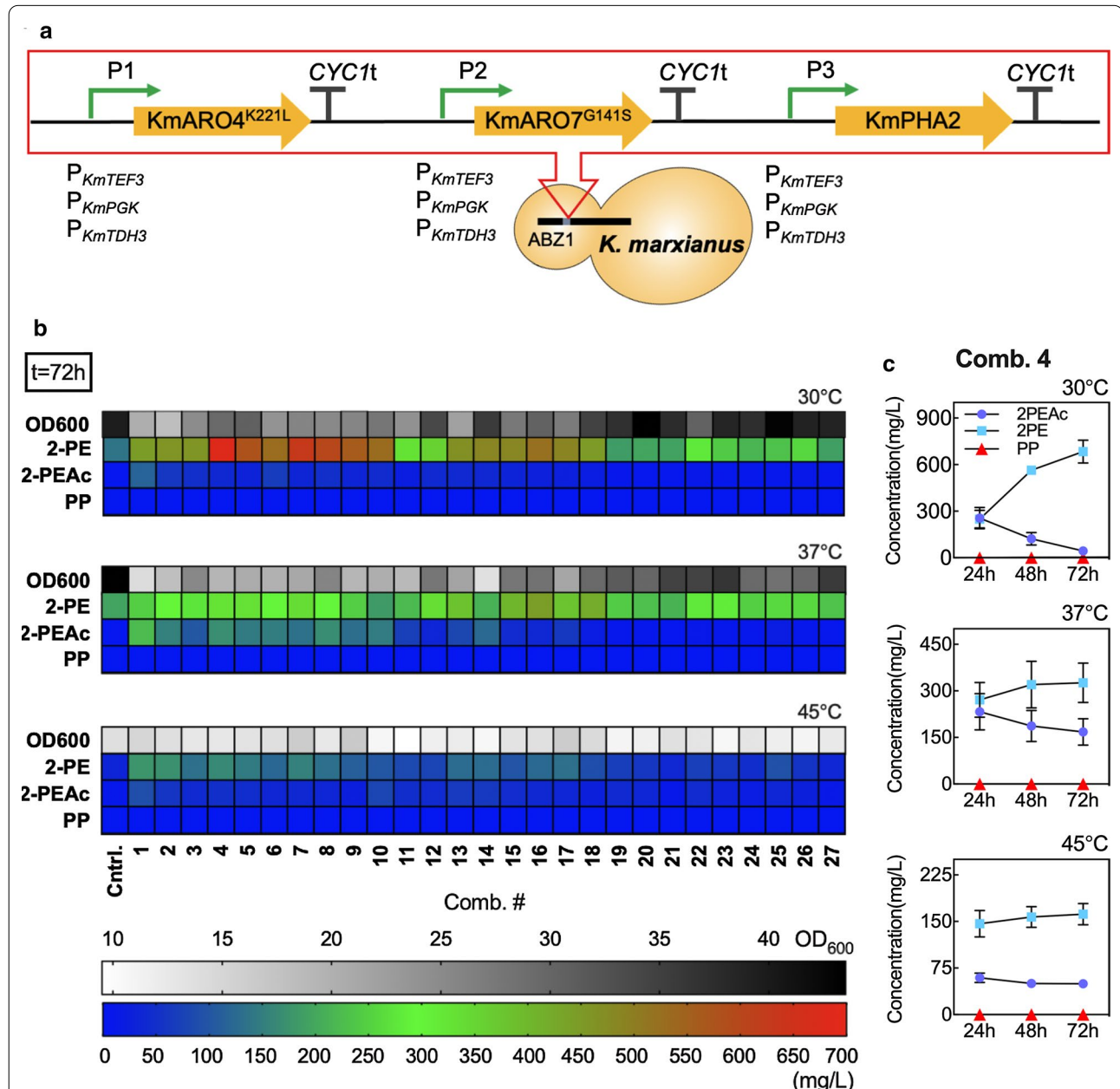


Fig. 5 Shikimate pathway refactoring for enhanced 2-PE biosynthesis. **a** Schematic representation of pathway constructs for the library consisting of all 27 combinations of P_{KmTEF3} , P_{KmPGK} , and P_{KmTDH3} promoters for expression of $KmARO4^{K221L}$, $KmARO7^{G141S}$, and $KmPHA2$. **b** OD₆₀₀, 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

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 Fig. 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 wild type pathway already produces upward of ~150 mg/L of 2-PE, thus providing a strong starting point for pathway engineering.

Figure 5b presents the OD₆₀₀ 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 (Additional file 1: Figure S5). 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 (Fig. 5b, direct comparison shown in Additional file 1: Figure S6). The large effect of *KmARO4^{K221L}* overexpression was observable in the combinations which have a 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 (Additional file 1: Figure S7).

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 OD₆₀₀ after 72 h of cultivation resulted in a Pearson coefficient, *r*, of -0.89 (Additional file 1: Figure S8). A similar correlation with a Pearson coefficient, *r*, of -0.74 was also found between total 2-PE titer (mg/L) and OD₆₀₀.

Given that *K. marxianus* CBS 6556 *ura3Δ his3Δ* exhibited comparable growth rate at 30 and 45 °C (Additional file 1: Figure S3), we used the refactored pathway library to investigate any potential temperature effects on both biomass accumulation and 2-PE biosynthesis. Figure 5b and Additional file 1: Figure S7 show that OD₆₀₀ 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 (Fig. 5b, Additional file 1: Figures S5, and S9). 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 (Additional file 1: Figure S9). 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 (Fig. 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.

Ehrlich pathway engineering and fed-batch operation

To further increase 2-PE biosynthesis, we explored additional mutations to increase Ehrlich pathway flux and reduce ester production (Fig. 6a). We first disrupted *KmARO8*, knockout of which is known in *S. cerevisiae* to upregulate *ARO10* expression and increase Ehrlich pathway flux [45]. 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 the 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 the batch culture at 30 °C (Additional file 1: Figure S10). 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

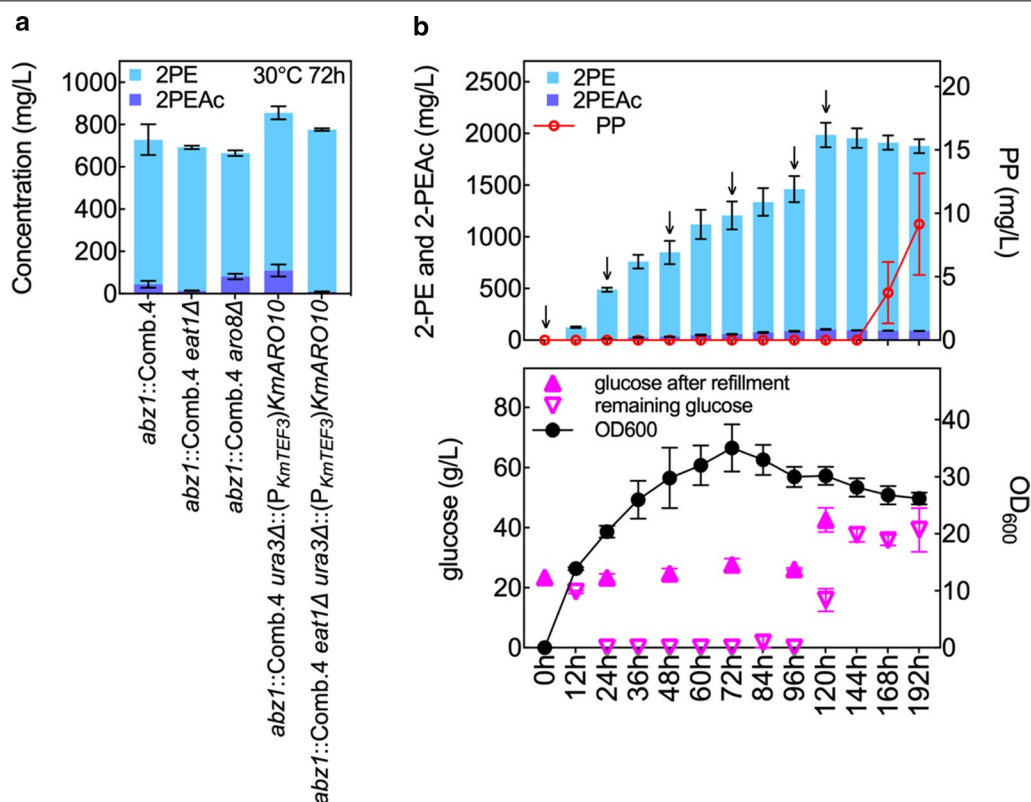


Fig. 6 Ehrlich pathway engineering and cultivation optimization. **a** Screening of Ehrlich pathway mutations for enhanced 2-PE biosynthesis. **b** 2-phenylethanol (2-PE), 2-phenylethyl acetate (2-PEAc), phenylpyruvate (PP) and biomass (OD₆₀₀) accumulation from D-glucose in fed-batch operation of the optimal strain, *K. marxianus* CBS 6556 eat1Δ his3Δ ura3Δ::(P_{KmTEF3})KmARO10, abz1::(P_{KmTEF3})KmARO4^{K221L}-(P_{KmPGK})KmARO7^{G1415}-(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

every 24 h, replenishing the culture to 20 g/L (Fig. 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 endpoint to the process.

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 [30, 46]. 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 [47], random simultaneous integration of multiple genes [48], genome-editing by CRISPR-Cas9 [27, 38, 39], and gene regulation by CRISPRi [8]. Missing from these techniques is the ability to integrate multiple genes into desired sites in the genome at high efficiency. With a CRISPR-Cas9 induced double-stranded break at the target 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. The integration of three genes into the *ABZ1* locus using a single genetic insert 4994 bp in length occurred at an efficiency of 58.3%, statistically similar to the efficiency of integrating a single-gene insert (1730 bp) and a two-gene insert (3343 bp) at the same site (Fig. 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 (Additional file 1: Figure S11). 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 (Figs. 4 and 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 [17–19, 49]. 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 [18]. 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 *KmARO4*, *KmARO7*, and wild type *KmPHA2* (Fig. 5).

Previous work in *S. cerevisiae* suggested that *ARO4* and *ARO7* are negatively regulated by tyrosine, thus reducing flux through the Shikimate pathway [10, 50–52]. A K229L mutation alleviated this feedback 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 [10]. A G141S or T226I mutation to *ScARO7* also relieved the suppression induced by tyrosine [51, 52]. 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 [10]. 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. marxianus* CBS 6556 *ura3Δ his3Δ* increased the 2-PEAc titer from 31 ± 7 to 202 ± 16 mg/L, a nearly sevenfold increase (Fig. 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 (Fig. 3). Overexpressing the *S. cerevisiae* variants in *K. marxianus* produced similar results, that is, alleviating *ARO4* inhibition was beneficial but overexpression of the *ARO7* mutant did not increase pathway

flux. 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 [18]. 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 (Additional file 1: Figure S12), 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 [8, 53]. 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 [27]. 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.

Kluyveromyces marxianus' fast growth and thermotolerance to temperature upward of 45–50 °C are valuable traits beneficial to bioprocess design [46]. As such, we investigated the effect of increased culture temperature on 2-PE biosynthesis using the refactored pathway library. Data presented in Fig. 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_{KmTEF3})*KmPHA2*) at 30 °C (Fig. 5 and Additional file 1: Figure 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 (Fig. 5 and Additional file 1: Figure S5).

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, ~twofold more than that at 30 °C [54]. 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 (Fig. 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 [42, 55]. As for *K. marxianus* CBS 712, the enzymatic activity of xylose reductase and xylitol dehydrogenase was not hampered by temperatures upward to 41 °C [56]. 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 (Fig. 5b and Additional file 1: Figure S7).

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 (Fig. 6b). Similar to other reports, in our long-term cultures we observed that 2-PE was toxic to *K. marxianus* [14, 57]. 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 [14]. 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 the 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 the enhancement of 2-PE biosynthesis [17, 19]. In *Y.*

lipolytica, overexpression of the homologous transketolase γ TKT and heterologous phosphoketolases BbxfpK and AcxpkA, which aimed to increase E4P formation from the pentose phosphate pathway, also increased the 2-PE production [19]. 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 [17].

Conclusions

The Shikimate pathway can be a productive source of aromatic metabolites useful in various applications, including stilbenoids and flavonoids used for pharmaceuticals and nutraceuticals. Here, we created a library of strains with refactored Shikimate pathway to explore the overproduction of 2-PE, but the same library could also be used to optimize the production of other aromatics. In addition, the multigene integration strategy created here expands the metabolic engineering toolbox for *K. marxianus*, a stress-tolerant yeast that is a promising host for the industrial production of biochemicals.

Methods

Molecular cloning and reagents

All molecular cloning reagents and enzymes were purchased from NEB. Plasmids created and used in this work are shown in Additional file 1: Table S2. 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).

***K. marxianus* strains, media, and cultivation**

Kluyveromyces 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. 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. 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.

CRISPR-Cas9 gene disruption and integration plasmids

CRISPR plasmids (pCRISPR) were constructed using pIW601 (Addgene ID 98907) linearized at PspXI and reassembled 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>) [58] and the uniqueness across the *K. marxianus* genome using a CRISPR/Cas9 target online predictor, CCTop (<https://crispr.cos.uni-heidelberg.de/>) [59].

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.

Enzyme screening and Shikimate pathway refactoring plasmids

All plasmids used for enzyme screening were constructed by replacing EGFP in pIW1135 with the new gene (Table S2). 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 gene. 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 *ABZI* locus. 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} [40] 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 *ABZI* flanking a three-gene overexpression insert. Integration was accomplished as described below.

***K. marxianus* transformation**

Kluyveromyces marxianus transformations were accomplished using a modified protocol described in [27]. 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 5000g for 1 min in a sterile 1.5 mL Eppendorf tube. After washing twice with 1 mL of sterile ddH₂O, 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 diH₂O in each well. Cells were then harvested by centrifugation at 5000g 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 inoculation in to the appropriate selective liquid medium or by plating directly on selective solid agar medium as described in the following section.

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 S11). 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 Fig. 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; the binding location of the upstream genomic primer was adjusted on purpose to maintain a band size difference between positive and negative results. For single-gene integration, in the absence of integration, amplification occurred with the two genomic primers to produce a 1640 bp amplicon, while successful gene integrations produced an 831 bp amplicon. Positive multigene integration produced a ~1110 bp amplicon, while the negative result produced a 2023 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 the 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).

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 expressing different fluorescent proteins. 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 tenfold 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).

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 5000g 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 ddH₂O. Two-microliter of the cell solutions was further diluted in 100 µL ddH₂O 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 Fig. 4 are shown in Figures S14-S17. 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.

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 5000g, 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 10,000g 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 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.

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 (Additional file 1: Figure S13).

Glucose assay

The concentration of residual glucose was quantified using a commercially available glucose assay kit (GAGO-20; Sigma-Aldrich). 200 µL of the 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 H₂SO₄. 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 ddH₂O 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 (Additional file 1: Figure S18). For sample preparation, 500 µL cell cultures were centrifuged down at 5000g for 1 min 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 ddH₂O. 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”.

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 the 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.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13068-020-01852-3>.

Additional file 1. Additional Figures and Tables.

Abbreviations

2-PE: 2-Phenylethanol; CRISPR: Clustered regularly interspaced short palindromic repeats; Cas9: CRISPR-associated protein 9; E-4-P: Erythrose-4-phosphate; PEP: Phosphoenolpyruvate; PP: Phenylpyruvate; Adh: Alcohol dehydrogenase; 2-PEAc: 2-Phenylethyl acetate; ATase or Atf: Alcohol acetyltransferase; PAR: Phenylacetaldehyde reductase; GRAS: Generally regarded as safe; NHEJ: Non-homologous end joining; HR: Homologous recombination; CRISPR a/i: CRISPR activation and interference; G-6-P: Glucose 6-phosphate; DAHP: 3-Deoxy-D-arabionoheptulosonate-7-phosphate; PABA: Para-aminobenzoate; Trp: Tryptophan; Phe: Phenylalanine; ARO3/ARO4: DAHP synthase; ARO7: Chorismate mutase; PHA2: Prephenate dehydratase; ARO8/ARO9: Aromatic aminotransferase; ARO10: Phenylpyruvate decarboxylase; EAT1: Ethanol acetyltransferase; sgRNA: Single guide RNA; EGFP: Enhanced green fluorescence protein; YFP: Yellow fluorescence protein; DSRED: Red fluorescence protein isolated from *Discosoma*; ANOVA: Analysis of variance; OD: Optical density; TAL: Triacetic acid lactone; BLAST: Basic local alignment search tool; 5-FOA: 5-Fluoroorotic acid; EDTA: Ethylenediaminetetraacetic acid disodium salt dihydrate; LB: Luria-Bertani; GC: Gas chromatography; HPLC: High-performance liquid chromatography; SD: Synthetic defined; DTT: Dithiothreitol; PEG: Polyethylene glycol; Tris-HCl: Tris(hydroxymethyl)aminomethane hydrochloride; TE: Tris-EDTA; Ex/Em: Excitation/emission wavelength in nm; FID: Flame ionization detector; RT-qPCR: Reverse transcription quantitative PCR; GAPDH: Glyceraldehyde 3-phosphate dehydrogenase.

Acknowledgements

We thank the Department of Energy and US National Science Foundation for funding to support this work.

Authors' contributions

ML, XL, and IW conceived the study and designed the experiments. ML and XL designed and characterized the promoters. ML conducted the pathway refactoring and metabolic engineering experiments with the assistance of XL, MMC, SDK. XS analyzed media effects. ML, XL, IW, and NAD wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by DOE DE-SC0019093, NSF 1803630 and NSF 180367.

Availability of data and materials

All data generated and analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All author consent to publish this work.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Chemical and Environmental Engineering, University of California Riverside, Riverside, CA 92521, USA. ² Department of Chemical and Biomolecular Engineering, University of California Irvine, Irvine, CA 92697, USA. ³ Center for Industrial Biotechnology, University of California Riverside, Riverside, CA 92527, USA.

Received: 14 September 2020 Accepted: 9 December 2020

Published online: 06 January 2021

References

- Fabre CE, Blanc PJ, Goma G. 2-Phenylethyl alcohol: an aroma profile. *Perfum Flavor*. 1998;23:43–5.
- Keasling JD, Chou H. Metabolic engineering delivers next-generation biofuels. *Nat Biotechnol*. 2008;26(3):298–9.
- <https://www.grandviewresearch.com/industry-analysis/flavors-fragrances-market>. Accessed 3 Dec 2020.
- van Summeren-Wesenhagen PV, Marienhagen J. Putting bugs to the blush: metabolic engineering for phenylpropanoid-derived products in microorganisms. *Bioengineered*. 2013;4(6):355–62.
- Etschmann M, Bluemke W, Sell D, Schrader J. Biotechnological production of 2-phenylethanol. *Appl Microbiol Biotechnol*. 2002;59(1):1–8.
- Krings U, Berger RG. Biotechnological production of flavours and fragrances. *Appl Microbiol Biotechnol*. 1998;49(1):1–8.
- Cao M, Gao M, Suástegui M, Mei Y, Shao Z. 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.
- Löbs AK, Schwartz C, Thorwall S, Wheelton I. Highly multiplexed CRISPRi repression of respiratory functions enhances mitochondrial localized ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *ACS Synth Biol*. 2018;7(11):2647–55.
- Lin JL, Wheelton 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.
- Luttik MAH, Vuralhan Z, Suij E, Braus GH, Pronk JT, Daran JMG. Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: quantification of metabolic impact. *Metab Eng*. 2008;10(3):141–53.
- Shen L, Nishimura Y, Matsuda F, Ishii J, Kondo A. 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):34–9.
- Guo D, Zhang L, Kong S, Liu Z, Li X, Pan H. Metabolic engineering of *Escherichia coli* for production of 2-phenylethanol and 2-phenylethyl acetate from glucose. *J Agric Food Chem*. 2018;66(23):5886–91.
- Chen XM, Kobayashi H, Sakai M, Hirata H, Asai T, Ohnishi T, Baldermann S, Watanabe N. 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.
- Gao F, 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):332–9.
- Rodriguez A, Kildegaard KR, Li M, Borodina I, Nielsen J. Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. *Metab Eng*. 2015;31:181–8.
- Suástegui M, Yu Ng C, Chowdhury A, Sun W, Cao M, House E, Maranas CD, Shao Z. Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in *Saccharomyces cerevisiae* for high production of polymer and drug precursors. *Metab Eng*. 2017;42:134–44.
- Hassing EJ, de Groot PA, Marquenie VR, Pronk JT, Daran JMG. Connecting central carbon and aromatic amino acid metabolisms to improve de novo 2-phenylethanol production in *Saccharomyces cerevisiae*. *Metab Eng*. 2019;56:165–80.
- Kim TY, Lee SW, Oh MK. Biosynthesis of 2-phenylethanol from glucose with genetically engineered *Kluyveromyces marxianus*. *Enzyme Microb Technol*. 2014;61–62:44–7.
- Gu Y, Ma J, Zhu Y, Ding X, Xu P. Engineering *Yarrowia lipolytica* as a chassis for de novo synthesis of five aromatic-derived natural products and chemicals. *ACS Synth Biol*. 2020;9(8):2096–106.
- Gao M, Cao M, Suástegui M, Walker J, Rodriguez Quiroz N, Wu Y, Tribby D, Okerlund A, Stanley L, Shanks JV, Shao Z. Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics. *ACS Synth Biol*. 2017;6(1):29–38.
- Kong S, Pan H, Liu X, Li X, Guo D. De novo biosynthesis of 2-phenylethanol in engineered *Pichia pastoris*. *Enzyme Microb Technol*. 2020;133:109459.
- Abdel-Banat BMA, Hoshida H, Ano A, Nonklang S, Akada R. 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):861–7.
23. Löbs AK, Lin JL, Cook M, Wheeldon I. 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):1274–81.
 24. Nonklang S, Abdel-Banat BMA, Cha-aim K, Moonjai N, Hoshida H, Limtong S, Yamada M, Akada R. High-temperature ethanol fermentation and transformation with linear DNA in the thermotolerant yeast *Kluyveromyces marxianus* DMKU3-1042. *Appl Environ Microbiol*. 2008;74(24):7514.
 25. Medeiros ABP, Pandey A, Christen P, Fontoura PSG, de Freitas RJS, Soccol CR. Aroma compounds produced by *Kluyveromyces marxianus* in solid state fermentation on a packed bed column bioreactor. *World J Microbiol Biotechnol*. 2001;17(8):767–71.
 26. Groeneveld P, Stouthamer AH, Westerhoff HV. Super life—how and why ‘cell selection’ leads to the fastest-growing eukaryote. *FEBS J*. 2009;276(1):254–70.
 27. Löbs AK, Engel R, Schwartz C, Flores A, Wheeldon I. CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *Biotechnol Biofuels*. 2017;10(1):164.
 28. Belloch C, Querol A, García MD, Barrio E. Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome-c oxidase II gene. *Int J Syst Evol Microbiol*. 2000;50(1):405–16.
 29. Lane MM, Morrissey JP. *Kluyveromyces marxianus*: a yeast emerging from its sister’s shadow. *Fungal Biol Rev*. 2010;24(1):17–26.
 30. Löbs AK, Schwartz C, Wheeldon I. Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. *Synth Syst Biotechnol*. 2017;2(3):198–207.
 31. Schwartz C, Cheng JF, Evans R, Schwartz CA, Wagner JM, Anglin S, Beitz A, Pan W, Lonardi S, Blenner M, Alper HS, Yoshikuni Y, Wheeldon I. Validating genome-wide CRISPR–Cas9 function improves screening in the oleaginous yeast *Yarrowia lipolytica*. *Metab Eng*. 2019;55:102–10.
 32. Schwartz C, Curtis N, Löbs AK, Wheeldon I. Multiplexed CRISPR activation of cryptic sugar metabolism enables *Yarrowia lipolytica* growth on cellobiose. *Biotechnol J*. 2018;13(9):1700584.
 33. Cao M, Fatma Z, Song X, Hsieh PH, Tran VG, Lyon WL, Sayadi M, Shao Z, Yoshikuni Y, Zhao H. A genetic toolbox for metabolic engineering of *Issatchenkia orientalis*. *Metab Eng*. 2020;59:87–97.
 34. Cao M, Gao M, Ploessl D, Song C, Shao Z. CRISPR–mediated genome editing and gene repression in *Scheffersomyces stipitis*. *Biotechnol J*. 2018;13(9):1700598.
 35. Dalvie NC, Leal J, Whittaker CA, Yang Y, Brady JR, Love KR, Love JC. Host-informed expression of CRISPR guide RNA for genomic engineering in *Komagataella phaffii*. *ACS Synth Biol*. 2020;9(1):26–35.
 36. Young EM, Zhao Z, Gielesen BEM, Wu L, Gordon DB, Roubos JA, Voigt CA. Iterative algorithm-guided design of massive strain libraries, applied to itaconic acid production in yeast. *Metab Eng*. 2018;48:33–43.
 37. Babaei M, Borja Zamfir GM, Chen X, Christensen HB, Kristensen M, Nielsen J, Borodina I. Metabolic engineering of *Saccharomyces cerevisiae* for rosmarinic acid production. *ACS Synth Biol*. 2020;9(8):1978–88.
 38. Nambu-Nishida Y, Nishida K, Hasunuma T, Kondo A. 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.
 39. Rajkumar AS, Varela JA, Juergens H, Daran JMG, Morrissey JP. Biological parts for *Kluyveromyces marxianus* synthetic biology. *Front Bioeng Biotechnol*. 2019;7:97.
 40. Lang X, Besada-Lombana PB, Li M, Da Silva NA, Wheeldon I. Developing a broad-range promoter set for metabolic engineering in the thermotolerant yeast *Kluyveromyces marxianus*. *Metab Eng Commun*. 2020;11:e00145.
 41. Schwartz C, Shabbir-Hussain M, Frogue K, Blenner M, Wheeldon I. Standardized markerless gene integration for pathway engineering in *Yarrowia lipolytica*. *ACS Synth Biol*. 2017;6(3):402–9.
 42. Fu X, Li P, Zhang L, Li S. 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):2715–29.
 43. Wu XL, Li BZ, Zhang WZ, Song K, Qi H, Dai J, Yuan YJ. Genome-wide landscape of position effects on heterogeneous gene expression in *Saccharomyces cerevisiae*. *Biotechnol Biofuels*. 2017;10(1):189.
 44. Lee KS, Kim JS, Heo P, Yang TJ, Sung YJ, Cheon Y, Koo HM, Yu BJ, Seo JH, Jin YS, Park JC, Kweon DH. Characterization of *Saccharomyces cerevisiae* promoters for heterologous gene expression in *Kluyveromyces marxianus*. *Appl Microbiol Biotechnol*. 2013;97(5):2029–41.
 45. Romagnoli G, Knijnenburg TA, Liti G, Louis EJ, Pronk JT, Daran JMG. Deletion of the *Saccharomyces cerevisiae* ARO8 gene, encoding an aromatic amino acid transaminase, enhances phenylethanol production from glucose. *Yeast*. 2015;32(1):29–45.
 46. Thorwall S, Schwartz C, Chartron JW, Wheeldon I. Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. *Nat Chem Biol*. 2020;16(2):113–21.
 47. Pecota DC, Rajgarhia V, Da Silva NA. Sequential gene integration for the engineering of *Kluyveromyces marxianus*. *J Biotechnol*. 2007;127(3):408–16.
 48. Heo P, Yang TJ, Chung SC, Cheon Y, Kim JS, Park JB, Koo HM, Cho KM, Seo JH, Park JC, Kweon DH. Simultaneous integration of multiple genes into the *Kluyveromyces marxianus* chromosome. *J Biotechnol*. 2013;167(3):323–5.
 49. Sekar BS, Lukito BR, Li Z. Production of natural 2-phenylethanol from glucose or glycerol with coupled *Escherichia coli* strains expressing L-phenylalanine biosynthesis pathway and artificial biocascades. *ACS Sustain Chem Eng*. 2019;7(14):12231–9.
 50. Hartmann M, Schneider TR, Pfeil A, Heinrich G, Lipscomb WN, Braus GH. Evolution of feedback-inhibited B/a barrel isoenzymes by gene duplication and a single mutation. *Proc Natl Acad Sci USA*. 2003;100(3):862.
 51. Schnappauf G, Krappmann S, Braus GH. Tyrosine and tryptophan act through the same binding site at the dimer interface of yeast chorismate mutase. *J Biol Chem*. 1998;273(27):17012–7.
 52. Krappmann S, Lipscomb WN, Braus GH. Coevolution of transcriptional and allosteric regulation at the chorismate metabolic branch point of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*. 2000;97(25):13585–90.
 53. Kruis AJ, Levisson M, Mars AE, van der Ploeg M, Garcés Daza F, Ellena V, Kengen SWM, van der Oost J, Weusthuis RA. Ethyl acetate production by the elusive alcohol acetyltransferase from yeast. *Metab Eng*. 2017;41:92–101.
 54. McTaggart TL, Bever D, Bassett S, Da Silva NA. Synthesis of polyketides from low cost substrates by the thermotolerant yeast *Kluyveromyces marxianus*. *Biotechnol Bioeng*. 2019;116(7):1721–30.
 55. Lertwattanasakul N, Kosaka T, Hosoyama A, Suzuki Y, Rodrussamee N, Matsutani M, Murata M, Fujimoto N, Tsuchikane K, Limtong S, Fujita N, Yamada M. Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. *Biotechnol Biofuels*. 2015;8(1):47.
 56. Signori L, Passolunghi S, Ruohonen L, Porro D, Branduardi P. Effect of oxygenation and temperature on glucose-xylose fermentation in *Kluyveromyces marxianus* CBS712 strain. *Microb Cell Fact*. 2014;13(1):51.
 57. Fabre CE, Blanc PJ, Goma G. Production of 2-phenylethyl alcohol by *Kluyveromyces marxianus*. *Biotechnol Prog*. 1998;14(2):270–4.
 58. <https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>. Accessed 3 Dec 2020
 59. <https://crispr.cos.uni-heidelberg.de/>. Accessed 3 Dec 2020.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.