Ollscoil na hÉireann, Corcaigh

The National University of Ireland, Cork University College Cork

School of Microbiology

Head of School: Dr David Clarke,

Supervisor: Professor John P. Morrissey

Development of synthetic biology tools for *Kluyveromyces marxianus* and their application to produce aromatic molecules in this yeast

Thesis presented by

Joel Abidemi Akinola

for the degree of

Doctor of Philosophy

In

Microbiology

July 2023

Table of Contents

Declaration

I, Joel Abidemi Akinola, declare that I have not obtained a degree from University College Cork, National University of Ireland, Cork or elsewhere on the basis of this Ph.D. thesis and that the results presented in this thesis were derived from experiments undertaken by me. Most of the research presented in this thesis was carried out at University College Cork but some data were generated by me at Delft University of Technology, The Netherlands.

Acknowledgments

I thank my creator for giving me the opportunity to complete this project with a sound mind and health. For guiding me through these challenging times and for surrounding me with people that constantly edify me.

Very importantly, I am eternally grateful to Professor John Morrissey for giving me the opportunity to pursue this PhD under his supervision, for his guidance and for his patience throughout this project.

I would like to thank my family for their support throughout this project. I thank you Mum and Dad for your prayers and for believing in me. I especially thank my wife Khadija Mu'azzam for being my number one support system. Khadija, I thank you for holding the fort in all aspects of our lives during the final months of this project, for listening to my worries, and for consoling me during times of disappointments. You are my rock!

I would like to acknowledge the financial support of the CHASSY project, funded by the European Union and the Lilly Research Scholarship at UCC. I would also like to thank all the partners within the project who contributed with ideas and suggestions to my research throughout the years, especially Professor Jean-Marc Daran who hosted me as a visiting student at Delft University of Technology, The Netherlands.

Recite in the name of your Lord who created –

Created man from a clinging substance.

Recite, and your Lord is the most Generous –

Who taught by the pen –

Taught man that which he knew not.

Quran 96:1-5

Abstract

The experimental work in this thesis was carried out as part of the European Union-funded Horizon 2020 project CHASSY (https://www.ucc.ie/en/eri/projects/chassy.html). CHASSY was a collaborative research project made up of industry and academic partners who set out to unlock the potential of yeasts as microbial cell factories for producing valuable compounds. While *Saccharomyces cerevisiae* is the most studied yeast for which there are many genetic engineering tools available and is thus commonly used for synthetic biology (SynBio) applications, it has certain limitations for specific applications. This has in recent times led to the exploration of the industrially attractive non-saccharomyces yeast, *Kluyveromyces marxianus* for various synthetic biology applications. This yeast has a QPS/GRAS status that drives its application in the food industry, it is thermotolerant, able to metabolise and grow on several feedstocks and SynBio tools for achieving heterologous gene expression and genome engineering by CRISPR technologies have been developed for it. Those recent advancements have contributed to research on this yeast in regard to improving our understanding of its physiology and for SynBio applications. To further improve genome editing tools for the yeast, a counter-selectable amidase (*amdS*) gene from bacteria was developed in this thesis for marker-dependent integration. Close to 100% of the transformants obtained had an *amdS* marker cassette integrated when the cassette was transformed. A CRISPR-based marker-free system was also developed and optimised for simultaneous editing of more than one locus in the yeast's genome. This system is highly useful because it eliminates genetic engineering restrictions that arise as a result of the need for markers, and it was made possible thanks to the yeast's homologous recombination machinery which allows homology directed repair (HDR). Thus, the system was optimised by improving HDR through the inactivation of non-homologous end-joining, the expression of recombinases and chemical synchronisation of yeast cells into the S/G2 phase of cell cycle. The optimisation allowed efficient marker-free deletion of specific genes, integration at specific loci and multi-loci integration at repeat sequences that are distributed around the yeast's genome.

The developed tools in this work are entirely based on the well-established yeast toolkit standard which uses Golden Gate cloning technique for DNA assembly, making it possible to use them in combination with pre-existing tools and they are transferable between SynBio applications. Therefore, the tools were used for unprecedented engineering of *K. marxianus* for heterologous production of a group of plant aromatic compounds called phenylpropanoids. The biosynthesis of these compounds takes the aromatic amino acids (AAA), phenylalanine, and tyrosine as precursors. Therefore, the key enzymes of the shikimate pathway which is responsible for AAA biosynthesis and the glycolytic and pentose phosphate pathways which

supply phosphoenolpyruvate and erythrose 4-phosphate as precursors of the shikimate pathway were overexpressed in order to increase AAA production. The phenylpropanoid pathway begins with the conversion of tyrosine to coumaric acid by means of a single step reaction catalysed by tyrosine ammonia lyase or phenylalanine in two reaction steps where phenylalanine ammonia lyase first converts phenylalanine to cinnamic acid and then cinnamic acid 4-hydroxylase and cytochrome P450 reductase convert cinnamic acid into coumaric acid. Given that coumaric acid is an important phenylpropanoid from which many other phenylpropanoids are produced, its production was first engineered into *K. marxianus* in with the expectation that it will be possible to further engineer coumaric acid producers to produce more complex phenylpropanoids. This work resulted in the production of coumaric acid at a yield of over 40mg/g glucose in 5L batch bioreactors with 2.5L working volumes of minimal glucose medium. Subsequently, plant genes that catalyse the derivatisation of coumaric acid into an important flavonoid called naringenin (coumaric acid:coenzyme A ligase, chalcone synthase and chalcone isomerase) were expressed in a coumaric acid producer. This however, only produced very low amounts of naringenin with minor decrease in coumaric acid yield, giving rise to a study in which the depletion of coumaric acid was the objective. Since the derivatisation requires coenzyme A and malonyl-CoA, increased biosynthesis of both metabolites was engineered by overexpressing acetyl-CoA carboxylase, acetyl-CoA synthase, and the entire coenzyme A biosynthetic pathway, leading to naringenin yield of 1231µg/g glucose. Due to the fact that freely available intracellular AAAs are catabolised by the Ehrlich pathway in yeasts, a study was conducted to identify the genes that contribute *K. marxianus* phenylpyruvate decarboxylase activity, a key player that is responsible for driving flux through the pathway. That study revealed that *ARO10* contributed most of the enzyme activity with some coming from *PDC1* and provided information that enabled the reduction of the activity so that AAAs are available to enter the heterologous phenylpropanoid pathway. Together, this work demonstrates the potential of *K. marxianus* as a cell factory for producing phenylpropanoids, bringing us closer to the production of these compounds from diverse feedstocks.

Dissemination of materials from this thesis

1. Manuscripts in preparation for submission

Joel A. Akinola, John P. Morrissey. Synthetic biology of aromatic molecules: engineering phenylpropanoid biosynthesis in yeast cell.

This work is presented as chapter 1 in this thesis.

Joel A. Akinola, Arun S. Rajkumar, John P. Morrissey. Optimisation of coumaric acid production from phenylpyruvate in *Kluyveromyces marxianus.*

This work is presented as chapter 3 in this thesis.

Joel A. Akinola, Andrea Rizzo, Jean-Marc Daran, John P. Morrissey. Enzymatic and functional characterisations reveal the contribution of *Kluyveromyces marxianus* 2-oxo acid decarboxylase genes to glycolytic and Ehrlich pathways in the yeast.

This work is presented as chapter 4 in this thesis.

2. Oral presentations

Joel A. Akinola and John P. Morrissey. Production and derivatisation of naringenin in yeast. Elli Lilly, Cork, Ireland (26th June 2019).

Joel A. Akinola, Arun S. Rajkumar, John P. Morrissey. Engineering the yeast *Kluyveromyces marxianus* for production of aromatic molecules. DECHEMA Bioflavour conference: Biotechnology of Flavours, Fragrances and Functional Ingredients. Frankfurt, Germany (27th – 30th September 2022).

3. Poster presentations

Joel A. Akinola, Javier A. Varela, Arun S. Rajkumar, John P. Morrissey. Development of multiplex genome editing tools for *Kluyveromyces marxianus.* YEASTDOC Doctorial School in Yeast Industrial Biotechnology. Berlin, Germany (23rd - 27th September 2019).

Else-Jasmijn Hassing, Arun S. Rajkumar, Macarena Larroude, **Joel A. Akinola**, Philip A. de Groot1, Mario Beck, Vita Marquenie, Jack T. Pronk, John P. Morrissey and Jean-Marc G.

Daran. Engineering Yeasts to Produce Aromatics. European Summit of Industrial Biotechnology. Graz, Austria (18th - 20th November 2019).

Joel A. Akinola, Arun S. Rajkumar, John P. Morrissey. Development of *K. marxianus* for commercial production of flavonoids. 6th Applied Synthetic Biology in Europe. Edinburgh, Scotland (2nd- 4th November 2022).

Chapter 1. Synthetic biology of aromatic molecules: engineering phenylpropanoid biosynthesis in yeast cell factories

Joel A. Akinola, John P. Morrissey.

This chapter is in preparation for submission

JAA prepared the manuscript

JPM edited the manuscript

Abstract

Aromatic molecules are highly sought after compounds because they are useful in various aspects of life. Some are used as food additives and the bioactive nature of many makes them applicable as active pharmaceutical ingredients or as key ingredients in functional foods. Their complex structure in many cases can make production through chemical synthesis not feasible, meaning that extraction from plant materials is their only source. However, this comes with various challenges from the fact that they are produced in minute quantities to difficulties in downstream processing. Metabolic engineering through synthetic biology approaches is currently an active area of research that focuses on developing cell factories that can eventually be used for commercial scale production of aromatics. This review begins by giving an overview of the field of synthetic biology, illustrating how it is used to develop cell factories by means of the design-build-test-learn cycle approach. As the biological precursors of many aromatics, the biosynthesis of aromatic amino acids in plants and microorganisms were assessed to identify variations in these hosts. The suitability of yeasts as microorganisms for producing aromatics was discussed and the engineering strategies that have been established so far to produce aromatics and their precursors in yeasts were elaborated on, giving specific examples from the literature.

1. Introduction

Aromatic molecules (aromatics) are highly useful chemicals that have applications in many areas of our lives. They are used as food additives, bioactive ingredients in functional foods, active pharmaceutical ingredients, and in the manufacturing of household products. With these molecules currently a multi-billion dollars valued market globally and projected to have a compound annual growth rate (CAGR) of 5.72% between 2022 to 2029 according to Data Bridge Market Research (https://www.databridgemarketresearch.com/reports/global-aromatic-compounds-market), production through chemical synthesis, for those aromatics that can be produced by this means, is bound to increase. Chemical synthesis is however plagued by the difficult multiple steps that are required to produce complex aromatic molecules, the requirement for toxic chemicals and production of toxic byproducts (Milke *et al.*, 2018). Given that a lot of aromatics are naturally biosynthesised in plants, some valuable ones can be alternatively extracted from plant materials, but this comes with its own challenges such as difficult downstream processing involved in obtaining the specific aromatics desired because plants produce several of these molecules with highly similar physicochemical properties. Additionally, because aromatics are secondary metabolites in plants and are consequently produced in small amounts, commercial extraction requires large amounts of plant materials and there can be variability in yields and product quality (Liu

and Nielsen, 2019). Those production approaches are not sustainable and the global push for sustainable development has resulted in active endeavours by the scientific community to direct attention toward sustainable manufacturing. One major approach to achieving sustainable manufacturing of aromatics is synthetic biology (SynBio) in which host cells are genetically modified into cell factories for production. Although a major advantage of production in cell factories is the broadening feedstock options in contrast to chemical synthesis, which uses strict petroleum-derived precursors (Liu *et al.*, 2020), the adoption of this production approach is hindered by the difficulty of attaining cost competitiveness compared to chemical synthesis. Efforts to overcome this major hurdle by the community can be summarised into (i) developing microbial cell factories that can efficiently convert starting materials into the desired product through synthetic biology approaches, and (ii) reducing the cost of starting materials.

1.1. Synthetic biology approaches for designing cell factories

It has been proposed that generally the actual yield should be 85% of the theoretical yield (Peralta-Yahya *et al.*, 2012) and that titre should exceed 50g/L with specific productivity above 2g/L/h (Woodley, 2017) for a bioprocess to be commercialised. These are stringent performance requirements that are difficult to achieve in cell factories, especially for producing molecules that are foreign to the host organism. Some aromatics are bioactive and could be potentially toxic to the organism among other unknown effects. Towards the developments of cell factories that meet these criteria, knowledge of the physiology of a microorganism is taken into consideration before it is selected as a suitable host for development into cell factories commonly by a top-down approach in which metabolic and genetic engineering techniques are used to establish heterologous biosynthetic pathways in the host (Elani, 2021). Alternatively, the bottom-up approach involving the construction of artificial cell-like entities from inanimate matter can be used in cell factories development, but this approach by itself has only seen moderate advancements compared to the top-down approach mainly due our limited understanding of the absolute requirements for life (Mutschler *et al.*, 2019; Elani, 2021). Although the presence of naturally existing pathways that can be tapped into to benefit specific SynBio applications is a major advantage of the top-down approach, this approach is paradoxically challenged by the fact that the main goal of cell physiology is biomass formation and growth. Thus, metabolic engineering, facilitated by the widely adopted design-build-test-learn (DBTL) strategy, is used to divert flux towards pathways that contribute to the biosynthesis of desired products. The identification of modification/engineering targets is facilitated by modelling **(Design stage)**. The entire set of metabolic reactions present in the host organism, referred to as the genome-scale metabolic model (GEM) of that organism, is enumerated, mathematically represented based on reaction

stoichiometry and computed to predict the rates of metabolic reactions through constraint-based modelling for predicting metabolic engineering strategies for cell factory development (Chen, Li and Nielsen, 2022). These are powerful tools with potential to reduce the number of trial and error iterations involved in the DBTL cycle (Nielsen and Keasling, 2016) and their usefulness has been demonstrated in several microbial hosts (Passi *et al.*, 2022). With biosynthetic enzyme engineering being a time consuming undertaking that requires protein engineering expertise, GEMs have the potential to identify bottleneck reactions in pathways for which biosynthetic enzymes can be engineered for improved catalysis, thus reducing the number of enzymes to be engineered in order to improve flux through important biosynthetic pathways. However, the predictive power of GEMs relies on host-specific experimental data and knowledge of the host's metabolic network which may differ from model organisms upon which GEMs are commonly based. These factors, coupled with the fact that predicted engineering targets are often obvious to experts with knowledge of host organisms' physiology has prevented the need for widespread usage of GEMs for strain development, but omics data are now being applied to further constrain models to improve their predictive power (Passi *et al.*, 2022).

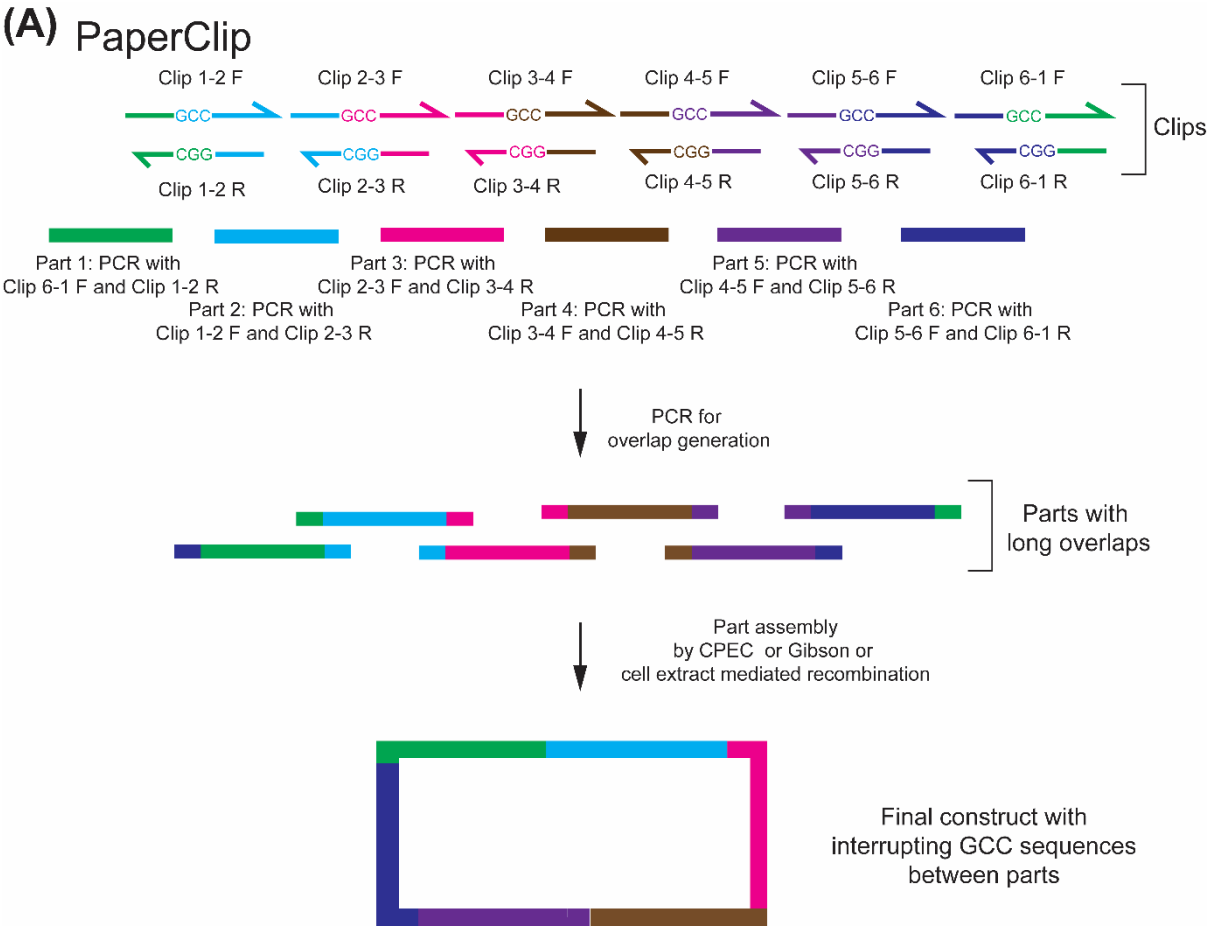
Once engineering targets are known, the next step is to establish the strategies in the host organism using genetic engineering tools (**Build stage**). The development of such tools is currently receiving a lot of attention because the knowledge of host organisms' physiology is usually adequate in the early stages of strain development to attain at least low production without the need for model-generated engineering targets. By viewing SynBio as a form of engineering, idempotent genetic engineering tools are developed based the engineering principles of abstraction and standardization to ensure that validated genetic elements are transferable between different applications and users, referred to as modular DNA assembly. Modular DNA assembly is achieved through DNA sequence-indiscriminate cloning methods that are classified into endonuclease-mediated assembly (e.g., Golden Gate assembly), site-specific recombination (e.g., Gateway), long-overlap-based assembly (e.g., Gibson assembly) and circular polymerase extension cloning (CPEC). Cloning methods that use DNA digesting enzymes work based on the fundamental principle that the enzymes expose single strand complementary sequences (overhangs) that can be annealed and ligated *in vitro,* while long-overlap-based assemblies is achieved *in vivo* by availing of endogenous homologous recombination repair machinery of *Saccharomyces cerevisiae*, *Bacillus subtilis* or *Escherichia coli*, which can take up overlapping linear DNA and assemble them into functional DNA constructs (Juhas and Ajioka, 2017). A major advantage of *in vivo* assembly is that large number and sizes of fragments can be assembled. This has enabled the construction of artificial prokaryotic genomes of up to 3.5Mb and has served as a powerful

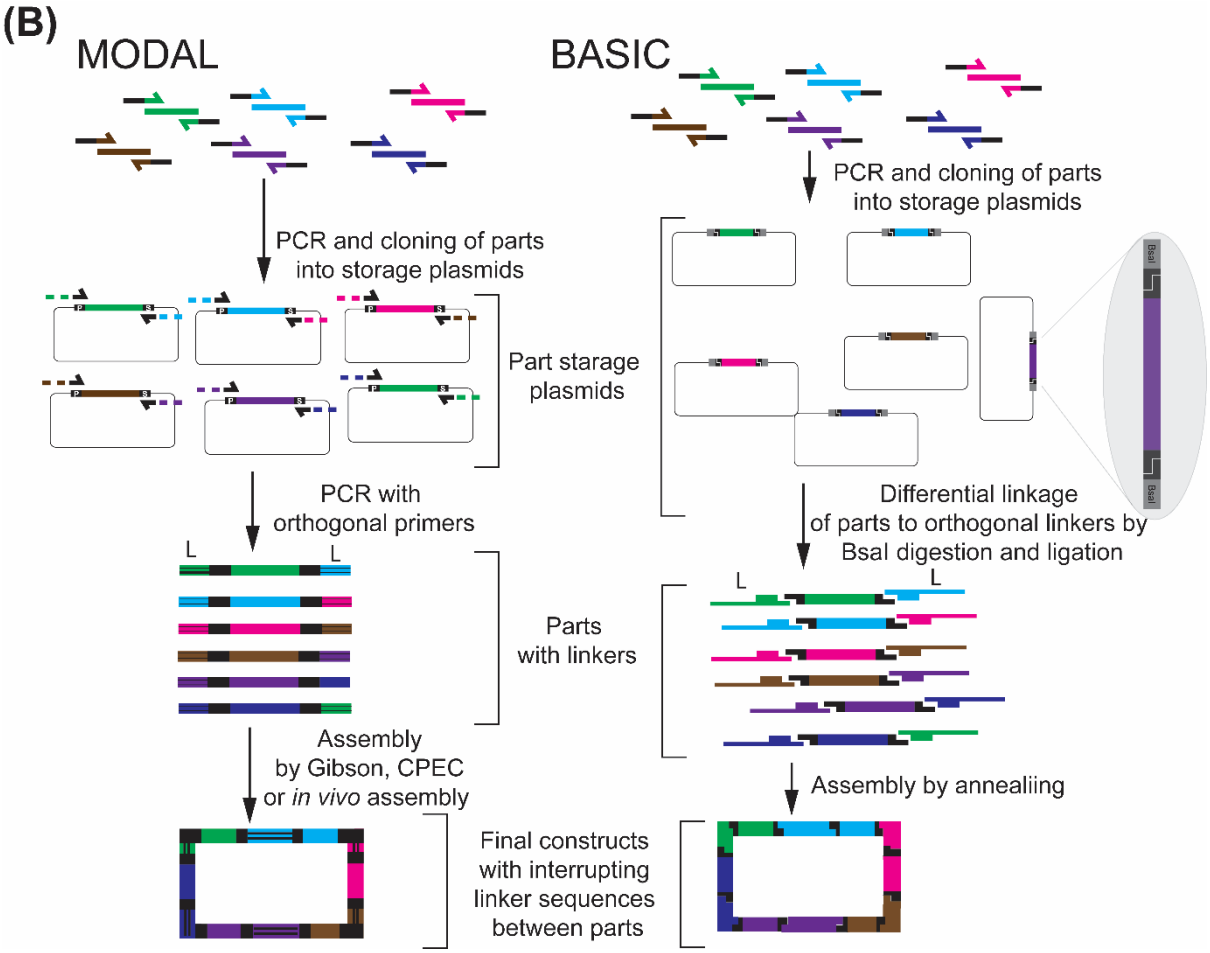
tool for complex SynBio applications especially for the bottom-up approach (Gibson *et al.*, 2008, 2010; Itaya *et al.*, 2008; Hutchison *et al.*, 2016). It was recently applied for the construction of supernumerary chromosomes of up to 100kb in *S. cerevisiae* (Postma *et al.*, 2021) and its application for the construction of "artificial yeast" based on the *S. cerevisiae* genome is under way (Richardson *et al.*, 2017; Dai *et al.*, 2020). Golden Gate assembly (GGA) and Gibson assembly (GA) are two commonly used *in vitro* cloning methods. GGA takes advantage of the ability of type IIS restriction endonucleases to cut DNA outside their recognition sites to expose overhangs of four bases (Engler *et al.*, 2009), while GA uses 5´ exonuclease activity to expose longer stretches of overhangs (Gibson *et al.*, 2009). The overhangs in both methods are then "glued" together based on sequence complementarity. CPEC is a PCR-based assembly in which DNA fragments act as primers for amplifying neighbouring fragments which act as templates, thereby generating single DNA molecules with all the template sequences (Quan and Tian, 2009). By assigning specific overhangs to functional DNA fragments based on a defined set of rules, the assembly of several fragments is achieved to generate genetic materials that can be used for specific functions in host organisms. The assembly is achieved through simultaneous DNA digestion and ligation or by PCR (CPEC) in practice, serving as the basis for several modular assembly frameworks/standards that are available for facilitating SynBio projects such as modular cloning (MoClo) which was initially based on GGA (Weber *et al.*, 2011) and later on GA (Akama-Garren *et al.*, 2016), modular overlap-directed assembly with linkers (MODAL) (Casini *et al.*, 2014) and its successor, Biopart Assembly Standard for Idempotent Cloning (BASIC) (Storch *et al.*, 2015), and PaperClip (Trubitsyna *et al.*, 2014). The core operation mechanisms of these frameworks are illustrated in figure 1. SynBio tools for applications in specific microbial (Engler *et al.*, 2014; Lee *et al.*, 2015; Moore *et al.*, 2016; Geddes, Mendoza-suárez and Poole, 2019; Larroude *et al.*, 2019; Rajkumar *et al.*, 2019), plant (Sarrion-Perdigones *et al.*, 2013; Binder *et al.*, 2014) and mammalian (Duportet *et al.*, 2014) hosts have been developed based on these standards and can be extended to other hosts of choice.

Host genome editing is a necessary aspect of SynBio because it allows host manipulation at genetic level for the disruption or repression of genes that mediate competing pathways and for functional expression of genes that positively contribute to specific SynBio applications. Although gene expression on episomal vectors is widely used for metabolic engineering applications in various hosts, this requires the maintenance of selection pressure by antibiotics or auxotrophy which may not be scalable, may be unstable and cause metabolic burden in some hosts (Ko *et al.*, 2020). Foreign genetic materials such as functional gene sequences can be integrated into host genomes by means of genome editing tools, allowing

for more controlled gene expressions. Targeted genome editing can be achieved by traditional selection marker-based tools and works by flanking DNA with sequences that are homologous to specific loci in host genome. By introducing the flanked DNA in the presence of a selective pressure, the DNA is integrated into the host genome by homologous recombination. Reports that the generation of double stranded breakages (DSB) enhanced homologous recombination in plants (Puchta, Dujon and Hohn, 1993), mammalian (Rouet, Smih and Jasin, 1994) and microbial (Resnick and Martin, 1976; Ayora *et al.*, 2011) hosts led to efforts by the scientific community to develop tools such as zinc-finger nucleases (ZFN) (Urnov *et al.*, 2010), transcription activator-like effector nucleases (TALENS) (Bogdanove and Voytas, 2011) and Cas enzymes for Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) (Jinek *et al.*, 2012) for targeted DSB generation. With the requirement of ZFN and TALENS for specific proteins for each targeted locus, these tools are difficult to establish. CRISPR on the other hand simply requires the functional expression of an endonuclease which is guided by RNA molecules. Thus, the operational difference between loci targeting is the provision of loci specific CRISPR RNA (CrRNA), making multi-loci targeting possible by simply providing multiple CrRNA. This has resulted in CRISPR overtaking ZFN and TALENS as the tool of choice for DSB generation and is even applied for other genome targeting applications such as CRISPR interference (CRISPRi) in which genes are repressed by the endonuclease acting as a physical obstacle to host's RNA polymerases to prevent transcription (Larson *et al.*, 2013). Although the development of CRISPR tools for hosts is faced by challenges including off-target effects, cytotoxicity, and low efficiency of multi-gene editing as recently reviewed (Li, Huang and Chen, 2022), they have been developed for many SynBio hosts.

Once constructed, host strains are tested to determine whether established strategies improve production (**Test/Learn stage**), adding to preexisting knowledge for strain construction. The complexity of biological systems warrants that several strategies be established which often features the testing of several gene expression systems and is complicated with increasing number of players in the SynBio application such as number of biosynthetic enzymes involved in a pathway. Thus, the performance of multiple strains must be tested. This requires their cultivation in industrially realistic conditions and screening, currently widely achieved by determining production titre through time-consuming chromatographic techniques. The delays involved at this stage represent a major bottleneck in strain development. High throughput cultivation and screening technologies are highly useful in mitigating this challenge as they speed up strain development by permitting simultaneous cultivation and screening of several strains generated by multiple strategies for decision making on how to proceed in order to foster the DBTL cycle (Rienzo *et al.*, 2021).


(A) PaperClip
Clip 1-2 F
Clip 2-3 F
Clip 3-4 F
Clip 4-5 F
Clip 5-6 F
Clip 6-1 F
GCC
CGG
Clips
Clip 1-2 R
Clip 2-3 R
Clip 3-4 R
Clip 4-5 R
Clip 5-6 R
Clip 6-1 R
Part 1: PCR with
Clip 6-1 F and Clip 1-2 R
Part 2: PCR with
Clip 1-2 F and Clip 2-3 R
Part 3: PCR with
Clip 2-3 F and Clip 3-4 R
Part 4: PCR with
Clip 3-4 F and Clip 4-5 R
Part 5: PCR with
Clip 4-5 F and Clip 5-6 R
Part 6: PCR with
Clip 5-6 F and Clip 6-1 R
PCR for
overlap generation
Parts with
long overlaps
Part assembly
by CPEC or Gibson or
cell extract mediated recombination
Final construct with
interrupting GCC sequences
between parts


(B)
MODAL
BASIC
PCR and cloning of parts into storage plasmids
PCR and cloning of parts into storage plasmids
Part starage plasmids
PCR with orthogonal primers
Differential linkage of parts to orthogonal linkers by BsaI digestion and ligation
L
L
L
L
Parts with linkers
Assembly by Gibson, CPEC or in vivo assembly
Assembly by annealiing
Final constructs with interrupting linker sequences between parts

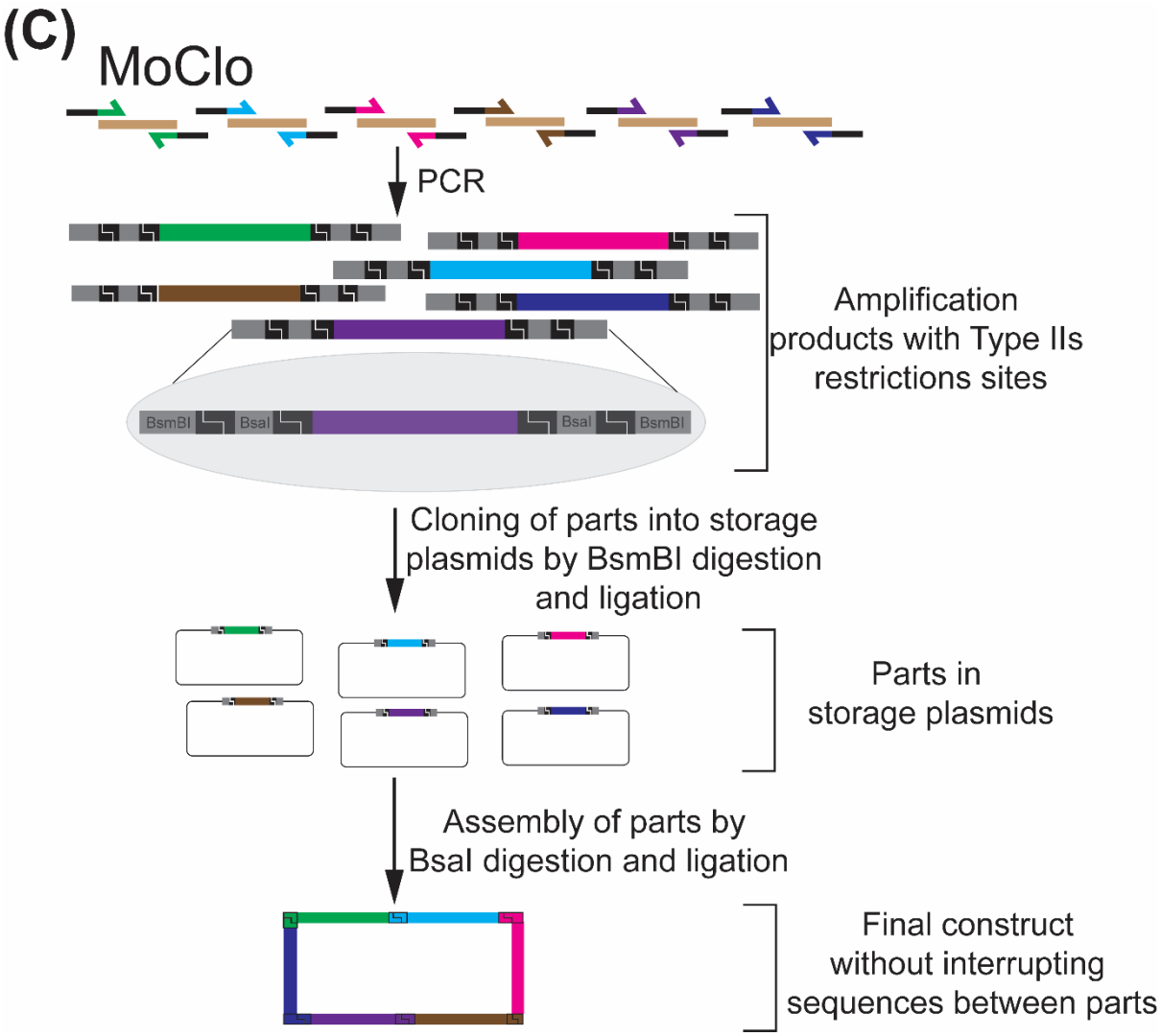


Figure 1. Modular assembly frameworks/standards for SynBio. (A) PaperClip. Uses short double stranded DNA molecules called “Clips” to prime and amplify the parts. Clips are homologous to the 5’ and 3’ ends of parts to be assembled for annealing to the part sequence. During the first PCR step, a clip denatures into forward primer for a part and reverse primer for the immediately preceding part (according to the desired concatenation order of parts in the final construct) to generate parts with long overlaps for subsequent assembly by CPEC (PCR-based assembly in which overlapping sequences of a part serve as primers for amplifying its neighbours), Gibson assembly or by cell extract-mediated recombination. A limitation of PaperClip is that GCC sequence is introduced between parts. (B) MODAL. Parts are first amplified with primers that have 25bp tails for incorporating prefix (P) and suffix (S) sequences at 5’ and 3’ end of parts, respectively. This enables part cloning into storage plasmids. Part storage plasmids are then used as templates for amplifying the parts using orthogonal primers with 40bp tails (represented as broken half arrows). These tails form overlapping sequences (“linkers”, L). The linker for each part is predetermined according to the desired concatenation order of parts in the final construct. Overlapping parts are then assembled by Gibson assembly, CPEC or even in vivo assembly; BASIC is similar to MODAL with the difference is that prefix and suffix sequences now contain inwards-facing BsaI sites to enable the addition of orthogonal linkers (L) by BsaI digestion and ligation. A disadvantage of MODAL and BASIC is the introduction of interrupting sequences between parts by linkers. (C) MoClo. Parts are first amplified with primers that have tails for incorporating inward facing Type IIs (BsaI and BsmBI) restriction sites. Parts can then be cloned into storage plasmids by BsmBI digestion and ligation. This removes the BsmBI sites but leaves behind the BsaI site in storage plasmids for parts assembly by BsaI digestion and ligation. A major advantage of MoClo is that interrupting sequences are not introduced. Type IIS overhangs are only 4bp, which can limit assembly efficiency, but adaptation of MoClo for Gibson assembly can improve efficiency due to long overlaps.

1.2. Reducing the cost of starting materials

The starting materials used for bio-based production which are supplied as components of the growth media for cultivating the host are an important cost factor to be considered in commercializing a bioprocess. Therefore, a lot of attention is currently directed at finding cheap feedstock. Due to high demand for renewable and sustainable energy sources, biofuel production has received and continues to receive a lot of attention in this area and the developments seen for feedstock provision for biofuel production are extendable to the production of other bioproducts such as aromatics. Although first-generation feedstocks from sugar crops, which can include food crops, can compete with food supply and is socially unsustainable, can be used as starting materials for producing value-added compounds, the movement of biofuel production from such feedstock (first-generation production) to second-generation production which by definition does not compete against food supplies (Aditiya *et al.*, 2016; Robak and Balcerek, 2018) means that the commercial production of aromatics is likely to also conform to this trend.

Lignocellulosic biomass (LCB) is a promising feedstock option due to its abundant availability and, being plant-based, is regarded as renewable. However, it is composed of complex polysaccharides and aromatic polymers that cannot be metabolized by most host organisms and must be broken down into metabolizable monomers (Mahmood *et al.*, 2019; Bertacchi *et al.*, 2022). The recalcitrant nature of LCB makes its deconstruction difficult, possibly energy-intensive and is the principal challenge in attaining economic viability for commercializing LCB-based processes (Mahmood *et al.*, 2019). Sources of LCB include forestry, dedicated energy crops, agricultural residues, algae and industrial and household wastes. LCB recalcitrance varies between sources as a factor of physico-chemical composition. It is mainly composed of cellulose as the integral polymer of D-glucose monomers linked by β-(1,4) glycosidic bonds, hemicellulose as a hetero-polymer of glucose, galactose, mannose, arabinose, xylose, and rhamnose monomers linked by β-(1,4)/or β-(1,3)-glycosidic bonds, and lignin as a hetero-polymer of aromatic monomers called monolignols (coumaryl alcohol, coniferyl alcohol and sinapyl alcohols) linked by alkyl-aryl, alkyl-alkyl, and aryl-aryl group linkages (Ponnusamy *et al.*, 2019). Celluloses form the embedded microfibrils of lignocellulose matrices and are consequentially resistant to enzymatic hydrolysis while hemicelluloses are amorphous, physically weak but act as physical barrier to enzyme accessibility to celluloses and are easily deconstructed. Lignin is hydrophobic, structurally rigid and is linked to carbohydrates to form lignin-carbohydrate complexes (LCC) (Tarasov, Leitch and Fatehi, 2018). The nature of LCC is an important factor determining LCB recalcitrance (Bertacchi *et al.*, 2022).

The deconstruction LCB involves the depolymerization of component polymers using enzymes that break linkages to release the monomeric sugars which are assimilable as starting materials for bioprocesses. This process requires initial physical, chemical, thermal and biological pre-treatment methods that are applied individually or in combination to destabilise the structural barriers so as to make the enzymes accessible to the linkages (Mahmood *et al.*, 2019; Mankar *et al.*, 2021). The pretreated biomass can then be subjected to enzymatic reactions that hydrolyse the linkages with the choice of enzyme(s) determined by the characteristic components of the particular LCB. It is common to take advantage of the catalytic capability of more than one enzyme by using them as cocktails for achieving close to complete degradation of polymers (Houfani *et al.*, 2020). While LCB is abundant around us, this process of deconstruction could depose it from its position as a cheap feedstock due to the cost accrued in the process and may even become unsustainable due to energy requirements. To curb these challenges, microorganisms that are able to hydrolyse natural biopolymers and enzymes are being investigated as cheap and ecofriendly means for deconstructing lignocellulose, but the time factor associated with this method currently limits its industrial adoption (Sharma, Xu and Qin, 2019). A more drastic approach of genetically modifying plants to reduce the complexity of lignocellulose has also been proposed (Xie and Peng, 2011). However, in addition to the fact that this represents a considerable move away from taking advantage of the abundant availability of LCB as its major upside, this approach could be ecologically detrimental given that lignocellulose is an important component of plant cell wall and modification of cell wall complexity is associated with poor plant health (Velvizhi *et al.*, 2022). Furthermore, the cultivation of lignocellulosic crops to serve as feedstock can threaten food security and biodiversity due to monocropping (Evans, Kelley and Potts, 2015; Hasegawa *et al.*, 2015).

The input of waste materials into manufacturing serves as a major win in the efforts to attain a circular economy. Municipal Solid Wastes (MSWs) are simply defined as everyday items that are thrown away as garbage at homes, hospitals, schools and businesses. Given its high contribution to wastes globally (Jain *et al.*, 2022), MSWs are receiving a lot of attention in regard to the production of added-value products from them. Organic Fraction of Municipal Solid Wastes (OFMSW) in particular are potential feedstock for bioprocesses but require refining before application for this purpose. As part of the EU Research and Innovation Horizon 2020 program, the BioRen project focused on the development of MSW refining (pretreatment method and hydrolysis) and fermentation for biofuel production towards a circular economy (Kowalski *et al.*, 2022). One of the outputs of this project was the successful development of an efficient technology for converting paper/cardboards wastes into cellulosic ethanol which involves separation of these materials from MSW, physical-mechanical

pretreatments by shredding, heating, stirring and washing in addition to chemical treatments with dilute acids was used to enhance the saccharification of cellulosic compounds present in the paper/cardboards wastes (Verhe *et al.*, 2022). The successful pilot scale production of ethanol by simultaneous saccharification and *S. cerevisiae* fermentation demonstrated the potential of the authors' method for deriving bioprocess feedstock from OFMSW and may be applicable to production of other value-added compounds like aromatics.

Although LCB has potential to provide cheap feedstock, the cost involved in its valorisation as discussed above could mean that while it represents an eco-friendly feedstock that can reduce carbon footprint and contribute positively to sustainable manufacturing, it might turn out to be a more expensive alternative to first-generation feedstock. This is likely to limit adoption. In the meantime, bioprocess side streams from various manufacturing industries that will normally be disposed of can be used as cheap feedstocks. Additionally, this represents an eco-friendly and possibly profit-boosting approach to process waste management for manufacturing plants. For example, biodiesel side stream glycerol waste was recently used as cheap carbon source for producing the biodegradable surfactants, rhamnolipids in a *Pseudomonas aeruginosa* host as replacements for nonbiodegradable surfactants (Baskaran *et al.*, 2021). Dairy processing operations produce side streams that are rich in nutrients and can thus be used as feedstock for production of value-added products. The precipitation and removal of milk casein during cheese-making produces cheese whey while second cheese whey is the liquid remaining after whey cheese separation during the production of cottage and ricotta cheeses (Asunis *et al.*, 2020). Cheese whey and second cheese whey are the main by-products of the dairy industry with 9-10L of whey resulting from the production of 1kg of cheese, majorly consist of lactose and are high-polluting agents, making their disposal a problem for the dairy sector (Zotta *et al.*, 2020; Figueroa Pires *et al.*, 2021). Microbial fermentations are, however, now used to valorise these dairy side streams to obtain value-added products including biofuels, beverages and biopolymers (Zotta *et al.*, 2020; Figueroa Pires *et al.*, 2021).

2. Variations in the aromatic amino acid biosynthetic pathways of plants and microorganisms

Although plants are natural sources of biological aromatics, plant cell factories are not commonly developed for their production due to challenges such as insufficient genetic engineering tools and potentially difficult downstream processing challenges. Plastid specificity of biosynthetic reactions is a major challenge in plant metabolic engineering. In particular, the presence of plastid import signal sequences in enzymes of the shikimate pathway suggests that the pathway occurs in chloroplast (Mousdale and Coggins, 1985;

Smart and Amrhein, 1987; Schmid *et al.*, 1990; Doong, Ganson and Jensen, 1993; Herrmann and Weaver, 1999; Mustafa and Verpoorte, 2005; Weber, Schwacke and Flügge, 2005; Tzin and Galili, 2010; Maeda and Dudareva, 2012). Thus, the engineering of plant cells for increased production of aromatics from metabolites of the pathway will require extensive modifications and while plastid engineering is a solution, tools for achieving it are limited (Fuentes, Armarego-Marriott and Bock, 2018). Nonetheless, plants serve as interesting biological systems from which a wealth of knowledge can be obtained to inform on metabolic engineering strategies especially for producing specific plant aromatics. Thus, we explore some key differences and similarities in the aromatic amino acids (AAA) biosynthetic pathways of plants and microorganisms in this section. The pathway is schematised in figure 2.

AAAs are produced in the shikimate pathway which begins with the condensation of phosphoenolpyruvate (PEP) coming from the glycolytic pathway and erythrose 4-phosphate (E4P) from the pentose phosphate pathway (PPP) to form 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) through the action of DAHP synthase (figure 2). DAHP synthases are present as isoforms in plants, yeasts and bacteria. *Escherichia coli* has three- AroF, AroG and AroH and the model plant *Arabidopsis thaliana* also has three- Dahps1 (At4G39980), Dahps2 (At4G33510), Dahps3 (At1G22410) and *S. cerevisiae* has two isoforms- Aro3 and Aro4. *E. coli* AroG contributes most of DAHP synthase activities with some contributions from AroF and AroH (Tribe, Camakaris and Pittard, 1976). In *S. cerevisiae*, while the disruption of both *ARO3* and *ARO4* is sufficient for AAA auxotrophy, individual inactivation of the genes is not, showing that the genes can compensate for each other (Teshiba *et al.*, 1986). In plants, most of the AAAs produced in the pathway are directed towards biosynthesising specialised aromatic metabolites and structural polymers unlike in microorganisms in which these amino acids are mostly used for protein synthesis. Accordingly, 30-45% of plant organic matter is derived from phenylalanine (Singh, Lewis and Towers, 1998). This coupled with the fact that more than 20% of fixed carbon is estimated to go towards the biosynthesis of aromatic compounds in plants (Metcalf, 1987; Herrmann, 1995; Tohge *et al.*, 2013) are major disparities between plants and microorganisms. Thus, the flux towards AAA biosynthesis in plants may require special regulation of the DAHP synthase reaction unlike those seen in microorganisms to ensure adequate precursor supply for producing aromatic metabolites. Indeed, carbon flow into the shikimate pathway is linked to photosynthetic electron flow (photosystem I) through the DAHP synthase activity requirement for reduced thioredoxin (Entus, Poling and Herrmann, 2002). As in other organisms, DAHP synthase is a major regulatory checkpoint of the shikimate pathway in plants and the transcription of DAHP synthase genes increases during bacterial infection by the plant pathogenic bacteria

Pseudomonas syringae, physical wounding, and UV-B treatment (Keith *et al.*, 1991; Ramani, Patil and Jayabaskaran, 2010). This suggests that plants regulate flux through the pathway in response to environmental stimuli at the point of DAHP synthase and its transcription and enzyme activity increases within hours in such conditions (Dyer *et al.*, 1989).

The activities of DAHP synthase isoforms are generally regulated by allosteric feedback inhibitions by AAAs in microorganisms: *E. coli* AroF, AroG and AroH are sensitive to tyrosine, phenylalanine, and tryptophan, respectively (Camakaris and Pittard, 1974; Panina *et al.*, 1999). *S. cerevisiae* Aro3 is sensitive to phenylalanine and Aro4 to tyrosine (Paravicini, Schmidhetni and Braus, 1989). Activities in higher plants were thought to not be subject to major allosteric control based on reports that *in vitro* activities of the enzymes from different plants are only weakly inhibited or even become activated by AAAs, (Herrmann and Weaver, 1999; Tzin and Galili, 2010; Maeda and Dudareva, 2012). Although the mechanisms of DAHP synthase regulation in plants is poorly understood, the expression of bacterial feedback insensitive DAHP synthase in *A. thaliana* and *Solanum lycopersicum* (tomato) resulted in the accumulation of shikimate, the three AAAs and their derivatives (Tzin *et al.*, 2012, 2013), indicating that DAHP synthase, like in microorganisms, is a limiting reaction in the plant shikimate pathway. Studies are now looking at plant DAHP synthases from the viewpoint that they are structurally similar to the DAHP synthase enzyme of *Mycobacterium tuberculosis* in that the three plant DAHP synthases have at least two allosteric effector binding sites. *M. tuberculosis* DAHP synthase is classified as a type II enzyme and requires simultaneous binding of phenylalanine and tryptophan for its allosteric inhibition (Webby *et al.*, 2010). In contrast, *E. coli* and yeast DAHP synthase isoforms are type I due to the presence of a single allosteric effector binding site (Hartmann *et al.*, 2003). A recent report that AAA supplementation reduced DAHP synthase activity in young seedlings but not in mature leaves showed that DAHP synthase activities can be inhibited *in vivo* by AAAs (Yokoyama *et al.*, 2021). Given that *DAHPS1* is predominantly expressed in mature leaves and *DAHPS2* in young seedlings, it was further shown *in vitro* that Daphs2, but not Daphs1 and Daphs3, was inhibited by tyrosine and tryptophan. This explains the difference in AAA effect on DAHP synthase activity in the two plant tissues and suggests that Dahp2 is under feedback inhibition. Furthermore, a highly complex regulation of DAHP synthase in which chorismate (a downstream intermediate of the shikimate pathway and the last common precursor of all AAAs) and caffeate (a phenylpropanoid intermediate) strongly inhibited the three DAHP synthase isoforms was revealed, but arogenate (a downstream intermediate below chorismate and the precursor of AAAs) counteracted Dahps1 and Dahps3 inhibitions. Although Kanaris and colleagues were unable to independently repeat the inhibitory effect of tyrosine on Dahps2 *in vitro*, these authors also found *in vivo* evidence suggesting that Dahps2

is the likely candidate of the three isoforms inhibited by tyrosine (Kanaris *et al.*, 2022). They detected reduced accumulation of Dahps2-yellow fluorescent protein fusion protein in the chloroplast not detected for the other two isoforms upon tyrosine treatment. Based on previous report that Dahps1 and Dahps2 interact with 14-3-3 omega protein in *A. thaliana* (Chang *et al.*, 2009), they explored the possibility that this protein may play a role in the cytosolic sequestration of Dahps2 upon tyrosine treatment, thereby limiting its importation into the chloroplast. Their experiments revealed that the C-terminal motif of Dahps2 is important for its tyrosine-regulated binding with 14-3-3 omega protein and they found evidence suggesting that the deletion of this motif results in Dahsp2 interactions with 14-3-3 omega protein to become unresponsive to modulation by tyrosine (Kanaris *et al.*, 2022). Another notable difference between DAHP synthases of plant versus microorganisms is that the plant enzymes show much higher affinity (K_m) toward E4P than PEP as has been demonstrated for *A. thaliana* (Yokoyama *et al.*, 2021), *Spinacia oleracea* (spinach) (Doong *et al.*, 1992) and *Daucus carota* (carrot) (Suzuki, Sakuta and Shimizu, 1996) compared to the *S. cerevisiae* enzymes which show higher affinity for PEP than E4P (Schnappauf *et al.*, 1998)

DAHP is further converted in a five-step reaction that is catalysed by Aro1 in yeasts and culminates in the formation of 5-enolpyruvoyl-shikimate 3-phosphate (ESPS). The *S. cerevisiae* Aro1 is a pentafunctional enzyme bearing catalytic domains for each of the five bioconversions and is encoded from a single open reading frame (ORF). In contrast, the reactions are catalysed by five monofunctional enzymes encoded by five individual ORFs in bacteria, and 3 monofunctional enzyme and a bifunctional enzyme in plants. DAHP is converted to 3-dehydroquinate by 3-dehydroquinate synthase enzyme which exist as single isoforms Dhqs (At5g66120) in *A. thaliana* and AroB in *E. coli*. The third and fourth steps involving the dehydration of 3-dehydroquinate to 3-dehydroshikimate which is in turn dehydrogenated into shikimate are catalysed by a bifunctional enzyme encoded by a single ORF in *A. thaliana* (Dhq/Sdh, At3g06350) while *Nicotiana tabacum* (tobacco plant) has two of these genes with the enzymes bearing a 3-dehydroqinate dehydratase and a shikimate 5-dehydrogenase domains. The two reactions are catalysed by two separate enzymes in *E. coli* with a single isoform of 3-dehydroqinate dehydratase, AroD, and the nicotinamide co-factor-dependent shikimate 5-dehydrogenase existing as two isoforms, AroE and YdiB. AroE which only takes NADP as co-factor has catalytic efficiency of up to 4000-fold higher than YdiB, making AroE the main enzyme catalysing this reaction in *E. coli* (Michel *et al.*, 2003). YdiB takes NAD and NADP as co-factors but also has specificity for quinate, a structurally similar molecule to shikimate, thus making it a quinate/shikimate dehydrogenase. The quantification of quinate and shikimate production by wildtype *E. coli* showed that the disruption of this gene has no effect on shikimate production but decreased quinate

production while its overexpression increased quinate production (García *et al.*, 2017). Therefore, YdiB is thought to be involved in the biosynthesis of quinate but not shikimate. Based on the crystal structure of *A. thaliana* Dhq/Sdh and the interaction of the domains with substrates it was proposed that the mode of substrate binding is conserved between plant and microbial Dhq dehydratase enzymes, and that the arrangement of the two enzyme domains is suggestive of metabolic flux control by means of the close proximity of the two sites, thereby increasing the effective concentration of the intermediate substrate, 3-dehydroshikimate (Singh and Christendat, 2006).

Next, the phosphorylation of shikimate into shikimate 3-phosphate is catalysed by shikimate kinase which exists as two isoforms in *A.* thaliana- Sk1 (At2g21940) and Sk2 (At4g39540) and in *E. coli*- AroL and AroK. Although most of the enzyme activity is contributed by AroK in *E. coli* based on data that its inactivation considerably reduced the activity, AroL also contribute some activity (Chen *et al.*, 2012). A study that looked at the effect of fungal elicitors on the transcription of AAA biosynthesis genes in tomato cells showed that shikimate kinase transcripts rapidly increased after treatment with the elicitor, presumably in favour of increasing precursor supply to biosynthetic pathways of secondary metabolites as a defence response to the elicitor (Gorlach *et al.*, 1995). Similar to DAHP synthase, shikimate kinase activity in spinach is stimulated by reduced thioredoxins and is regulated by energy charge (Schmidt *et al.*, 1990). Thus, this reaction is suspected to serve as a regulatory step in the shikimate pathway of plants for controlling metabolic flux towards the biosynthesis of some secondary metabolites. In line with this notion, the single shikimate kinase isoform of *Petunia hybrida* (petunia) was recently reported to play a role in the production of anthocyanin and flavonoid molecules in this plant (Yuan *et al.*, 2022). Shikimate 3-phosphate and PEP are then converted into 5-enolpyruvylshikimate 3-phosphate (EPSP) by EPSP synthase which exists as two isoforms in *A. thaliana*- Epsps1 (At2g45300) and Epsps2 (At1g48860) and as a single isoform in *E. coli* (AroA). ESPS is further converted into chorismate through the action of chorismate synthase Aro2 in *S. cerevisiae* and AroC in *E. coli.* Chorismate is a central branch point molecule in plants and microorganisms because its serves as precursor to aromatic amino acids and other molecules such as tetrahydrofolate (vitamin B9), phylloquinone (vitamin K1) and salicylate. In regard to AAAs, the shikimate pathway divides at this point into the phenylalanine/tyrosine branch and the tryptophan branch to produce the respective AAAs.

Towards the phenylalanine/tyrosine branch, chorismate is converted into prephenate by chorismate mutase, which is a monofunctional enzyme in *S. cerevisiae* (Aro7) and in some bacteria such as *Bacillus subtilis*. In *E. coli,* however, the enzyme is bifunctional, catalysing, in addition to the conversion of chorismate to prephenate, the conversion of prephenate to

phenylpyruvate towards phenylalanine in the case of PheA and 4-hydroxyphenylpyruvate towards tyrosine in the case of TyrA. Chorismate mutases are allosterically regulated by AAAs, serving as another checkpoint for controlling flux through the shikimate pathway. Aro7 activity in *S. cerevisiae* is inhibited by tyrosine and stimulated by tryptophan to ensure balance between both branches for chorismate metabolism (Brown and Dawes, 1990), and in *E. coli*, PheA and TyrA are feedback inhibited by phenylalanine and tyrosine, respectively (Pohnert *et al.*, 1999; Lütke-Eversloh and Stephanopoulos, 2005). Such regulation prevents overdrive of one branch at the expense of the other by ensuring the diversion of chorismate from tryptophan synthesis to phenylalanine/tyrosine synthesis in high tryptophan conditions and the reduction of phenylalanine/tyrosine synthesis in high phenylalanine/tyrosine conditions. The conversion of prephenate to phenylpyruvate is catalysed by prephenate dehydratase Pha2, and to hydroxy-phenylpyruvate by prephenate dehydrogenase Tyr1 in *S. cerevisiae*. The three chorismate mutases present in *A. thaliana* have been identified as Cm1 (At3g29200), Cm2 (At5g10870) and Cm3 (At1g69370) (Mobley, Kunkel and Keith, 1999). Similar to Aro7, the activities of AtCm1 and 3 are sensitive to inhibition by phenylalanine and tyrosine, and AtCm1 to stimulation by tryptophan. However, AtCm2, which is reported to have similar specific activity (Eberhard et al., 1996) as and higher specificity (Westfall, Xu and Jez, 2014) for chorismate than AtCm1, is not sensitive to AAAs. The localisation of AtCm2 in the cytosol unlike the other two isoforms which are found in the plastid support speculations that cytosolic shikimate pathway exists in plants (extraplastidial shikimate pathway duplication) (Lynch, 2022). The better catalysis of AtCm2 than the other two enzymes could be required for driving this alternative pathway.

The final steps for AAA biosynthesis in plants involves the conversion of prephenate into arogenate by prephenate aminotransferase, Ppa-at (AT2G22250) (Maeda, Yoo and Dudareva, 2011). Arogenate dehydratase and dehydrogenase then convert arogenate into phenylalanine and tyrosine, respectively. *A. thaliana* has six Arogenate dehydratases- Adt1 (At1g11790), Adt2 (At3g07630), Adt3 (At2g27820), Adt4 (At3g44720), Adt5 (At5g22630) and Adt6 (At1g08250)) and two Arogenate dehydrogenase (TyrA1 (At5g34930) and TyrA2 (At1g15710)). Of the six Adt enzymes, three of them (Adt1, Adt2 and Adt6) are able to convert prephenate to phenylpyruvate as a side reaction (Cho *et al.*, 2007) and the presence of enzymes with similar features have been reported in *Oryza sativa* (rice) (Yamada *et al.*, 2008) and *P. hybrida* (Maeda *et al.*, 2010). Thus, although the arogenate route is preferred for tyrosine and phenylalanine biosynthesis in plants, the phenylpyruvate/4-hydroxyphenylpyruvate route is also present (Tzin and Galili, 2010). Some bacteria belonging to the cyanobacteria and Pseudomonas species also biosynthesise tyrosine and phenylalanine via arogenate (i.e amination of prephenate before

dehydration/dehydrogenation), but most microorganisms use the phenylpyruvate/4-hydroxyphenylpyruvate (dehydration/dehydrogenation of prephenate before amination) route. Phenylpyruvate and 4-hydroxyphenylpyruvate are finally converted into phenylalanine and tyrosine through reversible transamination by aromatic amino transferases Aro8 and Aro9 in *S. cerevisiae* and mainly TyrB in *E. coli*. There are still critical knowledge gaps regarding these transamination reactions, which is required for completing the biosynthesis of both amino acids via the phenylpyruvate/4-hydroxyphenylpyruvate route, in plants (Wang and Maeda, 2018). Although several plant aminotransferases have been characterised, they mainly participate in AAA catabolism, deaminating AAAs into aromatic 2-oxo acids towards the production of a diverse array of plant natural products including phenylpropanoids and alkaloids and also the plant hormone, auxin (Wang and Maeda, 2018). During the early times when plant AAA biosynthetic pathway was being elucidated, a partially purified protein from mung bean was reported to transaminate 4-hydroxyphenylpyruvate into tyrosine but not phenylpyruvate into phenylalanine using glutamate as amino donor (Rubin and Jensen, 1979). Since then, plant AATs that have been demonstrated to have aromatic 2-oxo acid aminotransferase activities for converting aromatic 2-oxo acids to AAAs mainly favour AAA deamination and often displayed these activities when coupled to the deamination of another AAA (Wang and Maeda, 2018). For example, a *P. hydrida* AAT that preferentially converts phenylpyruvate to phenylalanine but strongly favours tyrosine as the amino donor was identified (Yoo *et al.*, 2013). This is contrary to ScAro8 which mainly catalyses the amination of phenylpyruvate and 4-hydroxyphenylpyruvate into phenylalanine and tyrosine in *S. cerevisiae* and favours glutamate as an amino donor (Urrestarazu *et al.*, 1998).

Towards the tryptophan branch, chorismate is aminated into anthranilate, the first committed metabolite in tryptophan biosynthesis. This reaction is achieved by the action of a complex formed by anthranilate synthase and the N-terminal domains of indole 3-glycerol phosphate synthase, catalysing the fourth reaction in *S. cerevisiae,* or of anthranilate phosphoribosyl transferase, catalysing the second reaction, in *E. coli*. These N-terminal domains have glutamine amidotransferase activity for functional complementation of anthranilate synthase. The complex competes against chorismate mutase in the phenylalanine/tyrosine branch for chorismate and serves therefore as an important checkpoint in regulating tryptophan biosynthesis through feedback inhibition of anthranilate synthase by tryptophan. Anthranilate synthase is encoded by *TRP2* in *S. cerevisiae* and anthranilate phosphoribosyl transferase, which converts anthranilate and, the pentose phosphate, phosphoribosyl pyrophosphate (PRPP) into phosphoribosyl anthranilate (PRA) and inorganic pyrophosphate is encoded by *TRP4*. PRA is then isomerised into carboxyphenylamino-1-deoxyribulose 5-phosphate (CDRP) by PRA isomerase, encoded by *TRP1*. As the penultimate step, the C-terminal

domain of indole 3-glycerol phosphate synthase Trp3 converts CDRP to indole 3-glycerol phosphate. Interestingly, anthranilate synthase, glutamine amidotransferase and PRA isomerase exist as a single trifunctional enzyme in fungi outside the Saccharomycetaceae family (Gaertner and DeMoss, 1969; Keesey and Demoss, 1982; Kos *et al.*, 1985; Mullaney *et al.*, 1985). In *E. coli,* anthranilate synthase is encoded by *trpE* and the bifunctional glutamine amidotransferase/anthranilate phosphoribosyl transferase by *trpD*. In some bacteria such as *Serratia marcescens* and also the hyperthermophilic archaeon, *Sulfolobus solfataricus*, however, glutamine amidotransferase and anthranilate phosphoribosyl transferase are separately encoded (Nichols *et al.*, 1980; Tutino *et al.*, 1993). Furthermore, PRA isomerase and indole 3-glycerol synthase are found in *E. coli* as a bifunctional enzyme encoded by *trpC*. Intriguingly, it was recently discovered that in contrast to *S. cerevisiae* and the filamentous fungi *Aspergillus niger* in which anthranilate has feed-forward activation effect on indole 3-glycerol synthase domains, anthranilate effect on the *E. coli* enzyme at the N-terminal indole 3-glycerol synthase domain of TrpC is inhibitory (Chen *et al.*, 2018). Indole 3-glycerol phosphate is then finally converted into tryptophan by tryptophan synthase, encoded by the *S. cerevisiae TRP5*. Trp5 is a bifunctional enzyme with two active sites, one to cleave indole 3-glycerol phosphate into indole and glyceraldehyde 3-phosphate, and the other to condense indole and serine into tryptophan. In *E. coli,* tryptophan synthase exists as a tetrameric protein with *trpA* encoding the alpha subunits and *trpB* encodes the beta subunits.

Similar to *E. coli* and *S.* cerevisiae, tryptophan biosynthesis in plants commences with the formation of anthranilate from chorismate by anthranilate synthase-glutamine amidotransferase complex, but unlike *E. coli and S.* cerevisiae, plant glutamine amidotransferase activity is not linked to indole 3-glycerol phosphate synthase or anthranilate phosphoribosyl transferase. Instead, anthranilate synthase exists as a tetrameric enzyme complex with alpha and beta subunits (Poulsen, Verpoorte and Bongaerts, 1993). The alpha subunit functions as anthranilate synthase while the beta subunit is the glutamine amidotransferase for transferring an amino group from glutamine to the alpha subunit. As in yeasts and bacteria, anthranilate synthase activity is feedback inhibited by tryptophan through its binding of the alpha subunit and the expression of feedback resistant alpha subunit increases the biosynthesis of tryptophan and its derivatives (Tozawa *et al.*, 2001; Hughes *et al.*, 2004). Therefore, similar to in microorganisms, this step is a regulatory checkpoint in the plant pathway and the expression of the genes encoding the subunits are co-ordinately regulated in response to bacterial pathogen infiltration (Niyogi *et al.*, 1993). Based on the ability to complement *E. coli* mutant of anthranilate synthase, *A. thaliana* was reported to possess at least two functional alpha subunit-encoding genes- *ASA1* (At5g05730) and *ASA2*

(At2G29690), and two beta subunit-encoding genes- *ASB1* (At1G25220) and *ASB2* (At5G57890) (Niyogi and Fink, 1992; Niyogi *et al.*, 1993), but other genes encoding putative proteins of both subunit are found in *A. thaliana* genome (Tzin and Galili, 2010). A single functional isoform of anthranilate phosphoribosyl, Pat1 (At5g17990) which coverts anthranilate to PRA was reported (Elledge *et al.*, 1991), but a putative protein (At1G70570) is also suspected to catalyse this reaction. PRA is then isomerised into CDRP by 3 isoforms of phosphoribosyl anthranilate isomerase-Prai1 (At1g07780), Prai2 (At5g05590), Prai3 (At1g29410.2). Although two indole 3-glycerol phosphate synthase isoforms are present only I3gps1 (At2G04400) has been reported as functional and I3gps2 (At5G48220) remains putative (Li *et al.*, 1995). This enzyme generally exists by itself in plants unlike in microorganisms where it exists as the C or N-terminal domains of other enzymes in the pathway as discussed above. As it is required for the formation of indole rings in plants, indole-3-glycerol phosphate synthase is a highly important enzyme required not only for the biosynthesis of tryptophan but also for the phytohormone indole 3-acetic acid (auxin), with indole 3-glycerol phosphate acting as a branch point molecule (Ouyang, Shao and Li, 2000). The formation of tryptophan from indole 3-glycerol phosphate in plants is structurally similar to *E. coli* in that the cleavage and condensation steps are catalysed by two separately encoded alpha and beta subunits that complex into a tetrameric enzyme with the alpha subunit cleaving indole 3-glycerol phosphate to generate indole for condensation with serine to form tryptophan by the beta subunit. The alpha subunit protein is encoded by at least one functional gene, *TSA1* (At3g54640.1) but a putative gene (At4g02610) with over 60% homology with the functional proteins also exists (Radwanski, Zhao and Last, 1995; Zhang *et al.*, 2008). On the other hand, two functional genes (*TSB1,* At5G54810 and *TSB2*, At4G27070) are known to encode the beta subunit (Last *et al.*, 1991) in addition to two genes (At5g28237 and At5g38530) encoding the putative enzymes (Yin *et al.*, 2010).

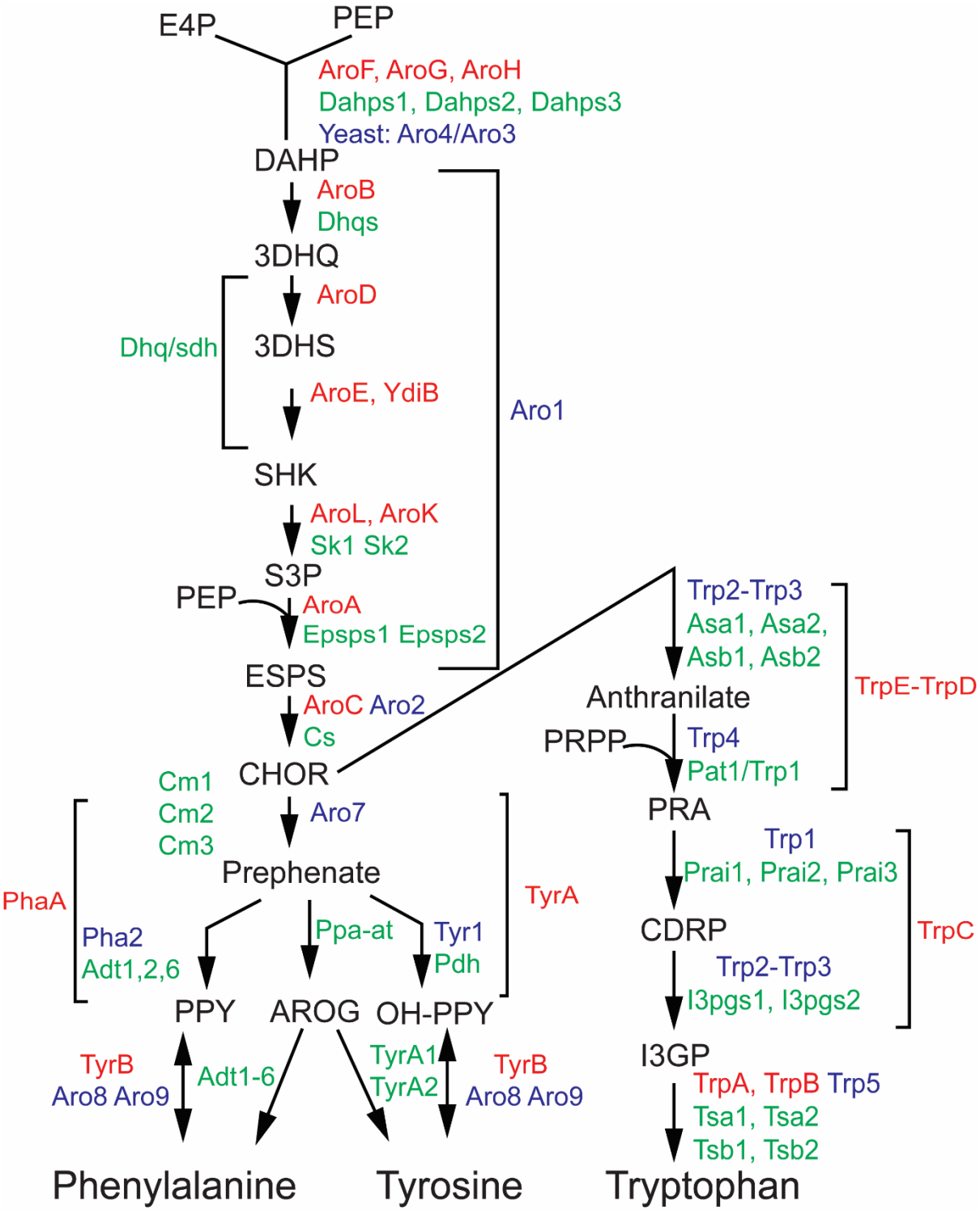


Figure 2. Biosynthesis of aromatic amino acids in plants and microorganisms. *S. cerevisiae* enzymes are shown in blue texts, *E. coli* in red and plant in green. In summary, E4P from pentose phosphate pathway and PEP from glycolytic pathway are the two precursors for aromatic amino acids biosynthesis. Chorismate is biosynthesised in the first part of the pathway, referred to as the shikimate pathway. Chorismate then goes on to serve as a branch point molecule from which each aromatic amino acid is produced. Chorismate is converted into prephenate towards the phenylalanine/tyrosine branch and to anthranilate towards the tryptophan branch. Detailed description of the pathway and enzymes catalysing reactions are given in the body text. Metabolite abbreviations: PEP, phosphoenolpyruvate; E4P, erythrose 4-phosphate; DAHP, 3-deoxy-D-arabino-heptulosonate-7-phosphate; 3DHQ, 3-dehydroquinate; 3DHS, 3-dehydroshikimate; SHK, shikimate; S3P, shikimate 3-phosphate; ESPS, 5-enolpyruvoyl-shikimate 3-phosphate; CHOR, chorismate; PPY, phenylpyruvate;

AROG, arogenate; OH-PPY, 4-hydroxyphenylpyruvate; PRA, phosphoribosyl anthranilate; PRPP, phosphoribosyl pyrophosphate; CDRP, carboxyphenylamino-1-deoxyribulose 5-phosphate; I3GP, indole 3-glycerol phosphate. Enzymes in green texts are from plants, those in red are from bacteria and in blue are from yeasts.

3. Yeasts as suitable cell factories for producing aromatics

With plants representing sources of most valuable aromatics, it is reasonable to combine SynBio and plant tissue culture for aromatics production (Chandran *et al.*, 2020). Although plants cells can, in theory, be metabolically engineered for increased aromatics production, the fact that plant genetic engineering has seen lesser advancement than microbial genetic engineering and the purification challenge, which is not a problem in microbial cell factories, are obvious obstacles to the application of plants as cell factories. Furthermore, strict regulation of aromatics biosynthesis, hinted by the fact that they are naturally biosynthesised at very low quantities in plants, means that detailed understanding of plant cell physiology in this regard, is required to achieve higher production titres (Courdavault *et al.*, 2021). The knowledge of physiology required may be specific for each aromatic molecule one desires to produce. Microorganisms serve as better hosts because they do not naturally produce many plant metabolites and can be engineered to solely produce compounds of interest which can make downstream process less challenging. Furthermore, regulations that limit biosynthesis in plant cells may be minimal in microbial host or even absent.

Deciding which microorganism is best suited for this purpose is not a trivial task and relies on available knowledge of the microorganism's biology. Bacteria and yeasts are widely developed as microbial cell factories due to the wealth of knowledge available on them. Although aromatics have been successfully produced in both microorganisms, the requirement for functional expression of membrane-bound proteins, which are distributed across the biosynthetic pathways of aromatics means that yeasts are best suited for producing aromatics (Liu and Nielsen, 2019). While this problem may be circumventable in bacteria in some cases, such as in the biosynthesis of phenylpropanoids, a large family of plant aromatics, from tyrosine as precursor instead of from phenylalanine which requires a cytochrome P450 enzyme (Hwang *et al.*, 2003; Watts, Lee and Schmidt-Dannert, 2004), it may not be possible to do so for the biosynthesis of many valuable aromatics. The decoration phenylpropanoids with hydroxyl functional groups by hydroxylases to form hydroxylated derivatives with attractive properties as later discussed requires the binding of these enzymes to the endoplasmic reticulum membrane (Cao *et al.*, 2020). The absence of this organelle in bacteria could limit the hydroxylation. The presence of endoplasmic reticulum in yeasts means that hydroxylation reactions could be more efficient in yeasts.

Given that its biology is well understood, it is genetically tractable and its GRAS (generally regarded as safe) status, *S. cerevsiae* is the most common yeast developed as a cell factory to produce aromatics and many other plant natural products. However, the last decade has ushered in genetic engineering tools that are now allowing the development of other yeasts species as cell factories. This is driven by specific phenotypes that they possess but are not present in *S. cerevisiae*. For example, the yeast *Kluyveromyces marxianus* is thermotolerant, fast growing and can grow on diverse substrates (Lane *et al.*, 2011; Lertwattanasakul *et al.*, 2015; Morrissey *et al.*, 2015). The ability of *K. marxianus* to grow faster than *S. cerevisiae* can reduce production time and the respiratory physiology of this yeast can be advantageous in that the direction of carbon towards ethanol production, which can reduce the yield on glucose of non-ethanol products, is lower than in *S. cerevsiae.* High flux through the pentose phosphate pathway was reported in *K. marxianus* (Jung *et al.*, 2016). This is a potentially beneficial metabolic feature of this yeast regarding precursor supply for aromatics production. *Yarrowia lipolytica* is an oleaginous yeast known for its superior ability to biosynthesise fatty acids which are derived from malonyl-CoA, a highly important metabolite necessary for producing some valuable plant flavonoids (Madzak, 2021). Thus, non-saccharomyces yeasts are being developed for producing various plant aromatics

4. Metabolic engineering of aromatic amino acids biosynthesis in yeasts

4.1. Engineering the shikimate pathway

In addition to the feedback inhibition of enzymes of the shikimate pathway for AAA biosynthesis discussed above, transcriptional regulation of the genes encoding key enzymes in the pathway also occurs (Cao *et al.*, 2020). While these regulations ensure that metabolic flux in cells is not overwhelmingly towards the shikimate pathway for proper cell functioning, they impede the production of aromatics in microbial cell factories due to inadequate precursor supply. Therefore, metabolic engineering of the shikimate pathway is required in developing microorganisms as cell factories for aromatics production. The negative transcriptional regulation of the genes encoding AAA biosynthetic enzymes is mediated by Ric1. Based on results that the deletion of *RIC1* in *S. cerevisiae* increased the transcription of *ARO1, ARO2, ARO3* and *ARO4,* Ric1, in addition to its initially associated function as a GTP exchange factor for Ypt6 and the association of its deletion with poorer thermotolerance (Bensen, Yeung and Payne, 2001), is suggested to be a transcription factor that represses the expression of these genes (Suástegui *et al.*, 2017). Although *RIC1* deletion increased shikimate production by 27% this strategy may not be suitable for increasing flux through the

shikimate pathway because the effect of its deletion on other genes and overall cell function is not fully understood. Therefore, a more widely used approach for increasing flux through the pathway is the overexpression of the genes encoding the enzymes of the pathway. This involves the deregulation of the entry point into the shikimate pathway through the overexpression of feedback resistant (FBR) DAHP synthases. This strategy is routinely employed in *S. cerevisiae* by overexpressing $ARO4^{FBR}$ (G226S, K229L and Q166K) to render it insensitive to inhibition by tyrosine with or without the deletion of the phenylalanine sensitive *ARO3* (Koopman *et al.*, 2012; Rodriguez *et al.*, 2015; Trenchard and Smolke, 2015), while in bacteria, $aroG^{FBR}$ and $aroF^{FBR}$ overexpression are adopted to render these enzymes insensitive to phenylalanine and tyrosine, respectively (Chandran *et al.*, 2003; Kogure *et al.*, 2016; Lee, Hung and Tsai, 2017). The possibility that genes and enzymes from bacteria, which are usually functionally expressible in eukaryotes, may possess certain advantages, has driven some researchers to bring bacterial genes into yeasts in recent years. With the overexpression of $aroG^{FBR}$ (P150L) in *S. cerevisiae* that is already expressing $ARO4^{FBR}$ (K229L) in addition to other modifications proving redundant, the most marked success with this approach has been reported for the expression of bacterial genes that are responsible for individually catalysing each of the five reactions from DAHP to ESPS. Catalysed by the pentafunctional Aro1 enzyme in yeasts, the tunning of the individual enzyme activity to alleviate bottlenecks in converting DAHP to EPSP will be a challenging task because it will require extensive engineering of the enzyme. The individual overexpression of the five genes from *E. coli* in *S. cerevisiae* strains with *ARO1* under native or deregulated transcription does not increase flux through the pathway, with the exception of shikimate kinase *aroL* (Rodriguez *et al.*, 2015; Hassing *et al.*, 2019). This suggests that the phosphorylation of shikimate into shikimate 3-phosphate by the shikimate kinase domain of Aro1 is a bottleneck in the pathway and the overexpression of *aroL* should be considered as a strategy for increasing flux through the pathway. However, the overexpression of *ARO1* mutant (D920A) has also been successfully used to increase shikimate production (Suástegui *et al.*, 2017).

4.2. Precursor supply: engineering pentose phosphate pathway to increase erythrose 4-phopsphate

The supply of precursors from central metabolic pathways to produce the first metabolite in the shikimate pathway (DAHP) is schematised in figure 3. As erythrose 4-phosphate (E4P) is one of the two precursors of the shikimate pathway and is produced in the pentose phosphate pathway (PPP), the metabolic engineering of the PPP for E4P production is commonplace for increasing AAA biosynthesis due to evidence that *S. cerevisiae* DAHP synthases showed lower affinity for E4P than the second precursor, phosphoenolpyruvate (Schnappauf *et al.*, 1998) and metabolic flux analysis revealed that E4P is the limiting

precursor in *S. cerevisiae* (Suástegui *et al.*, 2016). This is achieved by overexpressing genes of the non-oxidative phase of the PPP including *TKL1, TAL1, RKI1 and RPE1,* encoding transketolase, transaldolase, ribose 5-phosphate ketol-isomerase and ribulose 5-phosphate 3-epimerase, respectively. The overexpression of native *TKL1* and *TAL1* genes individually or together increased the production of aromatic alcohols in *S. cerevisiae* and *K. marxianus* (Hassing *et al.*, 2019; Rajkumar and Morrissey, 2020). Although *RKI1* overexpression was predicted *in silico* and confirmed experimentally to be a target for increasing shikimate production in *S. cerevisiae* (Suástegui *et al.*, 2017), the additional overexpression of *RKI1* and *RPE1* in *K. marxianus* did not further increase the AAA production, indicating that bottlenecks exist below the enzymes (Rajkumar and Morrissey, 2020), but it is also possible that their overexpression independently of *TKL1* and *TAL1* overexpression may not be a viable target for increasing E4P in *K. marxianus*.

As the central carbon metabolism-derived precursor of the non-oxidative phase of PPP, fructose 6-phosphate (F6P) availability can limit E4P biosynthesis. Thus, increasing the availability of F6P has also been adopted as a strategy for increasing E4P. The heterologous expression of phosphoketolase gene from *Bifidobacterium breve* (*Bbxfpk*) which converts fructose 6-phosphate from the glycolytic pathway to E4P and acetyl-phosphate was attempted for increasing E4P availability in combination with bacterial phosphotransacetylase (*pta*) for the subsequent conversion of acetyl-phosphate into acetyl-CoA, which can also enhance the production of acetyl-CoA derived products like polyketides (discussed below). This strategy improved flux through the shikimate pathway in *S. cerevisiae* as indicated by a 10% increase in the production of coumaric acid, an aromatic amino acid-derived molecule, but no improvement was observed in combination with the overexpression of native *TKL1* (Hassing *et al.*, 2019; Q. Liu *et al.*, 2019). Similar result was reported in *K. marxianus* where the expression of *Salmonella enterica pta* alongside *Bbxfpk* produced similar cumulative titre of aromatic alcohols and shikimate as the strain overexpressing *TAL1* and *TKL1* (Rajkumar and Morrissey, 2020). These results showed that the combination of this strategy and overexpression of the native PPP genes did not increase E4P supply. Although Gu and colleagues reported increase in aromatic alcohol production when bacterial phosphoketolases were expressed in *Yarrowia lipolytica*, these authors did not compare their result to production when genes of the non-oxidative phase of the PPP were additionally overexpressed (Gu, Ma, Zhu, Ding, *et al.*, 2020). F6P increase has also featured the deletion of glucose 6-phopsphate dehydrogenase *ZWF1*, which blocked the oxidative phase of the PPP, causing NADPH deficiency and severe growth defect (Curran *et al.*, 2013), and recently by the transcriptional repression of phosphofructokinase *PFK1/2* through expression from weaker promoters (Q. Liu *et al.*, 2019). Since xylose is assimilated via the PPP, it was

anticipated that non-saccharomyces yeasts that are naturally able to metabolize this carbon source could have higher availability of intracellular E4P. In line with that hypothesis, *Scheffersomyces stipitis* engineered by simply overexpressing the native transketolase gene (*TKT1*), *ARO4*^{FBR} and mutant *ARO1* (D900A) was reported to produce shikimate at a titre 7-fold higher than produced by *S. cerevisiae* (Gao *et al.*, 2017).

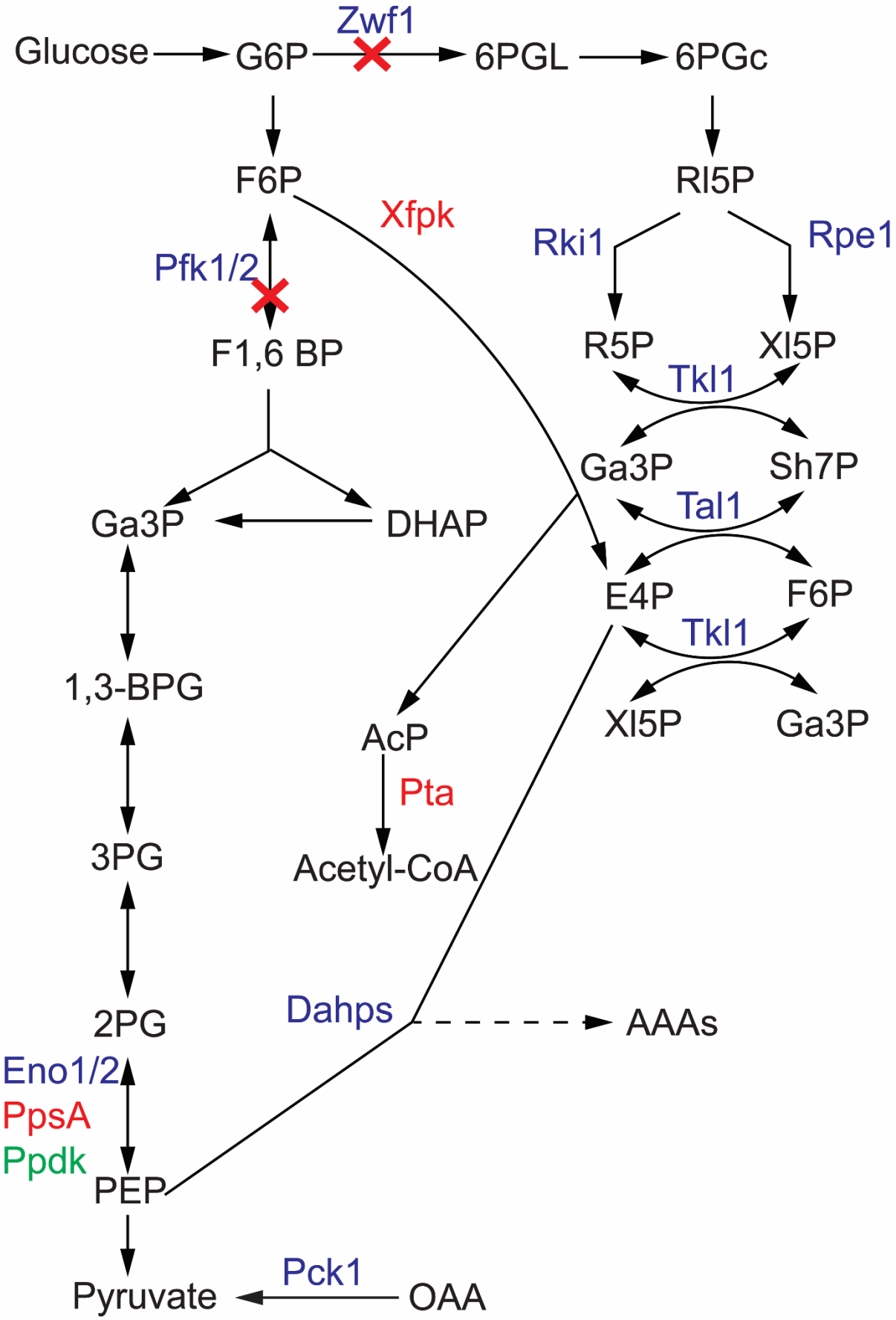


Figure 3. Supply of phosphoenolpyruvate and erythrose 4-phosphate precursors for aromatic amino acids biosynthesis in yeasts. Metabolite abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-

phosphate; F-1,6BP, fructose 1, 6-bisphosphate; Ga3P, glyceraldehyde 3-phosphate; DHAP, dihydroacetone phosphate; 1,3-BPG, 1, 3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; α-KG, α-ketoglutarate; Glu, Glutamate; Gln, glutamine; 6PGL, 6-phosphogluconolactone; 6PGc, 6-phosphogluconate; Rl5P, ribulose 5-phosphate; Xl5P, xylulose 5-phosphate; R5P, ribose 5-phosphate; Sh7P, sedoheptulose 7-phosphate; E4P, erythrose 4-phosphate; Enzyme abbreviations: Tkl1, transketolase; Tal1, transaldolase; Rki1, ribose-5-phosphate ketol-isomerase; Rpe1, ribulose-5-phosphate 3-epimerase; Xfpk, phosphoketolase; pta, phosphotransacetylase; Eno1/2, enolase; PpsA, PEP synthase; Ppdk, pyruvate orthophosphate dikinase; Zwf1, glucose 6-phopsphate dehydrogenase; Pfk1/2, phosphofructokinase; Pck1, PEP carboxykinase, Dahps, DAHP synthase. Enzymes in green texts are from plants, those in red are from bacteria and in blue are from yeasts. Reactions with Red “X” have been deleted or attenuated in order to increase E4P supply.

4.3. Precursor supply: engineering glycolytic pathway for phosphoenolpyruvate

The other precursor entering the shikimate pathway is phosphoenolpyruvate (PEP) which is produced in yeasts by the action of enolase *ENO1/2* on phosphoglycerate during glycolysis and by PEP carboxykinase *PCK1* during gluconeogenesis. PEP is also reversibly produced from pyruvate in bacteria and plants by the action of PEP synthase (PEPS) and pyruvate orthophosphate dikinase (PPDK). Although the increase of this precursor was widely sidetracked due to E4P being the limiting precursor, once E4P increase is engineered, PEP can become limiting. Therefore, increased PEP biosynthesis is underrepresented in the literature as a strategy for increasing precursor supply in yeasts. Recent success in increasing E4P led researchers to turn their attention to engineering PEP availability by increasing its biosynthesis and reducing its conversion by pyruvate kinase to pyruvate. With native PEP synthase gene (*ppsA*) commonly overexpressed for recycling pyruvate back into PEP in *E. coli* (Patnaik and Liao, 1994; Juminaga *et al.*, 2012; Wei, Cheng and Liu, 2016), recent attempt to increase PEP biosynthesis in *K. marxianus* by overexpressing this gene and genes encoding PEP biosynthetic enzymes from various organisms showed that the expression of *E. coli,* archaeal and fungal PEPS, PPDK from *A. thaliana,* and the overexpression of native *ENO1* all increased 2PE production (Rajkumar and Morrissey, 2020). The indispensable role of pyruvate means that increasing PEP availability for the shikimate pathway by deleting pyruvate kinase *CDC19/PYK2* would be detrimental to growth (Zelle *et al.*, 2010). However, (Q. Liu *et al.*, 2019) used a promoter screening library to identify a suitable promoter that allowed just enough pyruvate kinase expression for growth while ensuring optimal PEP supply to the shikimate pathway. Furthermore, the expression of mutant *PYK1* by (Hassing *et al.*, 2019) resulted in growth defect but increased 2PE production. These studies demonstrated that reduced conversion of PEP to pyruvate increased PEP supply to the shikimate pathway.

4.4. Directing chorismate towards the biosynthesis of specific aromatic amino acids

The fact that chorismate is a central metabolite that feeds into several pathways indicates that its uptake into those pathways is tightly regulated. As demonstrated above, the direction of chorismate towards the phenylalanine/tyrosine branch by chorismate mutase and the tryptophan branch by anthranilate synthase are under allosteric negative-feedback inhibition by the respective amino acids to avoid overdrive towards one route. Therefore, to engineer the direction of chorismate towards the phenylalanine/tyrosine route, feedback resistant chorismate mutase *ARO7* (G141S, T266I or T227L) is commonly overexpressed for the biosynthesis of prephenate, the precursor of tyrosine and phenylalanine. Since the biosynthesis of both AAAs compete for prephenate, further engineering strategies are established to direct prephenate towards either amino acid. It is known that the expression prephenate dehydrogenase Tyr1 is inhibited in the presence of phenylalanine (Mannhaupt *et al.*, 1989). It was reported that the heterologous expression of *TYR1* variant from *Medicago truncatula* (barrel medic), in order to partially bypass inhibition by phenylalanine increased the production of the tyrosine derivative, coumaric acid, by 30% (Q. Liu *et al.*, 2019). However, in their efforts to produce 2PE in *K. marxianus*, overexpression of the native prephenate dehydratase *PHA2* alongside *ARO9* by Rajkumar and Morrissey only marginally increased the titre (Rajkumar and Morrissey, 2020). The persistent production of tyrosol, the aromatic alcohol produced from tyrosine, showed that the increased prephenate biosynthesis achieved by overexpressing the shikimate pathway and *ARO7* was being directed towards tyrosine. This caused the authors to replace the promoter of *TYR1* with a weaker promoter, resulting in reduced tyrosol and increased 2PE. This *TYR1* promoter replacement strategy was also successfully applied for increasing 2PE production in *S. cerevisiae* (Hassing *et al.*, 2019).

Below Pha2 and Tyr1, the aromatic 2-oxo acid precursors, 4-hydroxyphenylpyruvate and phenylpyruvate are aminated into tyrosine and phenylalanine, respectively. This involves the transamination of the 2-oxo acids and is catalysed by aromatic amino transferase (Aat) enzymes Aro8 and Aro9. With Aro8 mainly converting the aromatic 2-oxo-acids into respective AAAs and Aro9 mainly catalysing the reverse reaction, the two reactions involve the transfer of an amino group from a donor to an acceptor. Although both Aats can catalyse the transfer of amino groups from various amino acids to aromatic 2-oxo acids to form AAAs (aromatic aminotransferase I) and the transfer of amino group from AAAs onto acceptors such as pyruvate, forming the respective aromatic 2-oxo-acid in the process (aromatic aminotransferase II), Aro8 is the main AATI and Aro9 the main AATII (Urrestarazu *et al.*, 1998). Indeed, *ARO8* deletion contributed positively to 2PE production in *S. cerevisiae* (Romagnoli, 2015) and the overexpression of Streptomyces glutamate oxidase was used to

increase intracellular levels of α-ketoglutarate in order to drive transamination towards phenylpyruvate for 2-phenyethanol production (Zhu *et al.*, 2021). The deletion of *ARO9* may also be considered for producing foreign aromatics because in contrast to the biosynthesis of aromatic alcohols which takes aromatic 2-oxo acids as direct precursors (discussed below), foreign aromatics take AAAs.

An approach for engineering yeasts for tryptophan production is the overexpression of $TRP2^{FBR}$ (S65R and S76L) to increase the direction of chorismate towards tryptophan (Graf, Mehmann and Braus, 1993). The requirement of this reaction for glutamine as amino donor has recently driven some authors to attempt the increase of glutamine by overexpressing *GLN1* in *S. cerevisiae,* which doubled the production of anthranilate (Kuivanen *et al.*, 2021). Although the disruption of chorismate mutase activity cannot be applied as a strategy for alleviating competition from the phenylalanine/tyrosine route as it can result in AAA auxotrophy, the repression of *ARO7* transcription may contribute to tryptophan production from chorismate.

5. Aromatic alcohols and aldehydes are the degradation products of aromatic amino acids via the native Ehrlich pathway

The increase in flux towards yeast aromatic amino acid biosynthetic pathway by means of the strategies described above generally increases the production of aromatic alcohols because freely available AAAs are degraded into these metabolites. Aromatic amino acids are degraded in the Ehrlich pathway through their deamination into aromatic 2-oxo-acids by AATII. This initial reaction in the pathway is driven by the thermodynamic pull created by the ability of yeast cells to keep intracellular aromatic 2-oxo-acids concentration low by means of their decarboxylation by phenylpyruvate decarboxylase Aro10 into aromatic aldehydes which are in turn further reduced or oxidized into aromatic acids or alcohols, depending on cellular redox status (Hazelwood *et al.*, 2008). This pathway degrades phenylalanine to 2-phenylethanol, tyrosine to tyrosol and tryptophan to tryptophol as illustrated in figure 4.

Although they can be quantified as proxies for detecting increase in intracellular AAAs, the usefulness of aromatic alcohols goes beyond this because of their strong, distinct, and pleasant smells. This makes them applicable as flavouring and fragrance agents. 2-phenylethanol has received a lot of attention, leading to the highest reported titre currently in *Y. lipolytica* at 22mM with complex media (Gu, Ma, Zhu and Xu, 2020), but titres of 13mM in *S. cerevisiae* (Hassing *et al.*, 2019) and even lower in *K. marxianus* (Kim, Lee and Oh, 2014; Rajkumar and Morrissey, 2020) have also been reported with minimal medium. Although tyrosol production is observed in the efforts to engineer 2-phenylethanol production,

dedicated engineering of tyrosol in yeasts has so far only been in *S. cerevsiae* with total titres of tyrosol and its glucoside derivative (salidroside) as high as 10 g/L (Guo *et al.*, 2020). In addition to the widely used strategies for increasing the flux of AAAs pathway, tyrosol engineering has been done by overexpressing *EctyrA*^{FBR} in combination with *PDC1* deletion to redirect carbon flux at the pyruvate branch point and alleviate carbon lose to alcoholic fermentation, and also by eliminating the direction of chorismate to phenylalanine and tryptophan by inactivating *PHA2* and *TRP3,* respectively (Guo *et al.*, 2019, 2020). Tyrosol derivatives was produced by the additional expression of bacterial FAD-dependent 4-hydroxyphenylacetate-3-monooxygenase to produce hydroxytyrosol (Bisquert *et al.*, 2022; H. Liu *et al.*, 2022) and *A. thaliana* UDP-glycosyltransferase ugt85a1 to glycosylate tyrosol and produce salidroside (Guo *et al.*, 2020).

The ability of yeasts to naturally degrade AAAs into aromatic alcohols via the Ehrlich pathway warrants that the engineering of yeasts to produce heterologous aromatics generally requires additional engineering of reduced flux through the pathway because it competes against heterologous pathways for the AAA precursors. The major role played by yeast native phenylpyruvate decarboxylase activity in driving flux through the pathway means that the pathway flux is successfully reduced by simply attenuating phenylpyruvate decarboxylase activity without having to modify the enzymes responsible for catalysing the transamination and oxidoreduction reactions (figure 4).

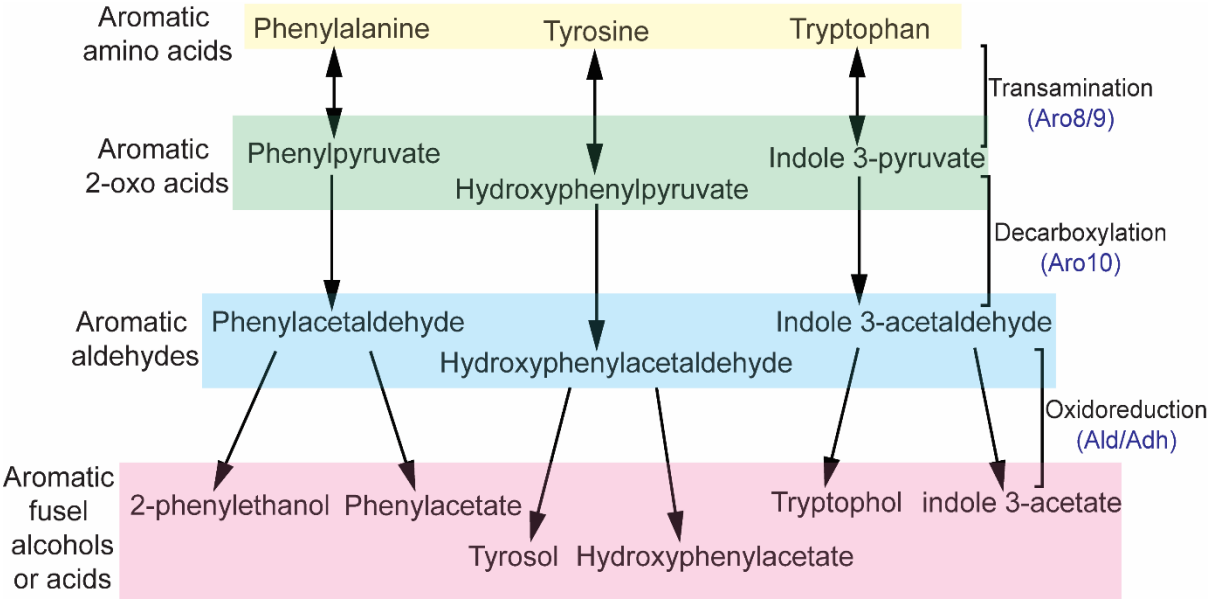


Figure 4. Degradation of aromatic amino acids in the Ehrlich pathway. Aromatic amino acids (in yellow background) are deaminated by the action of aromatic amino transferase II reaction. This reaction is catalysed by yeast Aro8 and Aro9 and forms aromatic 2-oxo acid products (in green background). This is followed by the decarboxylation of aromatic 2-oxo acids into their respective aromatic aldehyde (in blue background), mainly catalysed by phenylpyruvate decarboxylase Aro10. Aromatic aldehyde can then be oxidised or reduced into aromatic fusel alcohols or acids (in pink background) by alcohol/aldehyde dehydrogenases.

6. Production of phenylpropanoids in yeasts

Phenylpropanoids are a large and structurally diverse group of aromatic plant secondary metabolites produced from phenylalanine and tyrosine. They are so-called due to the presence of a phenyl (6 carbon) ring and three carbon aliphatic chain of coumaric acid, the central metabolite from which they are produced. Phenylpropanoids are structurally classified into flavonoids, monolignols, phenolic acids, stilbenes and coumarins. The biosynthesis of phenylpropanoids begins with the non-oxidative deamination of tyrosine directly into coumaric acid, or phenylalanine indirectly into cinnamic acid and subsequently into coumaric acid. These deamination reactions are catalysed by AAA ammonia lyases, tyrosine ammonia lyase (Tal) and phenylalanine ammonia lyase (Pal). In the phenylalanine route, cinnamic acid 4-hydroxylase, with the help of cytochrome P450 reductase, catalyses the hydroxylation of cinnamic acid into coumaric acid. These two metabolites can be further converted into other phenylpropanoids as illustrated in figure 5 and further discussed below.

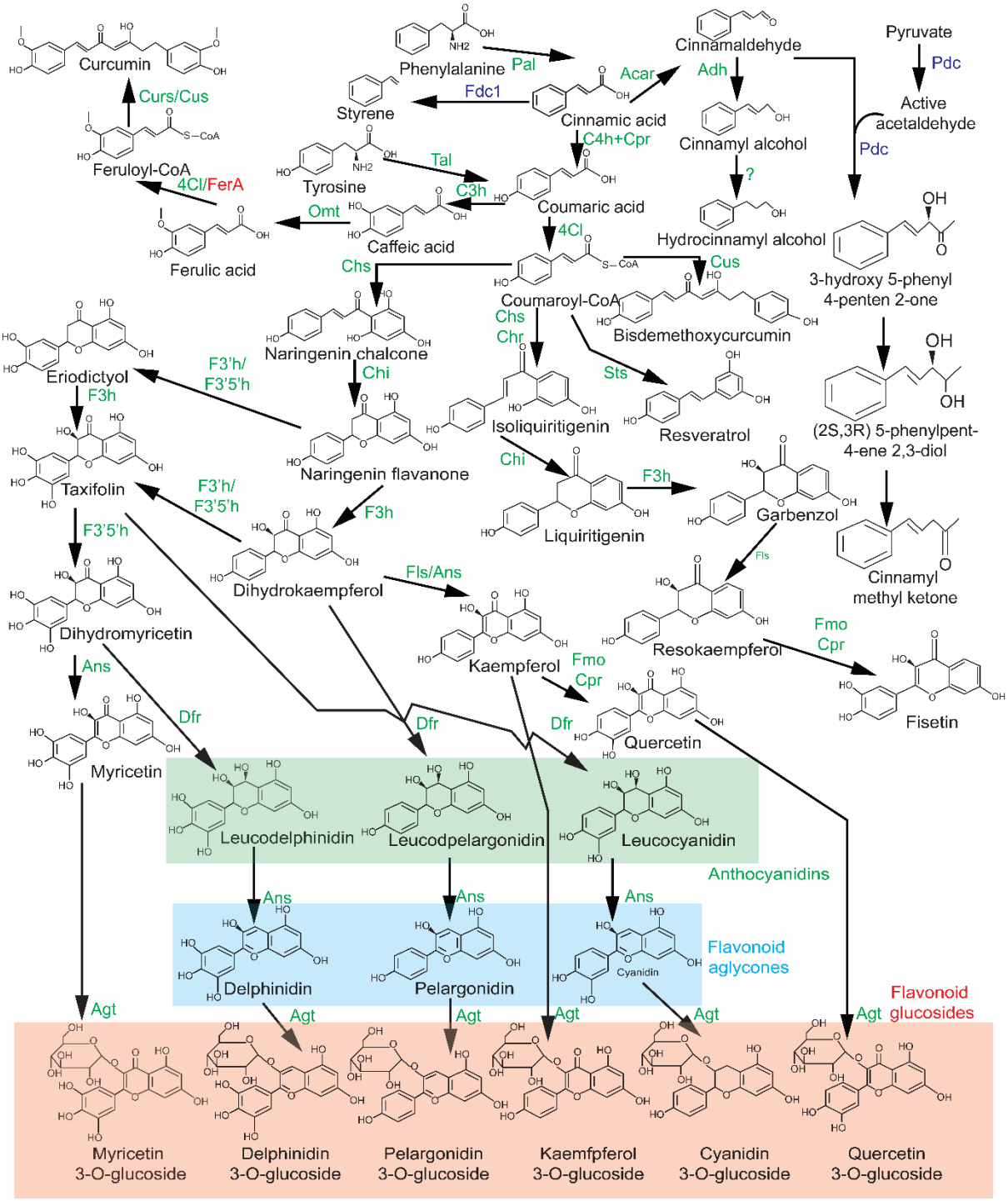


Figure 5. Biosynthesis of phenylpropanoids. Tyrosine and phenylalanine are the biological precursors of phenylpropanoids from which cinnamic and coumaric acids are biosynthesised by the action of tyrosine and phenylalanine ammonia lyases (Tal and Pal). Coumaric acid is produced via two alternative routes: in the first, Tal directly deaminates tyrosine into coumaric acid while the second involves the deamination of phenylalanine into cinnamic acid by Pal, followed by cinnamic acid hydroxylation by cinnamic acid 4-hydroxylase (C4h) and cytochrome P450 reductase (Cpr) to form coumaric acid. The production of more complex phenylpropanoids from cinnamic and coumaric acids is also illustrated. Aryl carboxylic acid reductase (Acar) and yeast native ferulic acid decarboxylase (Fdc1) catalyse cinnamic acid into cinnamaldehyde and styrene, respectively in yeast. Cinnamyl alcohol and hydrocinnamyl alcohol are cinnamaldehyde derivatives that have also been detected in yeast expressing Acar and yeast native Pdc activity produces cinnamyl methyl ketone from cinnamaldehyde. Coumaric acid is a more important precursor of phenylpropanoids than cinnamic acid

because a larger number of molecules are produced from coumaric acid. Thus, it is attractive to engineer yeasts for coumaric acid production with the expectation that it can be derivatised into other phenylpropanoids. Coumaric acid is converted into coumaroyl-CoA by coumaric acid:coenzyme A ligase (4Cl) or hydroxylated by coumaric acid 3-hydroxylase (C3h) into caffeic acid. The O-methylation of caffeic acid by O-methyltranferase forms ferulic acid which is ligated with coenzyme A by 4Cl or bacterial feruloyl-CoA synthetase (FerA) to form feruloyl-CoA. Coumaroyl-CoA and feruloyl-CoA then serve as precursors from which other phenylpropanoids are produced, including the polyketides curcumin, naringenin, liquiritigenin resveratrol and bisdemethoxycurcumin. Polyketides biosynthesis requires the action of polyketide synthase enzymes that catalyse the successive decarboxylative condensation of derivative units of coenzyme A. Chalcone synthase (Chs) forms naringenin or, in addition to chalcone reductase (Chr), forms liquiritigenin, stilbene synthase (Sts) forms resveratrol, curcuminoid/curcumin synthase (Curs/Cus) forms curcumin and bisdemethoxycurcumin. Polyketides can be derivatised into various hydroxylated derivatives by plant flavonoid hydroxylases (F3-h, F3'-h, F3'5'-h) and cytochrome P450 flavonoid monooxygenase (Fmo). Phenylpropanoids with even greater complexities can be produced from these hydroxylated derivatives by the action of dihydroflavonol 4-reductase (Dfr) to form anthocyanidins (in the green background) which are further converted into flavonoids aglycones by anthocyanidin synthase (Ans) (in the blue background) and then the addition of a glucose moiety by anthocyanin 3-O-glucosyltransferase (Agt) to aglycones forms flavonoid glucosides which are valuable as potential active pharmaceutical ingredients (in the orange background). Enzymes in green texts are from plants, those in red are from bacteria and in blue are from yeasts.

6.1. Phenylpropanoids that are derived from cinnamic acid

Cinnamic acid is itself a flavouring and fragrance agent used in foods and cosmetics in addition to some health benefits and is therefore valuable (Cao *et al.*, 2020). It is also the precursor for some other phenylpropanoids including cinnamaldehyde, hydrocinnamyl alcohol, cinnamyl alcohol and styrene. In an attempt to achieve *de novo* production of cinnamaldehyde, hydrocinnamyl alcohol and cinnamyl alcohol in *S. cerevisiae, A. thaliana PAL2*, aryl carboxylic acid reductase from Nocardia species and phosphopantetheinyl transferase (*entD*) from *E. coli* were expressed (Gottardi, Knudsen, *et al.*, 2017). The authors reported that while the strategy produced cinnamaldehyde, this metabolite was further reduced into cinnamyl alcohol by endogenous aldehyde dehydrogenase activity and then cinnamyl alcohol was converted by an unknown enzyme into hydrocinnamyl alcohol which made up the highest titre of the three cinnamic acid derivatives (Gottardi, Knudsen, *et al.*, 2017). In light of the fact that cinnamyl alcohol and hydrocinnamyl alcohol were less toxic on *S. cerevisiae* than cinnamic acid and cinnamaldehyde, it is possible that cinnamyl alcohol and hydrocinnamyl alcohol were produced in order to detoxify the other compounds. Furthermore, an unidentified compound whose peak area increased concomitantly with decreasing cinnamyl alcohol concentration over time in their initial study was subsequently identified as cinnamyl methyl ketone by the same research group (Gottardi, Grün, *et al.*, 2017). Cinnamyl methyl ketone is a product of carboligation reaction of cinnamaldehyde with active acetaldehyde mediated by pyruvate decarboxylase (Pdc), providing a route for biosynthesising the precursors for chemical synthesis of biologically active molecules such as the anti-cancer drug daunomycin. Cinnamaldehyde can be alternatively biosynthesised via cinnamate ligation to coenzyme A which is catalysed by 4-coumaric acid:coA ligase.

Cinnamoyl-CoA reductase then reduces cinnamoyl-CoA to cinnamaldehyde. This route is yet to be demonstrated in yeasts but has been applied for cinnamaldehyde production in bacteria (Bang *et al.*, 2016). The engineering of *S. cerevisiae* to produce styrene, the monomeric precursor of polystyrene from cinnamic acid has also been reported (McKenna *et al.*, 2014). This was achieved by co-expressing *AtPAL2* and yeast native ferulic acid decarboxylase *FDC1* gene, resulting in yield of 1.4 mg/g glucose.

6.2. Phenylpropanoids that are derived from coumaric acid

Most phenylpropanoids are produced from coumaric acid. Therefore, efforts on *de novo* production of phenylpropanoids in yeast have, in recent years, focused on the production of coumaric acid with the hope that this branch point molecule can be directed towards the biosynthesis of other phenylpropanoids, referred to in this review as the "coumaric acid platform approach". Its production in *S. cerevisiae* at 1.9g/L by simply overexpressing *Flavobacterium johnsoniaeu TAL* (*FjTAL*) in a strain with enhanced AAAs biosynthesis represented the first report of *de novo* coumaric acid production at such high titre (Rodriguez *et al.*, 2015). This coupled with the fact that coumaric acid production through the tyrosine route is particularly favourable because only a single gene is needed compared to three genes for the phenylalanine route encouraged others to attempt to produce coumaric acid and its derivatives via the tyrosine route by overexpressing *TAL* which has recently resulted in coumaric acid tire of 12.5g/L (154.9 mg/g glucose) to be produced in *S. cerevisiae* (Q. Liu *et al.*, 2019), representing the highest production reported for so far and it would be interesting to see how well this translates to the production of coumaric acid derivatives.

A concern in metabolically engineering is host response to the modifications and target compounds (Nielsen and Keasling, 2016). Systems biology based on omics-level characterization can be used to guide metabolic engineering decisions. Omics analysis of metabolically engineered *S. cerevisiae* strains suggested that the ability of CEN.PK to produce higher coumaric acid than S288C could be linked to difference in global gene expression and metabolites production in response to modifications (Rodriguez, Chen, *et al.*, 2017). The fact that CEN.PK showed fewer changes in transcriptional profile and intracellular metabolites concentrations than S288C suggested that CEN.PK is a more robust strain for producing coumaric acid and perhaps its derivatives. Interestingly, the enhancement of AAAs biosynthesis in both strains showed the greatest transcriptional changes by way of downregulating the genes that are involved in amino acid and sugar transport. Of these genes, the successful deletion of *TAT1, TPO1, ALP1, ADY1, GAL2* appeared to increase coumaric acid production in a CEN.PK strain with enhanced AAAs biosynthesis. Motivated by the fact that coumaric acid acts as growth and fermentation inhibitor of *S. cerevisiae* in

lignocellulosic hydrolysates, a recent multi-omics study elucidated how a Brazilian fuel ethanol industrial strain (SA-1) with high resistance to coumaric acid inhibition possesses this phenotype (Ciamponi *et al.*, 2022). This study provided differentially regulated genes that can serve as engineering targets for improving yeast tolerance to coumaric acid, especially in strategies where coumaric acid is fed to yeasts as starting materials, but also in the *de novo* production strategies where coumaric acid produced is used up by downstream enzymes for producing its derivatives.

An alternative to the "coumaric acid platform approach" is a wholistic approach where researchers set out to produce specific phenylpropanoids without the initial optimisation of the production of coumaric acid and other intermediates. This has resulted in successful production of various coumaric acid derivatives both *de novo* and by precursor feeding. Caffeic and ferulic acids and important polyketides like naringenin and resveratrol are examples of these derivatives. Even though the improvement of coumaric acid production in microbial cell factories is gradual, yeasts have been and are being successfully engineered to produce these derivatives, albeit in low quantities.

Caffeic acid has been linked to several health benefits including anticancer, antimicrobial, antiviral and antioxidant activities in addition to serving as precursor for ferulic acid which has similar bioactivities and is a hypoglycaemic and hypolipidemic agent (Monteiro Espíndola *et al.*, 2019; Stompor-Gorący and Machaczka, 2021). Caffeic acid is produced through the hydroxylation at the 3rd carbon position of the phenyl ring of coumaric acid by coumaric acid 3-hydroxylase. Isoforms of these enzymes have been identified in bacteria and plants and were expressed in *S. cerevisiae* for *de novo* and tyrosine supplemented production of caffeic acid production at 2.7mg/L and 289.4 mg/L, respectively (Liu *et al.*, 2019; Li *et al.*, 2020). Caffeic acid is further converted into ferulic acid by the action of O-methyltransferase (Omt). The dependence of the conversion of coumaric acid to caffeic acid on NADPH co-factor supply, which was largely ignored in the engineering of yeasts for the production of caffeic acid and its derivatives, recently motivated the engineering of co-factor supply for the production of these molecules in *S. cerevisiae* and was reported to produce caffeic acid titre of 5.5g/L and ferulic acid at 3.8g/L (Chen *et al.*, 2022).

Yeast engineering for polyketides production is attracting a lot of attention because polyketide are important intermediates to plant secondary metabolites that have potential applications as bioactive molecules. In particular, efforts by various research groups to produce naringenin and resveratrol have been seen in recent years. Naringenin is the precursor of flavonoids, which are a large family of molecules that are widely useful in many industries and resveratrol is known for its numerous health benefits. Central to the production of polyketides are polyketide synthases (PKSs). PKSs biosynthesise polyketides through successive

decarboxylative condensation of derivative units of coenzyme A (CoA). They are classified into types I, II and III. Type I PKSs are large proteins that have multiple domains that arrange into several modules with each catalysing a single decarboxylative condensation (modular type I PKS) or are clustered as a single module so that each domain repeatedly catalyses the condensations (iterative type I PKS) (Lim, Go and Yew, 2016). Type II PKSs are complexes of multiple enzymes with each harbouring a single and independent catalytic domain that is capable of iterative decarboxylative condensations while type III PKSs are homodimeric enzymes that conduct the iterative decarboxylative condensations and cyclization in a single active site.

Triacetic acid lactone (TAAL) is a simple but valuable polyketide that can be subjected to various reactions to derive useful molecules like sorbic acid and 1, 3-pentadiene for replacing several petroleum-derived molecules (Chia *et al.*, 2012). TAAL is naturally produced in *Gerbera hybrida* (barberton daisy) from which the type III PKS 2-pyrone synthase (2-PS) that is widely used for the heterologous production of TAAL in microorganisms was obtained. 2-PS produces TAAL from acetyl-CoA as the starter CoA derivative unit through two decarboxylative condensations using 2 molecules of malonyl CoA as extenders. Due to the higher tolerance of *S. cerevisiae* for TAAL compared to *E. coli*, efforts have been directed at producing this compound in yeasts at titres as high as 10g/L (Cardenas and Da Silva, 2014; Saunders *et al.*, 2015; Vickery *et al.*, 2018). More complex polyketides like naringenin and curcumin are biosynthesised from coumaric acid and ferulic acid, respectively. Coumaric acid and ferulic acid are initially converted into coumaroyl-CoA and feruloyl-CoA, respectively. These activated molecules can then be acted upon by type III PKSs to form the polyketides. However, the production of these complex polyketides in yeasts has been so far considerably lower than TAAL production. Curcumin is a polyketide that is produced in the rhizome of *Curcuma longa* (turmeric) and has documented pharmaceutical benefits. Recently, the expression of *Oryza sativa* (rice) or *C. longa* PKSs (curcuminoid/curcumin synthase) in addition to *At-4CL1* or feruloyl-CoA synthetase FerA from the bacteria *Pseudomonas paucimobilis* in *S. cerevisiae* produced 2.7mg/L of curcumin, which represents the first demonstration of curcumin production from ferulic acid in yeast (Rainha *et al.*, 2022).

Chalcone and stilbene synthases (Chs and Sts) are type III PKSs that convert coumaroyl-CoA into naringenin chalcone and resveratrol, respectively. These enzymes are the most studied type III PKSs and have revealed the presence of a Cys-His-Asn catalytic triad. Their catalysis as illustrated in figure 6 begins with the loading of coumaroyl-CoA onto the enzyme by binding to the sulphydryl group of the cysteine residue of the triad to form the starter unit. This is followed by three iterations of decarboxylative condensations of three malonyl-CoA molecules to the starter unit, assisted by the histidine and asparagine residues. Chs and Sts

can convert coumaroyl-CoA into different substrates (naringenin chalcone and resveratrol, respectively) because Chs catalysis C6 to C1 while Sts catalyses C2 to C7 cyclisation in the coumaroyl triacetyl thioester intermediate. The structural simplicity of type III PKSs makes them suitable candidates for biocatalyst engineering (Cai and Zhang, 2018) and are applied for heterologous production of polyketides in microbial hosts. Although the biosynthesis of naringenin and resveratrol requires 3 molecules of malonyl-CoA which is a highly important metabolite that is also required for proper cell function such as in fatty acid biosynthesis, they have been quite easily produced *de novo* in *S. cerevisiae* without the need to engineer malonyl-CoA biosynthesis, although at low titres. Early report showed that a *S. cerevisiae* strain with enhanced AAA biosynthesis and constitutively expresses the genes for naringenin biosynthesis via the phenylalanine route produced 400μM (109mg/L) of naringenin in batch bioreactor (Koopman *et al.*, 2012). A subsequent study look into naringenin production from supplemented coumaric acid, resulting in titres of up to 1.2g/L (Gao, Gao, *et al.*, 2020), but is not directly comparable to *de novo* biosynthesis. Another study looked at *de novo* production of naringenin by systematically separating the engineering strategies into three modules (Lyu *et al.*, 2017). Module 1 represented the naringenin pathway, malonyl-CoA supply was module 2 and the biosynthesis of tyrosine from glucose was module 3. Contrary to speculations that malonyl-CoA may be limiting and that its engineering should improve naringenin production the engineering of modules 1 and 3 but not 2 improved the production of naringenin with a maximum titre of 90mg/L obtained. Their strategy for increasing malonyl-CoA featured the heterologous expression of citrate lyase to increase cytosolic acetate, acetyl-CoA synthase to increase the conversion of cytosolic acetate into acetyl-CoA and acetyl-CoA carboxylase for improved conversion of acetyl-CoA to malonyl-CoA in addition to restraining fatty acid biosynthesis. A recent report, however, showed that the enhancement of fatty acid β-oxidation by expressing key genes of the pathway improved *de novo* naringenin production from tyrosine, resulting in maximum tire of 1.1g/L (Zhang *et al.*, 2021). It should be noted that the titres in both studies (Lyu *et al.*, 2017; Zhang *et al.*, 2021) were obtained in complex media which is not directly comparable to the titre reported for production in minimal medium in the 2012 study by Koopman et al. Thus, naringenin production by engineered *S. cerevisiae* in minimal medium so far could in fact still be low.

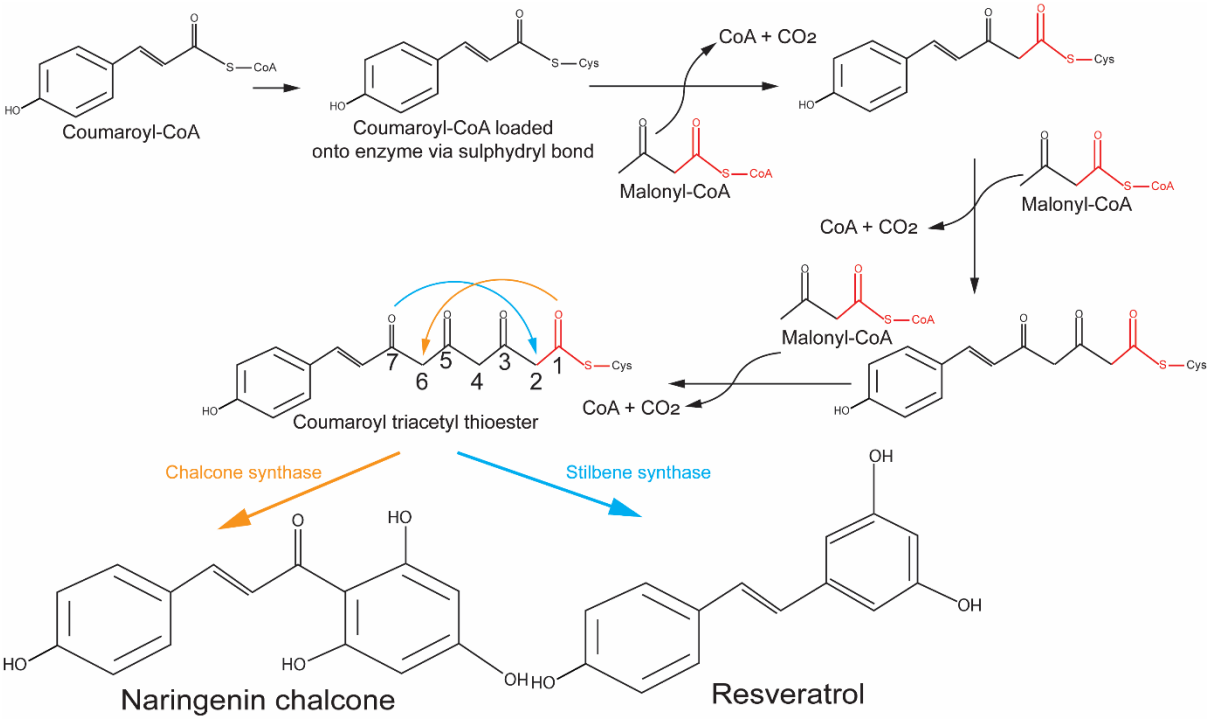


Figure 6. Reaction mechanism for the conversion of coumaroyl-CoA int of naringenin or resveratrol by polyketide synthases. Coumaroyl-CoA is loaded onto the polyketide synthase by means of a sulphydryl bond. Three molecules of malonyl-CoA are then added to form coumaroyl triacetyl thioester. The major difference between the two polyketide synthases, chalcone synthase and stilbene synthase, is in their cyclisation of coumaroyl triacetyl thioester. Chalcone synthase catalyses C1 to C6 while stilbene synthase catalyses C7 to C2 cyclisation of this intermediate into naringenin chalcone and resveratrol, respectively. This figure was adapted from previous review by others (Lim, Go and Yew, 2016).

Despite the use of complex medium and/or feeding with aromatic precursors, resveratrol production in *S. cerevisiae* in most earlier studies was below 10mg/L (Becker *et al.*, 2003; Beekwilder *et al.*, 2006; Shin *et al.*, 2011; Wang *et al.*, 2011). However, an industrial Brazilian sugar cane-fermenting *S. cerevisiae* expressing *A. thaliana 4-CL1* and *Vitis vinifera STS* produced 391mg/L of resveratrol when supplemented with 2.46 g/L of coumaric acid in complex medium (Sydor, Schaffer and Boles, 2010). The engineering of *S. cerevsiae* for resveratrol production has also followed the trend of increasing malonyl-CoA availability in recent times, which has contributed to higher and *de novo* production in minimal media. The overexpression of mutant version of yeast native acetyl-CoA carboxylase (ScAcc1 (S659A, S1157A)), which is insensitive to inhibitory phosphorylation by Snf1, in addition to enhancing AAA biosynthesis enabled CEN.PK lab strain of *S. cerevisiae* to produce 236 mg/L of resveratrol from glucose in flasks (Li *et al.*, 2015) and nearly 800mg/L in fed-batch bioreactors (Li *et al.*, 2016), representing a considerable improvements in production compared to earlier studies especially because the production was *de novo*. The expression of plant malonyl-CoA synthetase gene (*AAE13*), which increased malonyl-CoA concentration, partially complemented *ACC1* deletion in *S. cerevisiae* and increased resveratrol production by 2.45-

fold when fed with coumaric acid, was linked to drastic reductions in other major CoA compounds (Wang, Chen and Yu, 2014). Based on this, the authors proposed that the expression of malonyl-CoA biosynthetic enzymes should be combined with increased CoA biosynthesis. In line with this, the overexpression of *ScCAB1* encoding the first genes of CoA biosynthesis, pantothenate kinase, resulted in 2.8-fold increase in acetyl-CoA production and 3.5-fold increase in naringenin production titre (Liu, Zhang and Jiang, 2017).

The derivatisation of polyketides into downstream molecules has also been achieved in yeasts. This is exemplified by the 2017 work by Rodriguez and colleagues in which they set out to engineer *S. cerevisiae* for *de novo* production a portfolio of related flavonoids (Rodriguez *et al.*, 2017). The expression of *P. hydrida CHS* alone or in addition to a chalcone reductase (CHR) from *Astragalus mongholicus* (Mongolian milkvetch) in a strain with enhanced AAA and expressing *FjTAL* resulted in 1.55mg/L of naringenin and 5.3mg/L liquiritigenin. Chalcone isomerase (Chi) from *Medicago sativa* (Alfalfa) was also expressed in these strains for the isomerization of naringenin chalcone and isoliquiritigenin. The additional expression of *A. mongholicus* flavanone 3-hydroxylase (F3h) and *A. thaliana* flavonol synthase (Fls) produced 26.6mg/L of kaempferol and 0.5mg/L of resokaempferol in the naringenin and (iso)liquiritigenin producers, respectively. The production of kaempferol from naringenin by the F3h and Fls was particularly efficient given that naringenin production was only 1.55mg/L in the parent strain, compared to 26.6mg/L of kaempferol when the genes were expressed. Additionally, although the expression of these genes in the (iso)liquiritigenin producer produced very low titre of resokaempferol, an additional 11.4mg/L of kaempferol was produced from the naringenin that was produced because of incomplete reduction by Chr during isoliquiritigenin production. Finally, the expression of fusion proteins of *Catharanthus roseus* (Madagascar periwinkle) cytochrome P450 reductase (*CPR*) and cytochrome P450 flavonoid monooxygenases (*FMOs*) from *Fragaria ananassa* (strawberry) or *P. hybrida* in the kaempferol-producing background produced 16mg/L and 20.4mg/L of quercetin, respectively while their expressions in the resokaempferol-producing background strain produced 1.2mg/L and 1.65mg/L of fisetin, respectively. In line with the production of naringenin and kaempferol in the (iso)liquiritigenin producer, quercetin production was also detected in the resokaempferol-producing background strain.

Anthocyanins are highly interesting bioactive aromatics with multiple health benefits and have also been demonstrated to be derivable from naringenin in *S. cerevsiae* (Levisson *et al.*, 2018). By expressing F3h, anthocyanidin synthase (Ans) and anthocyanin 3-O-glucosyltransferase (Agt) from *A. thaliana* and dihydroflavonol 4-reductase (Dfr) from *G. hybrida* in a naringenin producer, the production of dihydrokaempferol, kaempferol, kaempferol 3-O-glucoside, pelargonidin and pelargonidin 3-O-glucoside was detected. The

combined and individual titres of kaempferol and kaempferol 3-O-glucoside (KKG branch) were higher than that of pelargonidin and pelargonidin 3-O-glucoside (PPG branch) in strains expressing the four enzymes (Levisson *et al.*, 2018). This showed that the direction of dihydrokaempferol by Ans toward the KKG branch was more efficient than direction by Dfr towards the PPG branch. The attempt to express alternative Dfr enzymes from *M. truncatula* and *Anthurium andreanum* (painter's-palette) did not improve pelargonidin and pelargonidin 3-O-glucoside production and the authors' speculation that *S. cerevisiae's* native glucosidase activity could hydrolyse pelargonidin 3-O-glucoside back into pelargonidin based on report that several native glucosidases can hydrolyse flavonoid glucosides back to their aglycone form (Schmidt *et al.*, 2011) was not corroborated by the deletion of *EXG1* and *SPR1* encoding β-glucosidases.

The substrate promiscuity of hydroxylases results in yet unpredictable biosynthesis of several hydroxylation derivatives of flavonoids so that it is difficult to control the production of specific flavonoids of interest in microbial hosts. This is a major challenge and the ability to control this reaction will highlight the advantage of synthetic biology of microbial hosts as a solution to the difficulty involved in downstream processing for obtaining pure flavonoids from plants. Hence, recent efforts are being focused on hydroxylases. Hydroxylases are cytochrome P450 enzymes and thus require cytochrome P450 reductase (Cpr) to act as an electron donor. The importance of Cpr is demonstrated by the fact that their enhancement is effective for improving the catalysis of cytochrome P450 enzymes (Urlacher and Girhard, 2019). Two independent studies were dedicated to the exploration of the potential hydroxylases for producing naringenin derivatives (Li *et al.*, 2020; Gao *et al.*, 2020). The comparison of four F3h showed that the enzyme from *Citrus sinensis* (sweet orange) resulted in the highest dihydrokaempferol production in naringenin-fed strains (Li *et al.*, 2020). Further expression of F3'h from *Silybum marianum* (milk thistle) converted dihydrokaempferol into taxifolin (dihydroquercetin) and the testing of F3'5'h from five plants with or without *S. marianum* Cpr revealed that *S. lycopersicum* F3'5'h produced the highest dihydromyricetin titre from taxifolin. The expression and choice of *CPR* was also highlighted to be important for dihydromyricetin production with *S. marianum* and *Glycine max* (soybean) *CPR* genes of the six genes tested revealed as the only ones able to increase dihydromyricetin titre (Li *et al.*, 2020), and the expression of F3'h from *S. marianum* or *G. hybrida* without heterologous Cpr did not produce eriodictyol until *CPR* from *S. marianum* or *C. roseus* were additionally expressed (Gao *et al.*, 2020). Finally, *de novo* production of these flavonoids was achieved by enhancing AAA biosynthesis and multi-copy expression of the biosynthetic genes for naringenin production via the tyrosine route (*FjTAL*) in a strain (BM-22) with multi-copy integration of *SlF3'5'H, GmCPR* in addition to single copy expression of *SmF3'H* and *CsF3H,*

and *TAT1* and *TPO1* disruption to generate strain BM-31. This strain produced maximum titres of 0.1mg/L, 61.1mg/L, 114.1mg/L and 246.4 for dihydrokaempferol, eriodictyol, dihydroquercetin and dihydromyricetin, respectively, in fed-batch bioreactor with complex medium. A recent study by these authors further showed that (i) the expression of anthocyanin transporter from *Malus domestica* (apple), *MdGSTF6,* increased the production of cyanidin 3-O-glucoside and delphinidin 3-O-glucoside by over 12-fold (ii) the screening of several Agt and Ans revealed that the combined expression of *Dianthus caryophyllus* (carnation) *AGT* and *P. hydrida ANS* produced the highest titres of both flavonoid glucosides. (iii) *EXG1* deletion increased the production of anthocyanin (Xu *et al.*, 2022). The establishment of these engineering targets in BM-31 (Li *et al.*, 2020) enabled this strain to produce 15.1mg/L, 8mg/L, 3mg/L, 3.5g/L and 0.7mg/L of cyanidin 3-O-glucoside, cyanidin, delphinidin 3-O-glucoside and delphinidin in fed-batch bioreactor with complex medium.

6.3. Production of phenylpropanoids and polyketides in non-saccharomyces yeasts

The production of polyketide phenylpropanoids has been particularly higher in *Y. lipolytica* than in *S. cerevisiae.* This is thought to be linked to the oleaginous nature of this yeast and the requirement of fatty acid biosynthesis for malonyl-CoA, suggesting that the yeast has high production of CoA derivative units. The expression of *G2PS1* encoding a functional 2-pyrone synthase in addition to the overexpression of the yeast's pyruvate bypass (*ACS1, ALD5, PDC2*) and *ACC1* resulted in TAAL titre of 36g/L (Markham *et al.*, 2018), representing the highest production reported yet. The authors also flagged the overexpression of *PEX10,* encoding a peroxisome biogenesis factor, as contributing positively to TAAL biosynthesis in that study and subsequently capitalised on this by overexpressing *PEX10* and *ACC1* in addition to *G2PS1* (strain YT-*PEX10, ACC1*). YT-*PEX10, ACC1* produced TAAL at 8.6g/L and served as basis for successful engineering and production of 28mg/L of naringenin, 48.7mg/L of resveratrol and 0.17mg/L of bisdemethoxycurcumin by coumaric acid feeding in flasks. Finally, they reported *de novo* production of naringenin at 124.1mg/L, achieved by expressing a *Saccharothrix espanaensis* TAL (*SeSAM8*) and polyketide synthase gene (*PKS1*) of *Huperiza serrata* (toothed clubmoss), and by overexpressing the native $ARO4^{FBR}$ and *PEX10*. These strategies translated into the production of 898mg/L of naringenin in glucose pulse-feeding mode, representing the highest microbial naringenin production so far (Palmer *et al.*, 2020).

Similarly, the highest *de novo* production of resveratrol in yeasts have been in *Y. lipolytica.* By expressing *FjTAL, At-4CL1* and *VvSTS* in the yeast without enhancing AAA biosynthesis, resveratrol was produced at 40mg/L (Sáez-Sáez *et al.*, 2020). This was doubled when FBR

Aro4 and Aro7 were overexpressed and further increased to 409mg/L by additional integration of 5 copies of *FjTAL, At-4CL1* and *VvSTS* in flasks and 12.4g/L in fed-batch bioreactor with minimal glucose media. However, the overexpression of a mutant *ACC1* (S667A, S1178A) did not increase production. This titre was recently doubled by the combination of strategies including multi-copy integration of a *Petroselinum crispum 4-CL1* (*Pc-4CL1*)-*VvSTS* fusion protein, the enhancement of coumaric acid precursor biosynthesis by combined expression of the tyrosine and phenylalanine routes, the deletion of the native diacylglycerol acyltransferase gene (*DGA1*) to reduced triacylglycerol biosynthesis and improve the availability of malonyl-CoA (M. Liu *et al.*, 2022). Speculation that inadequate coumaric acid biosynthesis may limit resveratrol production motivated the authors to express bacterial genes that can increase the supply of shikimate pathway erythrose 4-phosphate precursor and acetyl-CoA (the phosphoketolase pathway (figure 2)), including *Bifidobacterium brebifidum* (*BbFPK*), *B. adolescentis* (*BaFPK*) and *Clostridium acetobutylicum* (*CaFPK*), and phosphotransferase from *Bacillus subtilis* (*BsPTA*) and *E. coli* (*EcPTA*). Indeed, *CaFPK* which catabolises fructose 6-phosphate into acetyl phosphate and erythrose 4-phosphate, and *BsPTA,* which further converts acetyl phosphate into acetyl-CoA, increased coumaric acid production, but not resveratrol.

De novo-produced naringenin has also been derivatised in *Y. lipolytica* into hydroxylation derivatives like eriodictyol and taxifolin, by testing combinations of cytochrome P450 flavonoid 3′-hydroxylase (F3′h) and Cpr from several sources, in addition to enhanced AAA biosynthesis and increased malonyl-CoA biosynthesis (Lv *et al.*, 2019). The authors' report of maximum titre of naringenin, eriodictyol and taxifolin of 252.4mg/L, 134.2mg/L and 110.5mg/L, respectively, although in complex medium, represents an encouraging result that naringenin can be derivatised into downstream metabolites in this yeast. Given that naringenin production is higher in *Y. lipolytica* than in *S. cerevsiae* it would be interesting to see how targets identified in *S. cerevisiae* for anthocyanin production as discussed above translate to *Y. lipolytica*.

Although the production of phenylpropanoids in *K. marxianus* is yet to be reported, the simple polyketide, TAAL, has been produced in this yeast (McTaggart *et al.*, 2019). This work was dedicated to the exploration of the yeasts' performance in terms of growth and TAAL production from diverse carbon sources and at high temperatures when 2-PS from *G. hybrida* is expressed. A maximum TAAL production titre of 1.24g/L from 10g/L of xylose was achieved with a strain only expressing 2-PS from a high copy plasmid under the control of the *S. cerevisiae ADH2* promoter. Given the recent developments of SynBio tools for engineering *K. marxianus*, it is speculated that it will be possible to engineer this yeast to produce

phenylpropanoids, potentially allowing us to avail of the physiological advantages of the yeast for this purpose.

7. The yeast *K. marxianus* and its potential for biotechnological applications

K. marxianus is a homothallic hemiascomycetous yeast and is a phylogenetically close relative of *S. cerevisiae* (Lane and Morrissey, 2010). It is regarded as a sister species of *K. lactis* which is more commonly studied, and knowledge of its biology is often inferred for *K. marxianus.* Yeasts of the *K. marxianus* species, like *K. lactis,* are broadly characterised by their ability to assimilate lactose as carbon sources and are commonly found in food stuff without eliciting detrimental effects, earning them the United States GRAS (Generally Regarded As Safe) status and the European Union QPS (Qualified Presumption of Safety) status. This has opened *K. marxianus* up for many biotechnology applications due to reduced regulatory restrictions. This advantage alongside the fact that the yeast is able to metabolise more diverse substrates and is more thermotolerant than *S. cerevisiae* has attracted the interest of the scientific community in recent times, not only purely for biotechnological applications but also delving into studies to uncover underlying factors that convey these phenotypes. While the lactose assimilation capability is a defining phenotype for *K. marxianus*, not all members of the species have this ability and this was shown to be due to polymorphisms in the *LAC12* gene encoding the permease responsible for transporting of lactose into the cell (Varela *et al.*, 2017). Although strains that were unable to transport lactose possess *LAC12* orthologues, the gene is rendered non-functional for lactose transport by variation in 13 amino acids in the gene product. Furthermore, increased expression of *LAC12* genes has was reported in hypoxic condition and suggested to result in increased lactose uptake (de Paiva *et al.*, 2019). With *K. marxianus* strains having the ability to grow at higher temperature than many other yeasts, attempt to elucidate genes that are involved conveying thermotolerance by means of classical genetics showed that the deletion of *KLMX_70834* resulted in loss of growth at 47^{o}C (Montini *et al.*, 2022). Bioinformatic analysis revealed that the gene product may be involved in RNA-binding activity, but further investigation is required to reach a definitive conclusion on the role played by this gene in thermotolerance. It would be interesting to determine whether this gene can convey thermotolerance in non-thermotolerance yeasts like *S. cerevisiae*, further extending their biotechnological applications.

The properties of *K. marxianus* have driven its application for commercial biotechnology such as the production of bioethanol from side streams of the dairy industry, biomass or single cell

protein and enzyme production (Lane and Morrissey, 2010). Its use for various environmental applications such as paper waste and sludge treatment and removal of lactose and other sugars from wastes and sludge have been investigated (Hang, Woodams and Hang, 2003; Kádár, Szengyel and Réczey, 2004). The yeast also has potential for use in the food industry for the production of natural emulsifiers (del Carmen Vasallo *et al.*, 2006) and for baking (Caballero *et al.*, 1995). Its use for the production of aroma was proposed and it was suggested to be a suitable host for development as cell factory for the production of natural flavour and fragrance molecules (Morrissey *et al.*, 2015). To that effect, *K. marxianus* has seen pivotal developments in regard to the available SynBio tools and improved understanding of its physiology in recent years as will become apparent as you read on in this thesis.

Chapter 2. Development of synthetic biology tools for genomic integration in *Kluyveromyces marxianus*

Joel A. Akinola, Javier A. Varela, Jane McGrath, Jeremiah Burk, John P. Morrissey.

The data in this chapter are being prepared for inclusion in a manuscript

JAA designed and conducted the experiments, interpreted the data and prepared the manuscript

JAV designed the strategy for constructing the multiplex CRIPSR plasmids by Golden Gate assembly

JG constructed all deletion strains and conducted growth assessment on them in the *amdS* work as an undergraduate thesis under the supervision of JAA

JB constructed the complementation strains and conducted growth assessment on them in the *amdS* work as an undergraduate thesis under the supervision of JAA

JPM designed the *amdS* aspect of this work, provided supervision and edited parts of the manuscript.

Abstract

The application of yeasts for synthetic biology (SynBio) requires the use of genetic engineering for step-by-step modification of hosts. The selection of successfully modified cells is a highly important aspect in host genetic modification at each step. This is achieved through the linkage of desired modifications to selectable genes that confer a survival phenotype in the presence of selective pressure. With “unconventional” yeasts becoming widely used for SynBio applications, there is a need to address the selection of modified yeast cells during strain development. Although auxotrophic and dominant marker genes are routinely used for successful genetic engineering of *Kluyveromyces marxianus*, the number of available marker genes is limited and represents a major challenge in strain engineering. The solutions are: (i) increase the number of selectable genes available, (ii) recycle markers so that they can be used for other rounds of modifications, and (iii) use of marker-free systems. Marker recycling is an attractive modality because it allows a single marker to be reused multiple times and the possible protein burden that can arise from the expression of the marker gene which is usually no longer needed after selection is prevented; but better than this, is the elimination of the need for markers that is achieved in marker-free systems.

In this study, the issue of selectable markers in *K.* marxianus was first addressed by developing a gene called *amdS* from the filamentous fungi *Aspergillus nidulans* which encodes an enzyme that hydrolyses acetamide as a selectable marker in *K. marxianus*. Given that this gene has been demonstrated to be counter-selectable on a cytotoxic analogue of acetamide, fluoroacetamide, in close relatives of *K. marxianus* using direct repeats, attempts were made at developing the gene as a recyclable marker to enable its application for introducing multiple modifications into the yeast. However, because the counter-selection of *amdS* in *K. marxianus* was unsuccessful, we were unable to recycling the marker. Therefore, we moved on to address the issue by creating a marker-free system. Double stranded breakage (DSB) induction technologies such as clustered regularly interspaced short palindromic repeats (CRISPR) promise marker-free genome editing due to the lethality of unrepaired DSBs. With CRISPR-Cas9 systems already developed for *K. marxianus* for single gene disruption, the system was further developed in this work for marker-free genomic integration and simultaneous multi-loci genome editing. To achieve this, single plasmids harbouring Cas9, and up to 5 guide RNA (gRNA) expression systems were constructed. Testing of these plasmids for targeting *URA3, HIS3* and *LEU2* showed that the provision of DNA for repairing DSB by homology directed repair (repair DNA) increased the number of transformants obtained when a single locus is targeted in wildtype *K. marxianus*, but not for multi-loci targeting. An editing efficiency for multi-loci targeting of 38% was however obtained when

repair DNAs were transformed compared to 0% with no repair DNAs. Based on evidence that successful editing was mainly achieved by homology directed repair when repair DNAs are provided for multi-loci editing, various strategies were used to improve the number of transformants and editing efficiency, including the inactivation of non-homologous end joining, heterologous expression of recombinases and chemical synchronisation of cells into S/G2 cell cycle phase. By so doing, methods for achieving marker-free multi-loci integrations at the *URA3, HIS3* and *LEU2* loci and for multi-copy integrations at repeat (delta site) sequences distributed around the yeast's genome were developed. As proof-of-concept, the multi-copy integration system was used to increase the copy number of an *Arabidopsis thaliana* gene encoding chalcone synthase, a bottleneck enzyme in the biosynthesis of flavonoids, in a *K. marxianus* strain that already produces low yield of naringenin. The tools developed here are based on the Yeast Toolkit (YTK) standard, ensuring that they can be used in combination with other YTK parts, and are applicable for rapid development of *K. marxianus* for SynBio applications.

Introduction

The development of yeasts as cell factories for replacing the production of valuable chemicals currently produced through extraction from plants or by chemical synthesis from often unsustainable precursors is an active area of research due to demands for sustainable manufacturing. With *Saccharomyces cerevisiae* as the most studied yeast, the wealth of knowledge available for it has earned it its position as mainstay for yeast SynBio applications (Liu, Zhang and Nielsen, 2019). However, continued research efforts to improve our understanding of the physiology and genetic manipulation of non-saccharomyces yeasts have resulted in recent successes in developing them as cell factories for producing several valuable molecules (Patra *et al.*, 2021). The specific phenotypes that they naturally possess make them attractive for such applications (Löbs, Schwartz and Wheeldon, 2017) and *K. marxianus* is an example of these yeasts. *K. marxianus* is industrially attractive because of its thermotolerance, which can reduce cooling cost during industrial production, and fast growth. The ability of this yeast to metabolise diverse carbon sources means that feedstock choice for production is broader, such as process waste from several industries in support of the movement towards a circular economy. Although the natural phenotypes of a yeast may allow basal production in wildtype strains in some cases, genetic engineering is applied for strain optimisation. Thus, the first step towards availing of these advantages is to re-direct or re-engineer *K. marxianus* metabolism towards producing desired molecules through genetic engineering. This can involve the disruption of native genes and/or the introduction of genetic materials for the functional expression of genes encoding the enzymes involved in the

biosynthetic pathways of the molecules of interest, which can be foreign or native, as episomal vectors (plasmids) or by genomic integration. Due to possible expression instability of plasmid systems (Zhang, Moo-Young and Chisti, 1996), genomic integration is widely preferred for strain development. Thus, the development of cells as factories for molecule production generally requires that the genome of the cell be edited. Each editing step requires transformation, selection, and confirmation, making strain construction laborious and a major bottleneck in the Design-Build-Test-Learn (DBTL) iteration strategy for strain development.

Genome editing is traditionally achieved by homology flanks that foster genomic integration by means of homologous recombination with the aid of a selective pressure. Selectable genes are commonly used to select successfully modified cells. The linkage of a genome editing modification to a selectable gene ensures that unedited cells are eliminated by the selective pressure, allowing only edited cells to be obtained. At each step during development, strains are constructed by linking each modification to a selectable gene as illustrated for native gene disruption in figure 1A, but this can also be used for gene integration for heterologous expressions. A major challenge to strain development is the fact that the number of possible modifications is limited by the number of selectable genes available. This problem is minimised by the availability of yeast selectable genes for *S. cerevisiae* (Siewers, 2014), but only a few are available for non-saccharomyces yeasts (Cai, Gao and Zhou, 2019). For this purpose, counter-selectable marker genes are attractive because yeasts can be cured of them after selection, making them available for reuse and preventing possible protein burden arising from their expression which is usually no longer needed after selection. A commonly used counter-selectable marker is the *URA3* gene which encodes orotidine 5'-phosphate decarboxylase. The disruption of *URA3* prevents yeast growth in uracil-free medium (uracil auxotrophy), allowing the selection of strains with the desired modifications by associating the modification with the marker gene. The gene can then be removed for another round of modification by cultivating the strain on 5-fluoroorotic acid while the previous one is maintained, thereby selecting for cells that have lost the gene and have therefore regained uracil auxotrophy. The gene removal process is achieved through excision by homologous recombination mediated by direct repeat sequences flanking the integrated *URA3* marker (Alani, Cao and Kleckner, 1987; Längle-rouault and Jacobs, 1995; Kaneko *et al.*, 2009). Although other counter-selectable markers that work based on similar principle are available such as *MET15, LYS2,* and *CYH2* (Singh and Sherman, 1974; Chattoo *et al.*, 1979; Käufer *et al.*, 1980), site-specific homologous recombination-mediated counter-selection such as *LoxP-Cre* and Flp/FRT recombination systems have the potential to make any marker gene recyclable by flanking it with specific sequences and expressing *Cre* or Flippase (*Flp*) recombinases (Gueldener *et al.*, 2002; David and Siewers, 2015; Fraczek, Naseeb and Delneri, 2018). The site-specific activity

of *Cre* and *Flp* recombinases at *LoxP* and *FRT* sites, respectively, enables the excision of the marker gene without compromising the modification. However, a major limitation of this system is that repeated usage causes chromosomal rearrangements (Burgess and Kleckner, 1999; Delneri *et al.*, 2000; Solis-Escalante *et al.*, 2015).

Some selectable markers are available for *K. marxianus* and have been successfully applied for SynBio and for furthering our understanding of the yeast's physiology (Kim, Lee and Oh, 2014; Choo *et al.*, 2018; Rajkumar *et al.*, 2019). Marker recycling by means of the *LoxP-Cre* recombinase system was even applied in this yeast (Ribeiro *et al.*, 2007), but is plagued by the aforementioned limitation. The *amdS* gene from *A. nidulans* encodes for an amidase, which hydrolyses acetamide into acetate and ammonium (acetamidase). This gene has been used as a selection marker in *S. cerevisiae,* non-saccharomyces yeasts and in filamentous fungi (Kelly and Hynes, 1985; Yamashiro, Yarden and Yanofsky, 1992; Royer *et al.*, 1999; Read *et al.*, 2007; Schuster *et al.*, 2012; Solis-Escalante *et al.*, 2013; Fu *et al.*, 2016; Piva *et al.*, 2018; Hamilton *et al.*, 2020). This was made possible by the inability of these hosts to naturally hydrolyse acetamide. By expressing *amdS* in them they gain the ability to assimilate acetamide as nitrogen and carbon sources, enabling the application of the gene as a dominant marker. Once modification is successful, the *amdS* gene can be recycled by counter-selection on fluoroacetamide, a fluorine-containing analogue molecule of acetamide which is hydrolysed by the acetamidase to release the cytotoxic molecule, fluoroacetate. The application of *amdS* as marker is however not as straightforward in hosts that naturally assimilate and grow on acetamide because the growth phenotype must beforehand be eradicated, thereby permitting the use of an acetamidase gene as selectable marker as was recently demonstrated for *Yarrowia lipolytica* (Hamilton *et al.*, 2020). This yeast naturally grows on acetamide but the putative acetamidase genes were identified and disrupted, rendering it able to use its own acetamidase gene as marker (Hamilton *et al.*, 2020). *K. marxianus* also grows on acetamide unlike its evolutionary relative yeast, *K. lactis*, which grows poorly. While this phenotype gives wildtype *K. lactis* the advantage of using *amdS* as a dominant gene marker (Ooyen *et al.*, 2006; Read *et al.*, 2007), native acetamidase genes must be disrupted in *K. marxianus* to render the yeast unable to metabolise acetamide in order to apply acetamidase genes as marker in the yeast.

Technologies for introducing double stranded breakage (DSB) into genomes promise to eliminate the need for selectable markers altogether due to the lethality of unrepaired DSBs to cells. These breakages were previously achieved by transcription activator-like effector nucleases (TALENs) and zinc finger nucleases (ZFN), but the advent of clustered regularly interspaced short palindromic repeats (CRISPR) resulted in its widespread application for this purpose because it is easier to establish and suitable for multiplex genome editing (David and

Siewers, 2015; Cai, Gao and Zhou, 2019). CRISPR works by RNA-directed guiding of endonucleases to generate locus-specific DSBs (Jinek *et al.*, 2013), and although Cas9 represents a commonly used endonuclease, others have been developed for CRISPR applications such as Cas12a and MAD7. Genomic modifications are introduced during DSB repairs through endogenous cell repair mechanisms such as error-prone non-homologous end joining (NHEJ), or homology directed repair (HDR). HDR is routinely harnessed for genome engineering of *S. cerevsiae* by targeting specific loci for DSB generation and supplying repair DNA with homology flanks (Dicarlo *et al.*, 2013; Mitsui, Yamada and Ogino, 2019). The application of this technology for native gene disruption is illustrated in figure 1B but can also be used for gene integration. Although evidence that NHEJ is the predominant repair mechanism in many non-saccharomyces yeasts led to its application for gene disruptions (Cai, Gao and Zhou, 2019), this repair is not available for genomic integrations. Loss of function mutations in NHEJ machinery apparently unmasked HDR for efficient DSB repair in *K. marxianus* and other non-saccharomyces yeasts (Cai, Gao and Zhou, 2019). NHEJ-deficient *K. marxianus* strains are HDR-effective and have been applied for efficient marker and CRISPR-assisted marker-free genomic integrations (Nambu-Nishida *et al.*, 2017; Rajkumar *et al.*, 2019; Bever, Wheeldon and Da Silva, 2022). Importantly, Rajkumar and Morrissey recently provided a detailed and useful step-by-step protocol for achieving CRISPR-mediated marker-free single gene disruption of *K. marxianus* by HDR (Rajkumar and Morrissey, 2022).

Simultaneous genome editing has the potential to speed up strain construction (the build stage) because it allows the introduction of multiple genomic modifications in a single transformation step. The simplicity of CRISPR in that endonucleases are targeted to specific genomic loci by gRNA molecules is an advantage that allows for multiplex targeting. Once the endonuclease is being expressed in the cell, multiple gRNA molecules targeting more than one locus can be provided to direct endonuclease molecules to the loci for DSB generation. Cas9-based version of this technology, referred to as CRISPR-Cas9-assisted multiplex genome editing (CMGE) elsewhere (Wang *et al.*, 2018), has enabled simultaneous polygenic disruption and chromosomal integrations at multiple loci, and variations of it have been developed for *S. cerevisiae* (Jakočiūnas *et al.*, 2015; Mans *et al.*, 2015; Zhang *et al.*, 2019) and is being extended to non-saccharomyces yeasts (Malcı, Walls and Rios-Solis, 2020). Although the burden posed on cells by a single DSB is repairable by hosts' native repair machinery, enabling single loci targeting in many non-saccharomyces yeasts, multiple DSBs pose greater burden. Thus, multi-loci targeting is more challenging in these yeasts because multiple DSBs must be generated simultaneously, and yeast cells must be able to repair them in order to survive. Multi-loci targeting has been successful in non-saccharomyces species thanks to the implementation of various modalities such as the use of Cas12a which is more

sophisticated than Cas9 in its ability to process gRNA, ensuring a CRISPR system that does not rely on host RNA processing machinery (Swarts and Jinek, 2018; Paul and Montoya, 2020). Nonetheless, the repair of multiple DSBs remains challenging and limits the number of loci that can be targeted in non-saccharomyces hosts. Some wildtype strains of these yeasts are able to repair DSBs by both HDR and NHEJ, making it possible to simultaneously edit the genome by the introduction of indels at some target loci and by repair DNA integration at other target loci. Although the outcome of that approach is uncontrollable, it could suffice for multiple gene disruptions. Marker-free multi-loci integration of genes, however, requires greater control that may be achieved by inactivating repair by NHEJ so that all generated DSBs are only repairable by HDR, ensuring that repair DNAs for all targeted loci are integrated. Multi-loci targeting has been achieved in both wildtype and NHEJ inactivated strains with efficiencies of 85.6% in wildtype and 94.1% in *ku70/ku80 Y. lipolytica* for disrupting two genes (Gao *et al.*, 2016), 2.1% for triple integrations in *ku80 K. lactis* (Horwitz *et al.*, 2015) and 25% for double integrations in *ku70 K. marxianus* (Bever, Wheeldon and Da Silva, 2022).

In this study, we addressed the issue of markers in regard to the development of tools for genomic integrations in *K. marxianus.* Firstly, we constructed a *K. marxianus* strain that can be used as background for achieving marker-mediated genome editing by *amdS* selection on acetamide. Our attempt to counter-select the gene on fluoroacetamide was, however, unsuccessful due to low sensitivity of *amdS*-expressing *K. marxianus* strains to the toxicity of fluoroacetamide. As an alternative to marker-dependent integration, we developed a CRISPR-mediated marker-free integration system that allows multi-loci integration at the *URA3, HIS3* and *LEU2* loci and at repeat sequences distributed around the yeast's genome.

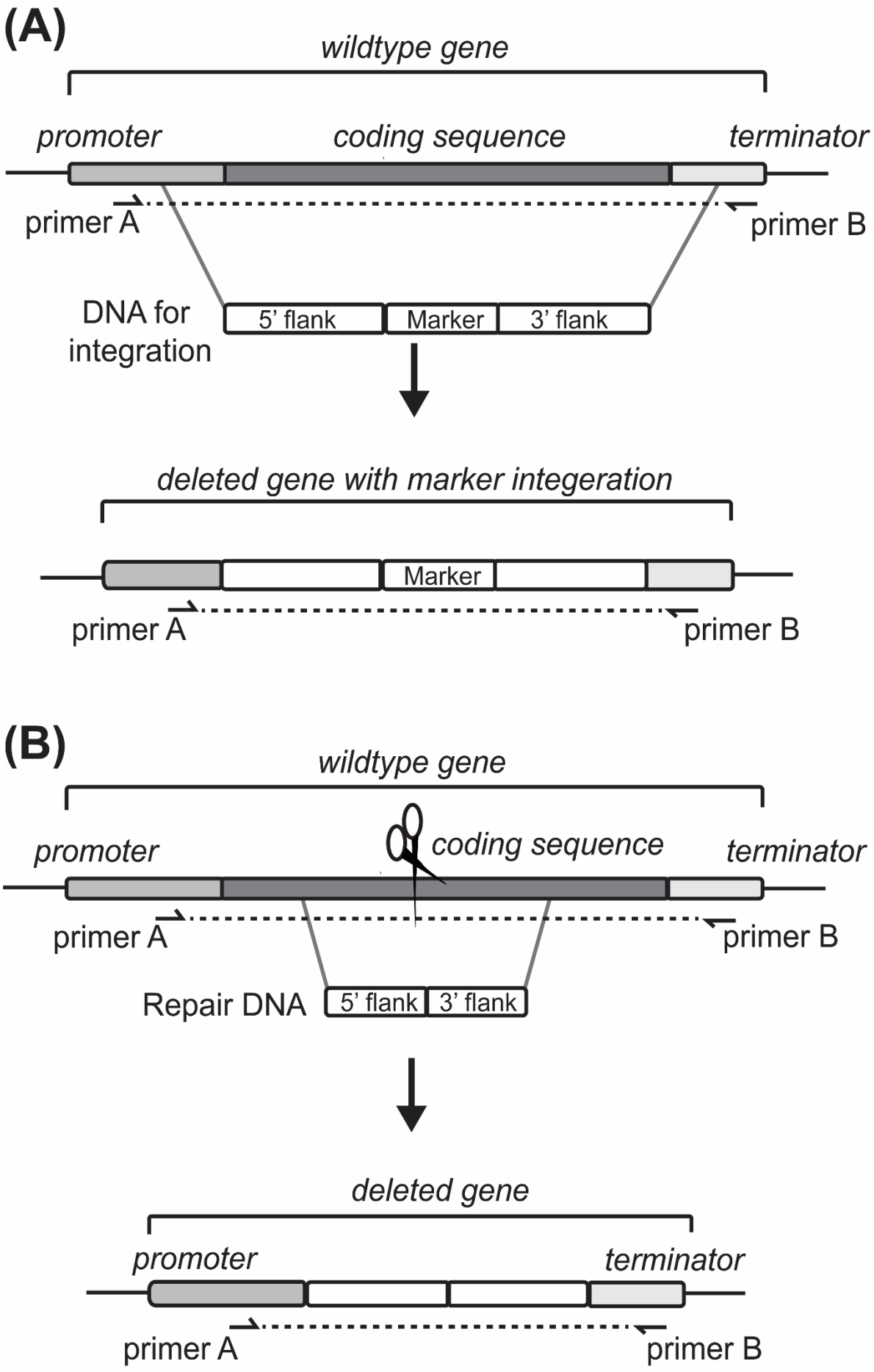

(A)
wildtype gene
promoter
coding sequence
terminator
primer A
primer B
DNA for integration
5' flank
Marker
3' flank
deleted gene with marker integeration
Marker
primer A
primer B
(B)
wildtype gene
promoter
coding sequence
terminator
primer A
primer B
Repair DNA
5' flank
3' flank
deleted gene
promoter
terminator
primer A
primer B

Figure 1. Strategies for genome editing. (A) Genome editing by flanking the DNA for integration with 5' and 3' sequences that are homologous to the target locus (represented as a gene here). The integration is achieved by homologous recombination. The DNA harbours a selectable marker that enables the selection of cells in which the DNA has been integrated. (B) Genome editing by technologies for introducing double stranded breakage (DSB) into genomes such as CRISPR and DSB repair by homology directed repair (HDR). CRISPR is used to target a locus (represented as a gene here) for DSB generation. By providing a DNA fragment with 5' and 3' flank sequences that are homologous to sequences outside the DSB site (repair DNAs), the DSB can be repaired by HDR, resulting in the integration of the repair DNA and the deletion of a segment of the locus. Both strategies delete the segment between the flanks, causing the gene to be disrupted. Primers A and B in both strategies represent forward and reverse locus-specific diagnostic primers to detecting genome editing.

Materials and methods

1. Strains and growth conditions

All yeast strains used in this study are listed in Table 1. Yeasts were routinely cultured at 30°C in YPD broth (2% yeast extract, 1% peptone and 2% glucose), minimal glucose medium (5g/L $(NH_4)_2SO_4$, 3g/L KH_2PO_4, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 15g/L glucose, and filter-sterilized vitamins and trace elements) brought to pH 5.5-6 with KOH (Verduyn *et al.*, 1992) or synthetic drop-out (SD) medium with the relevant supplements (0.15% ammonium sulphate, 0.19% yeast nitrogen base without amino acids or ammonia (Formedium), 2% glucose and the relevant drop-out supplement according to the manufacturer's recommendations (Formedium)). All liquid cultivations were done in 50mL media in 250mL Erlenmeyer flasks. Acetamide plates were made by supplementing minimal glucose medium with 0.6g/L acetamide (Sigma-Aldrich, Product No. 00160) as sole nitrogen source, with the absence of $(NH_4)_2SO_4$ compensated for by adding equimolar concentration of K_2SO_4. The growth phenotypes of yeast strain were determined by cultivation in liquid media or by spotting on solid media. For assessing growth phenotypes on acetamide plates, strains were precultured in minimal glucose media for 42 hours followed by further 24 hour cultivation in medium without nitrogen source to ensure that ammonium from the first cultivation was completely consumed so that acetamide was truly the only nitrogen source on the plates. *S. cerevsiae* BY4741 strains media were supplemented with 76mg/L methionine, 20mg/L uracil, and 20mg/L histidine. For assessing the sensitivity of strains to fluoroacetamide, minimal glucose media were supplemented with appropriate concentrations of fluoroacetamide (Santa Cruz Biotechnology, Catalogue No. sc-250002). Cultures were washed once with water, serial diluted to A 600nm of 1, 0.1, 0.01 and 0.001 and 3µL of each dilution was spotted on solid media. Solid media were made by 2% agar supplementation. Nutritional reagents and agar were purchased from Formedium, Hunstanton, UK. Bacterial transformation was with *E. coli* DH5α grown in LB broth (1% NaCl, 1% peptone, 0.5% yeast extract) or LB supplemented with 1.5% agar and appropriate antibiotic (50mg/L chloramphenicol, 100mg/L ampicillin, or 50mg/L kanamycin; Fisher Scientific (Dublin, Ireland)).

Table 1. Strains used in this study.

Strain	Genotype	Source
K. marxianus	NBRC177	NITE Biological Resource Centre
KmASR.005	*K. marxianus* NBRC1777 *dln4*	(Rajkumar *et al.*, 2019)
K. lactis	*K. lactis* CBS2359 wildtype	Westerdijk Institute
BY4741	*S. cerevisiae* S288C *his3leu2met15ura3*	(Brachmann *et al.*, 1998)
ScASR.001	BY4741 [pGREG505]	This work
ScASR.002	BY4741 [pGREG *amdS*]	This work
KmJA.266	KmASR.005 *kmar_20785*	This work
KmJA.267	KmASR.005 *kmar_40009*	This work
KmJA.268	KmASR.005 *kmar_50306*	This work
KmJA.269	KmASR.005 *kmar_80308*	This work
KmJA.270	KmASR.005 *kmar_80308kmar_20785*	This work
KmJA.271	KmASR.005 *kmar_80308kmar_40009*	This work
KmJA.272	KmASR.005 *kmar_80308kmar_50306*	This work
KmJA.295	KmJA.272; I4:: *KmPDC1pr-KMAR_50306-KmINU1ter* (*HphMX)*	This work
KmJA.296	KmJA.272; I4:: *KmPDC1pr-KMAR_80308-KmINU1ter* (*HphMX*)	This work
KmJA.281	KmJA.272; I4::*KmPDC1pr-amdS-KmINU1ter* (*HphMX*)	This work
KmJA.318	KmJA.272; I3::*amdS*	This work
KmJA.355	KmJA.272; I3:: *KmPGK1pr-Venus-KmPDC1ter* (*amdS* recyclable)	This work
KmJA.356	KmJA.272; I4:: *KmPGK1pr-Venus-KmPDC1ter* (*amdS* recyclable)	This work
KmJA.112	*K. marxianus* NBRC1777 *yku80*	This work
KmJA.098	KmJA.112; with *EcrecA* integrated at locus downstream of *KmHSP104*	This work
KmJA.099	KmJA.112; with *ScRAD52* integrated at locus downstream of *KmHSP104*	This work
KmJA.100	KmJA.112; with *ScGAL1pr-KmPGK1ter* integrated at locus downstream of *KmHSP104*	This work
KmJA.282	*K. marxianus* strain engineered for *de novo* production of flavonoids	This work

2. Construction of plasmids

2.1. CRISPR plasmids

All plasmids used in this study were constructed by Golden Gate assembly based on the Yeast Toolkit (YTK) standard (Lee *et al.*, 2015) and are listed in Table 2. Protospacer sequences for targeting genes were inserted into CRISPR plasmid pUCC001 according to our recent protocol (Rajkumar and Morrissey, 2022). Briefly, complementary oligonucleotides of the gRNA protospacer sequences predicted by the SgRNAcas9 software (Xie *et al.*, 2014) were designed and synthesised with 5′ overhang (5′-CGTC-3′) and 3′ overhang (5′-AAAC-3′) on the sense and antisense strands, respectively. The complementary oligonucleotides were denatured, annealed, and phosphorylated with 1 µL of T4 polynucleotide kinase (10 U/µL, NEB) in a total reaction volume of 10 µL. Fifty femto mole of the annealed protospacer insert was then put together with 100 ng of pUCC001 in a BsaI/T7 ligase Golden Gate reaction. Bacterial transformants of reaction products were selected on ampicillin supplemented LB medium and colonies harbouring plasmids with correctly inserted gRNA protospacer DNA were confirmed by Polymerase Chain reaction (PCR) using primer Bsa-R and the sense oligonucleotide of each gRNA protospacer sequence. In the development of marker-free tools, high copy number CRISPR plasmids used for single or simultaneous multi-loci targeting were constructed as illustrated in figure 8. pUCC001 plasmids harbouring *K. marxianus URA3, HIS3* and *LEU2* or no target were linearise by EcoRI restriction and used as PCR template from which the *ScTDH3pr-gRNA-ScCYC1ter* (gRNA transcriptional unit) sequence was amplified by means of Q5 High-Fidelity Polymerase (M4092L, New England Biolabs (NEB)) using primers JA603 to JA612. The PCR amplification products contained BsmBI binding sites to enable the generation of 5' and 3' overhangs for Golden Gate assembly. Thus, the primer pair used for amplifying each gRNA transcriptional unit sequence is determined by the desired order of the concatenated gRNA transcriptional units in the final CRISPR plasmid. All CRISPR targets and primers used in this study are listed on Table 3 and Table 4, respectively. Once purified, amplification products were put together in a BsmBI/T7 ligase Golden Gate reaction with pP7Cas9T1 and pMTU-DO-HPH-C4 to generate final CRISPR plasmids which contained 1 to 5 gRNA transcriptional units with targets or without targets (empty). In the final plasmids, pP7Cas9T1 provided *S. cerevsiae* codon optimised Cas9 gene from *Streptococcus pyrogenes* while pMTU-DO-HPH-C4 supplied the plasmid backbone containing a bacterial kanamycin resistance *KanR* gene, a high copy number replication origin for *K. marxianus* and yeast hygromycin resistance *hphMX* gene. Each gRNA in the final plasmid has its own *ScTDH3* promoter and *ScCYC1* terminator.

Table 2. Plasmids used in this study.

Name	Relevant description	Source
pUCC001	*AaTEF1pr-SpCas9-ScPHO5ter ScTDH3pr-HH-gRNA$_{empty}$-HDV-ScCYC1t-AmpR- hphMX*	Addgene ID. 124451
pUCC001-*KMAR_20785*	pUCC001 with *K. marxianus KMAR_20785* target insert	This work
pUCC001-*KMAR_40009*	pUCC001 with *K. marxianus KMAR_40009* target insert	This work
pUCC001-*KMAR_50306*	pUCC001 with *K. marxianus KMAR_50306* target insert	This work
pUCC001-*KMAR_80308*	pUCC001 with *K. marxianus KMAR_80308* target insert	This work
pGREG-505	*ScHIS3-ScCYC1ter-kanMX-ScLEU2-AmpR*	(Jansen *et al.*, 2005)
pGREG-505-TEF1	*ScTEF1pr-ScHIS3-ScCYC1ter-kanMX-ScLEU2-AmpR*	(Varela *et al.*, 2017)
pKmK.I3L	YTK level-I plasmid bearing 5' homology arm for targeting integration at a locus between *KmSWF1* and *KmARO1*	Addgene ID. 125032
pKmK.I3R	YTK level-I plasmid bearing 3' homology arm for targeting integration at a locus between *KmSWF1* and *KmARO1*	Addgene ID. 125065
pKmK.I4R	YTK level-I plasmid bearing 5' homology arm for targeting integration at chromosome IV:240042..241741 (locus downstream of *KmHSP104*)	Addgene ID. 125033
pKmK.I4R	YTK level-I plasmid bearing 3' homology arm for targeting integration at chromosome IV:240042..241741 (locus downstream of *KmHSP104*)	Addgene ID. 125066
pYTK084	YTK level-I plasmid bearing *KanR*-ColE1	
pYTK047	YTK level-I plasmid bearing bearing GFP dropout	
pYTK073	YTK level-I plasmid bearing ConRE' part	
pYTK001	Level I GFP dropout plasmid	Addgene ID. 65108
pL1 *KMAR_50306*	Level I plasmid of *KMAR_50306*	This work
pL1 *KMAR_80308*	Level I plasmid of *KMAR_80308*	This work
pUG-*amdSYM*	Template plasmid from which *amdS* was amplified for *in vivo* assembly with pGREG-505-TEF1 in BY4741	(Solis-Escalante *et al.*, 2013)
pL1_*amdS*	Level I plasmid of *A. nidulans amdS* codon optimised for *S. cerevisiae*	This work
pI4-MTU-DO HPH	Integrative plasmid with GFP dropout targeting locus downstream of *KmHSP104*; *hphMX* marker	Addgene ID. 160216
pL1_P2*amdS*T1 marker	Level I plasmid of *KmPDC1pr- amdS-KmINU1ter* as marker part	This work
pL1_P2*amdS*T1 recyclable marker	Level I plasmid of recyclable *KmPDC1pr-amdS-KmINU1ter* as marker part. The *amdS* transcriptional unit has the direct repeat sequence at the 3' end of *KmINU1ter.*	This work
pI4-P2-*KMAR_50306*-T1 HPH	pI4-MTU-DO HPH with insert *KmPDC1pr-KMAR_50306-KmINU1ter*	This work
pI4-P2-*KMAR_80308*-T1 HPH	pI4-MTU-DO HPH with insert *KmPDC1pr-KMAR_80308-KmINU1ter*	This work
pI4-P2-*amdS*-T1 HPH	pI4-MTU-DO HPH with insert *KmPDC1pr-amdS-KmINU1ter*	This work
pI3-MTU-DO-AMDS	Integrative plasmid with GFP dropout targeting locus between *KmSWF1* and *KmARO1*; *amdS* marker	This work
pI4-MTU-DO-AMDS	Integrative plasmid with GFP dropout targeting locus downstream of *KmHSP104*; *amdS* marker	This work
pI3-P1YT4-AMDS	pI3-MTU-DO AMDS with insert *KmPGK1pr- Venus-KmPDC1ter*	This work
pI4-P1YT4-AMDS	pI4-MTU-DO AMDS with insert *KmPGK1pr- Venus-KmPDC1ter*	This work

pI3-MTU-DO recyclable	AMDS	Integrative plasmid with GFP dropout targeting locus between *KmSWF1* and *KmARO1*; *amdS* recyclable marker	This work
pI4-MTU-DO recyclable	AMDS	Integrative plasmid with GFP dropout targeting locus downstream of *KmHSP104*; *amdS* recyclable marker	This work
pI3-P1YT4 recyclable	AMDS	pI3-MTU-DO AMDS recyclable with insert *KmPGK1pr-Venus-KmPDC1ter*	This work
pI4-P1YT4 recyclable	AMDS	pI4-MTU-DO AMDS recyclable with insert *KmPGK1pr-Venus-KmPDC1ter*	This work
pKmK.T4		Level I plasmid of *KmPDC1* terminator	Addgene ID. 125056
pKmK.P1		Level I plasmid of *KmPGK1* promoter	Addgene ID. 125034
pKmK.T1		Level I plasmid of *KmINU1* terminator	Addgene ID. 125053
pKmK.P2		Level I plasmid of *KmPDC1* promoter	Addgene ID. 125035
pYTK033		Level I plasmid of *Venus*	Addgene ID. 65140
pMTU-DO-HPH-C4		High-copy expression vector for *K. marxianus*, hygromycin selection	Addgene ID. 160343
pKmK.P7		Level I plasmid of *K. marxianus TEF1pr*	Addgene ID. 125040
pKmK.T1		Level I plasmid of *K. marxianus INU1ter*	Addgene ID. 125053
pKmK.P2		Level I plasmid of *K. marxianus INU1pr*	Addgene ID. 125035
pYTK036		Level I plasmid of *Streptococcus pyrogenes Cas9*	Addgene ID. 65143
pYTK033		Level I plasmid of *Venus*	Addgene ID. 65140
pYTK030		Level I plasmid of *S. cerevisiae GAL1pr*	Addgene ID. 65137
pTU-DO-URA		plasmid with GFP dropout with yeast *ScURA3* and bacterial *ampR* markers	This work
pKmK.T5		Level I plasmid of *K. marxianus PGK1ter*	Addgene ID. 125057
pKmK.T3		Level I plasmid of *K. marxianus KMXK_A3020ter*	Addgene ID. 125055
pP7Cas9T1		pTU-DO-URA with insert *KmTEF1pr-SpCas9-INU1t*	This work
pUCC001-*KmURA3*		pUCC001 with *K. marxianus URA3* target insert	Reference 15
pUCC001-*KmHIS3*		pUCC001 with *K. marxianus HIS3* target insert	Reference 15
pUCC001-*KmLEU2*		pUCC001 with *K. marxianus LEU2* target insert	Reference 15
pUCC001-*KmYKU80*		pUCC001 with *K. marxianus YKU80* target insert	Reference 15
pJA_U*gRNAempty*		*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{empty}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_M2 *gRNAempty*		*KmTEF1pr-SpCas9-KmINU1t* 2 X (*ScTDH3pr-HH-gRNA$_{empty}$-HDV-ScCYC1t*) *hphMX- KanR*	This work
pJA_M3 *gRNAempty*		*KmTEF1pr-SpCas9-KmINU1t* 3 X (*ScTDH3pr-HH-gRNA$_{empty}$-HDV-ScCYC1t*) *hphMX- KanR*	This work
pJA_M4 *gRNAempty*		*KmTEF1pr-SpCas9-KmINU1t* 4 X (*ScTDH3pr-HH-gRNA$_{empty}$-HDV-ScCYC1t*) *hphMX- KanR*	This work
pJA_U *URA3*		*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmURA3}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_U *HIS3*		*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmHIS3}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_U *LEU2*		*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmLEU2}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_M2 *UH*		*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmURA3}$-HDV-ScCYC1t ScTDH3pr-HH-gRNA$_{KmHIS3}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_M2 *UL*		*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmURA3}$-HDV-ScCYC1t ScTDH3pr-HH-gRNA$_{KmLEU2}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_M3 *UHL*		*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmURA3}$-HDV-ScCYC1t ScTDH3pr-HH-gRNA$_{KmHIS3}$-HDV-ScCYC1t ScTDH3pr-HH-gRNA$_{KmLEU2}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_M4 delta		*KmTEF1pr-SpCas9-KmINU1t* 4 X (*ScTDH3pr-HH-gRNA$_{delta}$ -HDV-ScCYC1t*) *hphMX- KanR*	This work
pJA_M5 delta		*KmTEF1pr-SpCas9-KmINU1t* 5 X (*ScTDH3pr-HH-gRNA$_{delta}$-HDV-ScCYC1t*) *hphMX- KanR*	This work

pIURA3-MTU-DO	Markerless integrative plasmid with GFP dropout targeting and deleting the locus of *K. marxianus URA3*	This work
pIHIS3-MTU-DO	Markerless integrative plasmid with GFP dropout targeting and deleting the locus of *K. marxianus HIS3*	This work
pILEU2-MTU-DO	Markerless integrative plasmid with GFP dropout targeting and deleting the locus of *K. marxianus LEU2*	This work
pIDelta-MTU-DO	Markerless integrative plasmid with GFP dropout targeting the locus of *K. marxianus* delta sites	This work
pIURA3-P2YT1	Markerless integrative plasmid with *KmPDC1pr- Venus-KmINU1*ter insertion targeting and deleting the locus of *K. marxianus URA3* (for *URA3* YFP cassette)	This work
pIHIS3-P2YT1	Markerless integrative plasmid with *KmPDC1pr- Venus-KmINU1*ter insertion targeting and deleting the locus of *K. marxianus HIS3* (for *HIS3* YFP cassette)	This work
pILEU2-P2YT1	Markerless integrative plasmid with *KmPDC1pr- Venus-KmINU1*ter insertion targeting and deleting the locus of *K. marxianus LEU2* (for *LEU2* YFP cassette)	This work
pIDelta-MTU-DO	Markerless integrative plasmid with *KmPDC1pr- Venus-KmINU1*ter targeting the locus of *K. marxianus* delta sites (for delta site YFP cassette)	This work
pL1_*EcrecA*	Level I plasmid of *S. cerevisiae* codon optimised *E. coli recA* gene	This work
pL1_*ScRAD52*	Level I plasmid of *S. cerevisiae RAD52*	This work
pL1_*Mock gene25*	Level I plasmid of 25bp DNA fragment for connecting YTK standard promoters to terminators	This work
pL1_NM *AtCHS3*	Level I plasmid of *S. cerevisiae* codon optimised *Arabidopsis thaliana CHS3*	This work
pJA_HR-L2.1	pI4-MTU-DO-G418 with insert *ScGAL1pr-EcrecA-KmPGK1ter*	This work
pJA_HR-L2.2	pI4-MTU-DO-G418 with insert *ScGAL1pr-ScRAD52-KmPGK1ter*	This work
pJA_HR-L2.3	pI4-MTU-DO-G418 with insert *ScGAL1pr-KmPGK1ter*	This work
pJA_LII RFP-DO 1	YTK Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 1 of level III (multigene) plasmids	This work
pJA_LII RFP-DO 2	YTK Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 2 of level III (multigene) plasmids	This work
pJA_LII RFP-DO 8	YTK Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 3 of level III (multigene) plasmids	This work
pJA_NM L2. 26	pJA_LII RFP-DO 1 with insert *KmTEF1pr-AtCHS3- KMXK_A03020ter*	This work
pJA_NM L2. 63	pJA_LII RFP-DO 2 with insert *KmPDC1pr- Venus-KmINU1*ter	This work
pJA_NM L2. 33	pJA_LII RFP-DO 8 with insert *KmTEF1pr-AtCHS3- KMXK_A03020ter*	This work
pJA_NM L3.54	pIDelta MTU-DO with insert *KmTEF1pr-AtCHS3- KMXK_A03020ter KmPDC1pr- Venus-KmINU1*ter *KmTEF1pr-AtCHS3-KMXK_A03020ter*	This work
pI4-MTU-DO-G418	Integrative plasmid with GFP dropout targeting locus downstream of *KmHSP104*; *KanMX* marker	Addgene ID. 160215

2.2. *amdS* marker integrative cassettes

Using Q5 High-Fidelity Polymerase*, KMAR_50306* and *KMAR_80308* were amplified from *K. marxianus* genome and inserted into pYTK001 to generate pL1_KMAR_50306 and pL1_KMAR_80308. *S. cerevisiae* codon optimised *amdS* was sourced as gene block from Twist Biosciences (San Francisco, U.S.A.) and similarly inserted into pYTK001 to generate pL1_*amdS.* These genes were then individually inserted into pI4-MTU-DO HPH integrative plasmid by BsaI/T7 ligase Golden Gate reaction including pKmK.P2*,* pKmK.T1 and pL1_*KMAR_50306,* pL1_*KMAR_80308* or pL1_*amdS* to generate plasmids from which cassettes for each gene could be generated by SgsI digestion for yeast integration (pI4-P2-*KMAR_50306*-T1 HPH, pI4-P2-*KMAR_80308*-T1 HPH and pI4-P2-*amdS*-T1 HPH). By using appropriate primers ensuring the presence of YTK standard overhangs for selection markers (type 6) in the amplification product, the entire *KmPDC1pr-amdS-KmINU1ter* was amplified from pI4-P2-*amdS*-T1 HPH and again inserted into pYTK001 to generate pL1_P2*amdS*T1 marker. For constructing pL1_P2*amdS*T1 recyclable marker, two products were amplified from pI4-P2-*amdS*-T1 HPH: (1) *KmPDC1pr-amdS-KmINU1ter,* making up the 5' segment*,* and (2) sequence at the 3' end of ConRE' and 5' end of *KmPDC1* promoter, making up the 3' segment. The final pL1_P2*amdS*T1 recyclable marker has one of the direct repeat sequences of 211bp at the 3' end of *KmINU1ter*. The direct repeat sequence is provided in figure 6. Final integrative plasmids bearing a bacteria functional green fluorescent protein transcriptional unit (GFP dropout) with *amdS* marker or recyclable marker were then constructed for I3 and I4 integrative cassettes as previously described elsewhere (Rajkumar *et al.*, 2019). Briefly, publicly available 5' and 3' homology arms (pKmK.I3L and pKmK.I3R for I3, and pKmK.I4L and pKmK.I4R for I4) were put together with pYTK084, pYTK047, pYTK073 and pL1_P2*amdS*T1 marker or pL1_P2*amdS*T1 recyclable marker. For final plasmids with *amdS* recyclable marker, the 5' direct repeat flank forms as pL1_P2*amdS*T1 recyclable marker assembles with ConRE' from pYTK073 while the 3' direct repeat flank is provided by pL1_P2*amdS*T1 recyclable marker. The direct repeat flanks should allow the excision of *KmPDC1pr-amdS-KmINU1ter* after integration into yeast genome as illustrated in figure 6. GFP dropout in the final integrative plasmids can be replaced by desired single or multiple transcriptional units in BsaI/T7 ligase or BsmBI/T7 ligase Golden Gate reactions, respectively. This allows for green/white selection of successfully assembled plasmids during bacterial transformation. These final plasmids are named pI3-MTU-DO-AMDS, pI4-MTU-DO-AMDS, pI3-MTU-DO AMDS recyclable and pI4-MTU-DO AMDS recyclable. *KmPGK1-Venus-KmPDC1ter* was then inserted into the plasmids by mean of BsaI/T7 ligase Golden Gate reactions including pKmK.P1 (*KmPGK1* promoter), pYTK033 (*venus*), pKmK.T4 (*KmPDC1*

promoter) and the plasmids to generate pI3-P1YT4 AMDS, pI4-P1YT4 AMDS, pI3-P1YT4 AMDS recyclable and pI4-P1YT4 AMDS recyclable. Correct assemblies of the plasmids were confirmed by colony PCR with primer pairs ASR_K001F/ASR_P001R and ASR_T004F/ASR_K007R and subsequently by NotI restriction.

Table 3. CRISPR targets in this study.

Target locus	Sequence	Source
KmYKU80	AACCCTCTGATATCGATACC	(Rajkumar et al. 2019)
KmURA3	AGGAAGTTACGAAAGAACCT	(Rajkumar et al. 2019)
KmHIS3	AGCGAAGCACTCTGGTTGGT	(Rajkumar et al. 2019)
KmLEU2	TTGAAACCTGAGTACGCCAA	(Rajkumar et al. 2019)
K. marxianus delta sites	AAAGCATTTGATAATTGAAT	This work
KMAR_20785	ATGTGGGGGTAGCTCTGGTG	This work
KMAR_40009	CCAGCTGACCCAACATACCT	This work
KMAR_50306	CTCTCTACTAAACTCTCTCC	This work
KMAR_80308	CGATGGTCGTCCCACTCCCA	This work

2.3. Markerless integrative cassettes.

First, markerless integrative plasmids harbouring a GFP dropout were constructed as previously described elsewhere (Rajkumar *et al.*, 2019) and above for the construction of *amdS* marker integrative plasmids. Briefly, the 5' and 3' homology arms for DSB repair by HDR were sourced by amplification from genomic DNA using Q5 High-Fidelity Polymerase and appropriate primers ensuring the presence of YTK standard overhangs in the amplification products for constructing the integrative plasmid. Importantly, the BsaI generated 5' overhangs of 3' homology arms (YTK part type 7) were substituted for that of yeast selectable markers (type 6) to enable the construction of the plasmid by BsaI/T7 Golden Gate reaction without a type 6 part. Homology arms were put together with pYTK084, pYTK047, pYTK073 in BsaI/T7 Golden Gate reactions to generate markerless integrative GFP dropout plasmids for each locus. The GFP dropout can be replaced by desired single or multiple transcriptional units in BsaI/T7 ligase or BsmBI/T7 ligase Golden Gate reactions, respectively. This allows for green/white selection of successfully constructed plasmids during bacterial transformation. The markerless plasmids were then put together with plasmids pKmK.P2 (*KmPDC1* promoter), pYTK033 (*Venus*) and pKmK.T1 (*KmINU1* terminator) in a BsaI/T7 ligase Golden Gate reaction to insert *KmPDC1pr-venus-KmINU1ter* into the plasmids, generating plasmids

from which YFP integrative cassettes can be generated through SgsI linearisation of the plasmids. Correct assemblies of the plasmids were confirmed by PCR with primer pairs ASR_K001F/ASR_P002R and ASR_T001F/ASR_K007R and subsequently by NotI restriction.

For constructing the plasmid used for integrating *Arabidopsis thaliana CHS*3 at delta sites, *S. cerevisiae* codon optimised *AtCHS3* bearing BsaI and BsmBI sites at its 5' and 3' ends was sourced as a gene block from Twist Biosciences (San Francisco, U.S.A.) to generate its Level I plasmid. Level II plasmids of *AtCHS3* were then constructed by cloning *KmTEF1pr-AtCHS3-KMXK_A03020ter* into pJA_LII RFP-DO 1 and pJA_LII RFP-DO 8 plasmids intended for transcriptional units at position 1 and at terminal position 3 of level III (multigene) plasmids, respectively. *KmPDC1pr-Venus-KmINU1ter* was likewise cloned into pJA_LII RFP-DO 2 intended for position 2. The three plasmids were then put together with the delta site integrative plasmid (pIDelta MTU-DO) in a BsmBI/T7 ligase Golden Gate reaction to construct the multigene integrative plasmid of 2 copies of *AtCHS3* and 1 copy of *Venus* (pJA_NM L3.54). The plasmid was confirmed by PCR with primer pairs ASR_K001F/ASR_P008R, ASR_T003F/ASR_P002R and ASR_T003F/ASR_K007R and subsequently by NotI restriction.

2.4. Integrative cassettes for expressing recombinases.

S. cerevisiae codon optimised *EcrecA* and *RAD52* genes were sourced as gene blocks from Twist Biosciences and cloned into level-I plasmids pL1_*EcrecA* and pL1_*ScRAD52*, respectively. The previously described dominant *KanMX* marker integrative plasmid, pI4-MTU-DO-G418, for yeast selection in G418 antibiotic (Rajkumar *et al.*, 2019) was put together in BsaI/T7 Golden Gate reactions with pYTK030 (*ScGAl1pr*) and pKmK.T5 (*KmPGK1ter*) in addition to pL1_*EcrecA*, pL1_*ScRAD52* or L1_*Mock gene25* to construct integrative plasmids pJA_HR-L2.1, pJA_HR-L2.2 or pJA_HR-L2.3, respectively. Correct assemblies of these plasmids were confirmed by PCR using primer pairs ASR_K001F/JA567, ASR_T005F/ASR_P002R and subsequently by NotI restriction. The subsequent integration of cassettes derived from these plasmids into the genome of *K. marxianus* allowed for galactose inducible expression of *EcrecA* and *ScRAD52*.

Table 4. Primers used in this study.

Name	Sequence	Usage
JA494	GCATCGTCTCATCGGTCTCACCCTTTCAGGCGCGCCATTTGGGACTCCAAAACTTATATTTTGAAC	Forward primer for amplifying *KmURA3* 5' homology arm from genomic DNA
JA183	ATGCCGTCTCACAGGTCTCACGTTCGTCTCATCAGTCTAGATGCGAATTCCTCAAAGCTGAAATCCTCCAAGATATC	Reverse primer for amplifying *KmURA3* 5' homology arm from genomic DNA
JA181	GCATCGTCTCATCGGTCTCATACATTTGGGACAACAATACAGAACTGTG	Forward primer for amplifying *KmURA3* 3' homology arm from genomic DNA
JA495	ATGCCGTCTCAGGTCTCATCGGCTGAGGCGCGCCAGTGAGTAGTAGTGGTTGATATTTGAAATCTA	Reverse primer for amplifying *KmURA3* 3' homology arm from genomic DNA
JA172	CCTTAGTAACTATTCCAATGGAGG	Forward primer priming immediately outside the 5' homology arm sequence of *KmURA3*
JA173	ATTCTTATGACATCGGCCATC	Reverse primer priming immediately outside the 3' homology arm sequence of *KmURA3*
JA490	GCATCGTCTCATCGGTCTCACCCTTTCAGGCGCGCCCTGTAATGGATGTGTTTTAGCATACCC	Forward primer for amplifying *KmHIS3* 5' homology arm from genomic DNA
JA189	ATGCCGTCTCACAGGTCTCACGTTCGTCTCATCAGTCTAGATGCGAATTCATTGAGATATGCCCCCCATGTA	Reverse primer for amplifying *KmHIS3* 5' homology arm from genomic DNA
JA186	GCATCGTCTCATCGGTCTCATACAGATTTGGGTTTAAAAAGAGAAAAAATCGGTG	Forward primer for amplifying *KmHIS3* 3' homology arm from genomic DNA
JA491	ATGCCGTCTCAGGTCTCATCGGCTGAGGCGCGCCACACCAAGTCGTTAAAAACTCTAGC	Reverse primer for amplifying *KmHIS3* 3' homology arm from genomic DNA
JA168	CCATAAATGTAATCAACCAATGGA	Forward primer priming immediately outside the 5' homology arm sequence of *KmHIS3*
JA169	AGAATCACTTGCAGGGATCA	Reverse primer priming immediately outside the 3' homology arm sequence of *KmHIS3*
JA492	GCATCGTCTCATCGGTCTCACCCTTTCAGGCGCGCCAATTGGGAAGCGAAGGTATTATCT	Forward primer for amplifying *KmLEU2* 5' homology arm from genomic DNA
JA135	ATGCCGTCTCACAGGTCTCACGTTCGTCTCATCAGTCTAGATGCGAATTCCCAGTAGCATCAATTGCAGC	Reverse primer for amplifying *KmLEU2* 5' homology arm from genomic DNA
JA138	GCATCGTCTCATCGGTCTCATACATGGCATCATCTAGATTGTGGAGAAA	Forward primer for amplifying *KmLEU2* 3' homology arm from genomic DNA
JA493	ATGCCGTCTCAGGTCTCATCGGCTGAGGCGCGCCAAACCATAAACAGAGCTTCTAAATACTCTG	Reverse primer for amplifying *KmLEU2* 3' homology arm from genomic DNA
JA170	GACAAGGAACCAAAGGAAATTATTC	Forward primer priming immediately outside the 5' homology arm sequence of *KmLEU2*
JA171	GGTATTGACTTGAACTCTATTGAATG	Reverse primer priming immediately outside the 3' homology arm sequence of *KmLEU2*

JA063	AAACAGCAGAGTTGTTAAGATTAGTTGAGGTTTTGGGTCCATATATCTGTCTATTGAAGACACATG TAGATATCTTGGAGGATTTCAGCTTTGAGTTTGGGACAACAATA	Forward primer for *KmURA3* repair DNA
JA064	GATCTCTTCCCTTTGCGAAAAGACCTCTACCAACAATAATGATGTCTGATCCACCGGCAACAACTT CATCCACAGTTCTGTATTGTTGTCCCAAACTCAAAGCTGAAATC	Reverse primer for *KmURA3* repair DNA
JA059	ATGACATACCCAGAAAGGAAAGCTTTTGTGTCTAGAATAACAAATGAGACAAAAATTCAGATAGCC ATATCCTTACATGGGGGGCATATCTCAATGATTTGGGTTTAAAA	Forward primer for *KmHIS3* repair DNA
JA060	AAAGTTACTCTTGCAGCTTCCGCAAATGACTCCAAGAAATGTGGTATCATTTCACATGAAAGATCA CCGATTTTTTCTCTTTTTAAACCCAAATCATTGAGATATGCCCC	Reverse primer for *KmHIS3* repair DNA
JA061	ATCAAAGTTTTGAACGCTATTTCAGAAGCACGCCCTTCTATCAAGTTTAATTTTGAACATCATTTGA TTGGTGGTGCTGCAATTGATGCTACTGGTGGCATCATCTAGAT	Forward primer for *KmLEU2* repair DNA
JA062	AGCGGAATCGATCAATTGATGTTGGACAGTCAATTGAGGGAACTCATTCTTGATGGTTTCTTCTAC TGTTTTTCTCCACAATCTAGATGATGCCACCAGTAGCATCAATT	Reverse primer for *KmLEU2* repair DNA
JA080	AGCAGCTGCTCATAGAAGT	Forward diagnostic primers for detecting *KmURA3* deletion
JA081	GGATCTGCCTACTCTCTTCAA	Reverse diagnostic primers for detecting *KmURA3* deletion
JA065	CCAGAAAGGAAAGCTTTTGTGTCTAGAATAAC	Forward diagnostic primers for detecting *KmHIS3* deletion
JA319	GCTGACTCACTTCTGTGATGATCATTAAAGC	Reverse diagnostic primers for detecting *KmHIS3* deletion
JA320	TCAAAGAATATCGTTGTCTTACCAG	Forward diagnostic primers for detecting *KmLEU2* deletion
JA321	TGCCAAAATCTCCTTTACAGC	Reverse diagnostic primers for detecting *KmLEU2* deletion
JA265	GCATCGTCTCATCGGTCTCATACATATAAGTCAAGTCAGAAGAAACCAAGGAGAACAACAGTAAA TACTTACTTATCTAGTGAATCCAACA	Forward primer for constructing 5’ homology arm for integration at delta site
KA266	ATGCCGTCTCAGGTCTCATCGGCTGAGGCGCGCCTGTTGGATTAATTAATTGGTAATTGATTTGT TGGATTCACTAGATAAGTAAGTATTT	Reverse primer for constructing 5’ homology arm for integration at delta site
JA267	GCATCGTCTCATCGGTCTCACCCTTTCAGGCGCGCCTTGATCCTGGACGTCCTACGACGTTGTCT AGGGCGCCAGATATTATACCAACTATAAAAGCTAAGGCTAA	Forward primer for constructing 3’ homology arm for integration at delta site
JA268	ATGCCGTCTCACAGGTCTCACGTTCGTCTCATCAGTCTAGATGCGAATTCGCCTTGATCCTAGTA TATCTAGGCTTTAGCCTTAGCTTTTATAGTTGGTATAATA	Reverse primer for constructing 3’ homology arm for integration at delta site
ASR_K001F	TTTGCTGGCCTTTTGCTC	Forward primer, priming at the 3' end of ColE1 in all of YTK backbones for checking level I, II and III plasmids
ASR_P002R	ATGCCGTCTCTCAGGGTCTCACATAGCAATTATTTGGTTTGGGTGT	Reverse primers priming at 3’ end of *KmPDC1* promoter
ASR_T001F	GCATCGTCTCATCGGTCTCATAGTCTGATCTGCTTACTTTACTAACGAC	Forwards primers priming at 5’ end of *KmINU1* terminator
ASR_K007R	CAGTCATCGGTATGATCTG	Reverse primer for checking integrative dropout plasmids and insertion of transcriptional units into the plasmids
ASR_T005F	GCATCGTCTCATCGGTCTCAATCCGTGCATCATTGACCTTTCTAA	Reverse primers priming at 5’ end of *KmPGK1* terminator
ASR_T003F	GCATCGTCTCATCGGTCTCATAGGATATAAAAACCGGTCACGTG	Reverse primers priming at 5’ end of *KMXK_A3020* terminator
ASR_K002R	CATCTGGATTTGTTCAGAACG	Reverse primers for checking Level I plasmids
ASR_K001R	ATTGGTAACTGTCAGACCAAGTTTA	Reverse primers for checking Level II plasmids
JA567	TAGCAATGAGCAGTTAAGCGTA	Reverse primer priming at 3’ end of *S. cerevisiae GAL1* promoter sequence; used to confirm successful insertion of recombinase genes into pI4-MTU-DO-G418 in this study.

JA603	CGTCTCACCAAAAGGGAGTTAGAATCATTTTGAATAAAAAA	Forward primer for amplifying gRNA transcriptional unit intended for single targeting CRISPR plasmid or for position 1 in multiplex CRISPR plasmids
JA604	CGTCTCACATCTCCAAAACCTTCTCAAGCAAGG	Reverse primer for amplifying gRNA transcriptional unit intended for position 1 in multiplex CRISPR plasmids
JA605	CGTCTCAGATGAAGGGAGTTAGAATCATTTTGAATAAAAAA	Forward primer for amplifying gRNA transcriptional unit intended for position 2 in multiplex CRISPR plasmids
JA606	CGTCTCAGAACCCAAAACCTTCTCAAGCAAGG	Reverse primer for amplifying gRNA transcriptional unit intended for position 2 in multiplex CRISPR plasmids
JA607	CGTCTCAGTTCAAGGGAGTTAGAATCATTTTGAATAAAAAA	Forward primer for amplifying gRNA transcriptional unit intended for position 3 in multiplex CRISPR plasmids
JA608	CGTCTCATACCCCAAAACCTTCTCAAGCAAGG	Reverse primer for amplifying gRNA transcriptional unit intended for position 3 in multiplex CRISPR plasmids
JA609	CGTCTCAGGTAAAGGGAGTTAGAATCATTTTGAATAAAAAA	Forward primer for amplifying gRNA transcriptional unit intended for position 4 in multiplex CRISPR plasmids
JA610	CGTCTCAACTTCCAAAACCTTCTCAAGCAAGG	Reverse primer for amplifying gRNA transcriptional unit intended for position 4 in multiplex CRISPR plasmids
JA611	CGTCTCAAAGTAAGGGAGTTAGAATCATTTTGAATAAAAAA	Forward primer for amplifying gRNA transcriptional unit intended for position 5 in multiplex CRISPR plasmids
JA612	CGTCTCATGCTCCAAAACCTTCTCAAGCAAGG	Reverse primer for amplifying gRNA transcriptional unit intended for single targeting CRISPR plasmid or for terminal position in multiplex CRISPR plasmids
JA128	GTACCTTGTGTAAGATGGCCCAGCGCTGGAGATGCAGTGGGATGGCTGGGGGCAGGTCAAGAC TATGGGCAGACTTTTGTCCACTCTTGCCCGACATTCGTCTCTGCCTT	Forward primer for *KmYKU80* repair DNA
JA129	ATGAGCTGGAAACGTTTGACCAGCAGTGTCCGTTTGAGGTGGTACTGGTTTCTCCCTCTGCCCTG CCCTCTGCCTTGCCAAAGGCAGAGACGAATGTCGGGCAAGAGTGG	Reverse primer for *KmYKU80* repair DNA
JA130	GGAGGGGTCCTATGGGACAA	Forward diagnostic primers for detecting *KmYKU80* deletion
JA131	AAAGATGAGCAAGATGAGGAGACGGAT	Reverse diagnostic primers for detecting *KmYKU80* deletion
JA344	ATTCCAAACTCTTCCCAGAATGGCTGATCCCCGAAGATAAATTGCCTGGAGCTGACGTCAAAAAT GTTTTGGCATGGCCTGTTACTTCAGGCTTTACCCTGACATGGTGG	Forward primer for *KMAR_20785* repair DNA
JA345	AGATATGTGTCTCATGAGACTAACGACCTTCTCGTCTTGGGCTTTCTTGCAAACGACCTGAATGCT GACTGGTCCACCATCCACCATGTCAGGGTAAAGCCTGAAGTAAC	Reverse primer for *KMAR_20785* repair DNA
JA346	GCATCGTCTCATCGGTCTCATATGACAAAGGGTTCGTATCTTG	Forward diagnostic primer for detecting *KMAR_20785* deletion
JA347	ATGCCGTCTCAGGTCTCAGGATCTAATCTTTCAAAGCCGATTTTGT	Reverse diagnostic primer for detecting *KMAR_20785* deletion
JA352	AACCCAAAGCTTCCACAAAGCTTCAGGAAGCAAGATCAGAATTGCCTGGTCATTGGGAATTGCCT GTTGATGTGTTGGAAAAATTTTCCATGACTGGAATTACTTCAACC	Forward primer for *KMAR_40009* repair DNA
JA353	CCTGAAAGGTTGGAGTTACGATCTGAATATGACCAGGGGCTCCTAAGTAAGTATTCTCTTTATAAA CCGGATATGTTCGAGGTTGAAGTAATTCCAGTCATGGAAAATTT	Reverse primer for *KMAR_40009* repair DNA

JA354	GCATCGTCTCATCGGTCTCATATGACAATTTCCGAAATCGCTC	Forward diagnostic primer for detecting *KMAR_40009* deletion
JA355	ATGCCGTCTCAGGTCTCAGGATCTAAGCATTCAAAACACTGTCGA	Reverse diagnostic primer for detecting *KMAR_40009* deletion
JA360	GGTTCTTGATACAGAGCCAATTGATGCTAGGAAGTTCTTGGTAGAGCTGCTTTCGGAGAAAGATT TAAAGATAGTCAATGGGTACACTATTGATGTGATAAGTCTGGATC	Forward primer for *KMAR_50306* repair DNA
JA361	CAGATTCGTCGTCATATCTCTTACCTGTTAATTGTAAAGCAACAGGGCCATTGACAAATTCGTTCG CGTCATATTTCAATGATCCAGACTTATCACATCAATAGTGTACC	Reverse primer for *KMAR_50306* repair DNA
JA580	GCATCGTCTCATCGGTCTCATATGACGGTCGACGCAG	Forward diagnostic primer for detecting *KMAR_50306* deletion and for amplifying the gene for insertion into pYTK001 to generate pL1 *KMAR_50306*
JA581	ATGCCGTCTCAGGTCTCAGGATTTAGCGTCTTTCCACGTTTAGTAC	Reverse diagnostic primer for detecting *KMAR_50306* deletion and for amplifying the gene for insertion into pYTK001 to generate pL1 *KMAR_50306*
JA368	CGATGCGTACAACAAGGAAAGAGAGGACCAGATCAAGGCGTACGATCCGGAGTTTGTAGCCAAG TGCGAGAAGTTGCTACCCAGCGCAGATGTGCGGTATAATGATGACG	Forward primer for *KMAR_80308* repair DNA
JA369	TTCATCGTGGTAACGCTTTCCTGTTAATTGCAAAGCGACAGGAGCGTTTTCGTATTCCTTGCTGTT GTATCTGATGGACCCGTCATCATTATACCGCACATCTGCGCTGG	Reverse primer for *KMAR_80308* repair DNA
JA576	GCATCGTCTCATCGGTCTCATATGACGTATTTAACGGAAAACCCCG	Forward diagnostic primer for detecting *KMAR_80308* deletion and for amplifying 5' segment of the gene for insertion into pYTK001 to generate pL1 *KMAR_80308*
JA577	CACGTCTCACTGACGTACGAGGCGTTCGTTGG	Reverse primer for amplifying 5' segment of *KMAR_80308* for insertion into pYTK001 to generate pL1 *KMAR_80308*
JA578	TTCGTCTCATCAGGTACATCGATTTTGTCCCGGAA	Forward primer for amplifying 3' segment of *KMAR_80308* for insertion into pYTK001 to generate pL1 *KMAR_80308*
JA579	ATGCCGTCTCAGGTCTCAGGATTCAAACTCTGGTAACATTCAAAAGTTCGG	Reverse diagnostic primer for detecting *KMAR_80308* deletion and for amplifying 3' segment of the gene for insertion into pYTK001 to generate pL1 *KMAR_80308*
JA547	GCATCGTCTCATCGGTCTCATACAATCCAGCGAATATACAGCGT	Forward primer for amplifying *KmPDC1pr-amdS-KmINU1ter* from pI4-P2 *amdS* T1 HPH for insertion into pYTK001 to generate pL1_P2*amdS*T1 marker and pL1_P2*amdS*T1 recyclable marker
JA548	ATGCCGTCTCAGGTCTCAACTCCACCTTTCTAAGGTTAGAATAGAAAT	Reverse primer for amplifying *KmPDC1pr-amdS-KmINU1ter* from pI4-P2 *amdS* T1 HPH for insertion into pYTK001 to generate pL1_P2*amdS*T1 marker
JA586	CACGTCTCACACCTTTCTAAGGTTAGAATAGAAATTTTTT	Reverse primer for amplifying *KmPDC1pr-amdS-KmINU1ter* from pI4-P2 *amdS* T1 HPH for insertion into pYTK001 to generate pL1_P2*amdS*T1 recyclable marker
JA587	TTCGTCTCAGGTGTTACAGAGATGTTACGAACCACTA	Forward primer for amplifying 3' segment of ConRE' and 5' segment of *KmPDC1* promoter from pI4-P2 *amdS* T1 HPH

JA588	ATGCCGTCTCAGGTCTCAACTCGGTGAGATTTGTTCATCCCG	Reverse primer for amplifying 3' segment of ConRE' and 5' segment of *KmPDC1* promoter from pI4-P2 *amdS* T1 HPH
ASR_K002R	CATCTGGATTTGTTCAGAACG	Reverse primers for checking Level I plasmids
ASR_I3_DS_F	ATTCCAGACACAGGTGGAA	Forward primer for checking cassette integration at I3
ASR_I3_DS_R	AAAGGAAGGATTCGGCGTTA	Reverse primer for checking cassette integration at I3
ASR_I4_DS_F	AGTAGTGAGTGACAGACAC	Forward primer for checking cassette integration at I4
ASR_I4_DS_R	GCAGTTTCTGTGCAGAAAA	Reverse primer for checking cassette integration at I4
ASR_K007R	CAGTCATCGGTATGATCTG	Reverse primer for checking integrative dropout plasmids and insertion of transcriptional units into the plasmids
ASR_P002R	ATGCCGTCTCTCAGGGTCTCACATAGCAATTATTTGGTTTGGGTGT	Reverse primer priming at 3' end of *KmPDC1* promoter
ASR_P001R	ATGCCGTCTCTCAGGGTCTCACATATTTGTATCTTTATATAGGTAGTGTGTA	Reverse primer priming at 3' end of *KmPGK1* promoter
ASR_T001F	GCATCGTCTCATCGGTCTCATAGTCTGATCTGCTTACTTTACTAACGAC	Forwards primer priming at 5' end of *KmINU1* terminator
ASR_T004F	GCATCGTCTCATCGGTCTCATAGGAGAGGGAGAGAGGATAAAGAGAT	Forwards primer priming at 5' end of *KmPDC1* terminator
Bsa-R	TACACGCGTTTGTACAGAAAAAAAAGAAAAATTTGA	Reverse primer for verifying protospacer sequence insertion into pUCC001.
JA623	GAATTCGATATCAAGCTTATCGATACCGTCGATGCCACAATCTTGGGAA	Forward primer for amplifying *amdS* for *in vivo* assembly with pGREG-505-TEF1 in BY4741
JA624	GCGTGACATAACTAATTACATGACTCGAGGTCGACTTATGGAGTAACAACGTTACC	Reverse primer for amplifying *amdS* for *in vivo* assembly with pGREG-505-TEF1 in BY4741

3. Yeast transformation and confirmation of genomic modifications.

A variation of the lithium acetate/PEG transformation method was used to introduce functional DNA molecules into yeasts throughout this work (Rajkumar and Morrissey, 2022). Briefly, yeast strains were cultured overnight in YPD and reinoculated at A 600nm of 0.5 in 50mL of fresh YPD in 250mL Erlenmeyer flasks and further cultured for 3 hours before harvesting for transformation. For experiments involving chemical synchronisation of cells into S/G2 phase, fresh YPD cultures were supplemented with 2.5mL of water (untreated) or 2M (20X) hydroxyurea (treated) after 1 or 3 hours and then further cultured for 2 hours before harvesting for transformation by washing once with water. For experiments involving heterologous expression of recombinases, YPD was entirely replaced with YPGalactose to induce the expression of recombinases in strains KmJA.098 and KmJA.099. For DSB repair by HDR during simple gene disruption, short (190bp) repair DNAs were designed and prepared as previously described (Rajkumar and Morrissey, 2022) while targeting for the purpose of cassette integration was achieved by first digesting integrative plasmids with SgsI (FD1894, Fisher Scientific) in 20µL reactions. Harvested cultures were transformed with 4.5µg of repair DNAs that was purified and concentrated by ethanol precipitation or the entire 20µL of unpurified digested integrative plasmids and appropriate CRISPR plasmid for DSB generation. For multiplex gene targeting, the repair DNAs or digested integrative plasmids for all targeted loci were transformed. For single and multi-loci targeting, 150ng and 500ng of CRISPR plasmids were transformed, respectively. For integrating marker-harbouring cassettes, 2µg of digested plasmids were simply used to transform cells. Transformants were selected on YPD plates supplemented with 400mg/L hygromycin or G418 or on acetamide plates depending on the appropriate selective pressure. For selection on acetamide, cultures for transformation were harvested by washing 50mL of YPD cultures three times with 25 mL of water to prevent the emergence of non-transformants due to remnant nitrogen sources from YPD. pGREG-505-TEF1 was used to establish the expression of *amdS* in BY4741 as previously described (Varela *et al.*, 2017). Briefly, pGREG-505-TEF was digested with SalI (FD0644, Fisher Scientific) and the *amdS* gene was amplified from pUG-*amdSYM* using primers JA623 and JA624. Transformation of BY4741 was subsequently conducted with undigested pGREG-505 or transformation with the digested plasmid and the *amdS* amplification products and selected on solid SD medium lacking leucine for *in vivo* assembly.

Successful modifications were confirmed in emerged transformants by PCR, detection of YFP signal and/or auxotrophy test by patching onto solid media missing appropriate nutrients. For gene deletions, diagnostic primers complementary to internal regions of the gene but outside

the deleted region were used (illustrated as primers A and B in figure 1) while cassette integrations were confirmed with primers A to D as illustrated in figures 4A and 12A. For determining fluorescence, YFP signal of isolates were determined by cultivation in YPD overnight, followed by reinoculation in fresh YPD by diluting down to 100 fold and further cultivating for 3 hours. Cultures were then harvested at mid-exponential growth phase by a single wash with water, resuspended in water and transferred at volumes of 200µL into 96-well plates. Using a Spark Multimode Microplate reader (Tecan), we determined culture A 600nm and fluorescence with excitation and emission of 485 and 535nm, respectively. After correcting for the autofluorescence of wildtype *K. marxianus*, the fluorescence signal was normalised to cell number by dividing by the A 600nm.

4. Counter-selection of *amdS* by fluoroacetamide treatment.

For fluoroacetamide treatment, strains were cultivated in 50mL liquid minimal glucose medium in 250 mL Erlenmeyer flasks overnight for 18hours after which cultures were diluted to A 600nm of 1 and 500 µL of these diluted cultures were spread on solid minimal glucose medium supplemented with 2.3 or 23g/L of fluoroacetamide (fluoroacetamide plates). Once well-defined isolates emerged on fluoroacetamide plates, their growth phenotypes on acetamide were assessed by replica plating to patch them onto acetamide plates.

5. Quantification of naringenin production titre

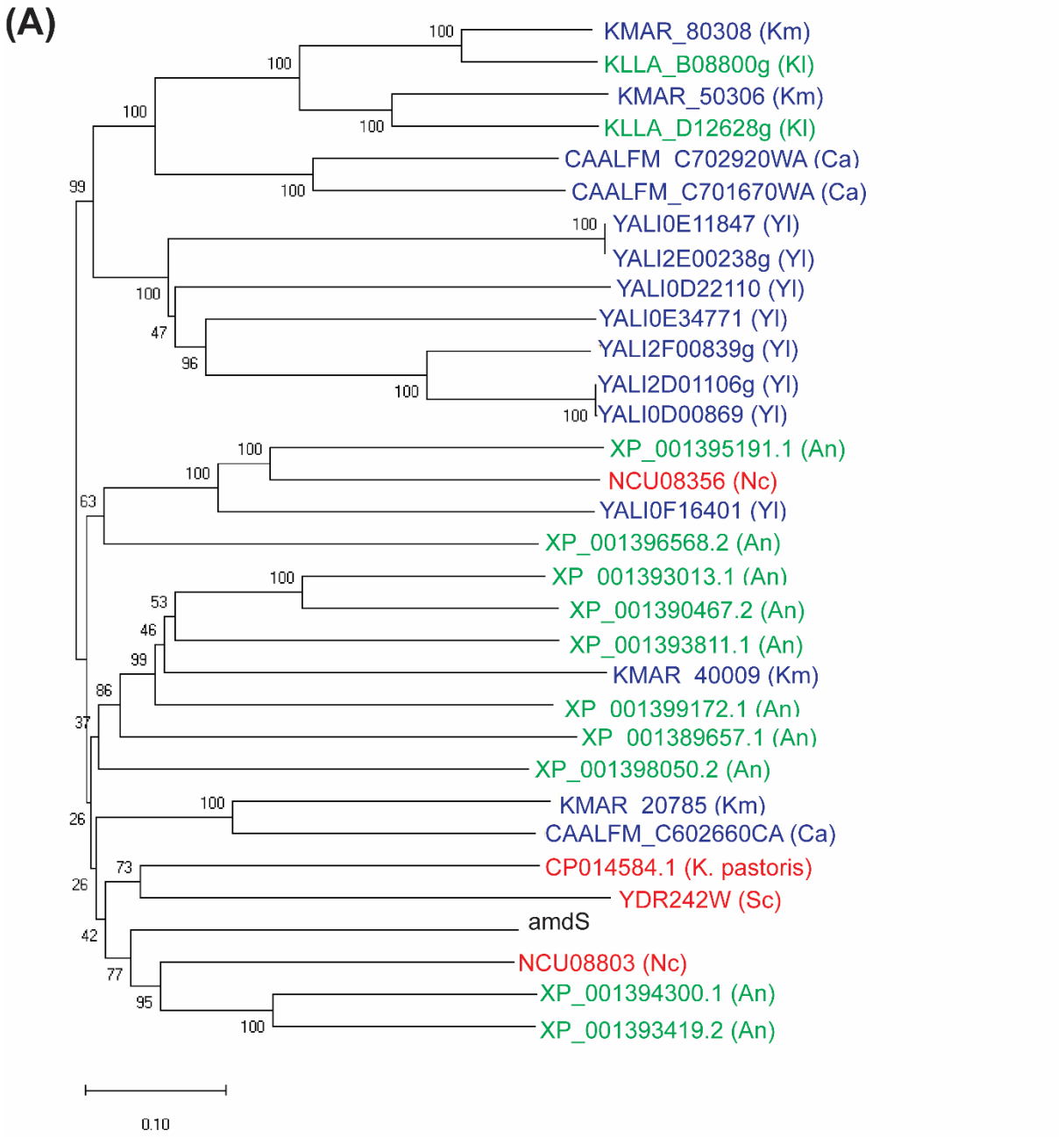
The concentrations of naringenin produced by modified *K. marxianus* strains in this work were determined by mixing 700µL of cultures with 96% ethanol at 1:1 ratio. The mixture was vortexed and pelleted by centrifugation using a benchtop centrifuge at 4°C for 5 minutes at maximum speed. Flavonoids concentrations in supernatants were the determined by an Agilent 1260 Infinity HPLC Infinity system with Zorbax Eclipse Plus C18 column (4.6 x 150mm, 3.5µm; Agilent, Little Island, Ireland) operated at 40°C with 20mM KH_2PO_4 brought to pH2 with phosphoric acid and supplemented with 1% acetonitrile/acetonitrile as the mobile phase at a flow rate of 0.8 mL/min as previously described (Koopman *et al.*, 2012). Detection was by a VWD100 multi-wavelength detector at an absorbance of 200nm.

Results

1. Development of *A. nidulans amdS* gene as a selectable marker in *K. marxianus*

1.1. Bioinformatic analysis of acetamidases in fungi.

S. cerevisiae, Neurospora Crassa and *Candida glabrata* are unable to metabolise acetamide (non-metabolisers) (Yamashiro, Yarden and Yanofsky, 1992; Solis-Escalante *et al.*, 2013; Fu *et al.*, 2016), *Aspergilus niger* and *K. lactis* metabolise it poorly (Kelly and Hynes, 1974; Read *et al.*, 2007), while *Y. lipolytica, C. albicans* and *Komagataella pastoris* are able to metabolise it (good metabolisers) (Fu *et al.*, 2016; Piva *et al.*, 2018; Hamilton *et al.*, 2020). Based on this information, homology search of *A. nidulans amdS* protein was used to identify putative acetamidase genes in the genomes of fungi from the three groups in addition to *K. marxianus* NBRC1777 genome. No hit was obtained for *C. glabrata.* To gain insight into the homology of protein hits from these yeasts, a tree showing the relationship based on sequence identity between the proteins was constructed by means of the neighbour-joining method (Figure 2A). KMAR_20785 and KMAR_40009 in *K. marxianus* showed high sequence identity with the amdS and hits from non-metabolisers and KMAR_50306 and KMAR_80308 with hits from good metabolisers and is distant from *amdS*. The two hits from *K. lactis* (KLLA0_D12628g and KLLA0_B08800g) are homologous to KMAR_50306 and KMAR_80308, indicating that these proteins are orthologues. To gain insight into why the *K. lactis* proteins do not confer acetamide metabolism in the yeast to a similar level as in *K. marxianus*, the sequence of amdS, the 4 *K. marxianus* and the two *K. lactis* hits were aligned alongside the sequences of known amidases from bacteria. The enzymes from these bacteria were chosen for the alignment because their important active site residues have been characterised. The alignment revealed that residues that form the catalytic triad (Ser-cisSer-Lys) (Wu *et al.*, 2020) in addition to Asp191 residue for which substitution was previously reported to reduce the specific activity of the amidase from *R. rhodochrous* (Kobayashi *et al.*, 1997) are conserved in amdS and all 6 *K. marxianus* and *K. lactis* proteins (figure 2B). Thus, we were unable to identify any important residue differences between the proteins and it remains unclear why the *K. lactis* proteins are unable to confer acetamide metabolism in this yeast.


(A)
KMAR_80308 (Km)
KLLA_B08800g (Kl)
KMAR_50306 (Km)
KLLA_D12628g (Kl)
CAALFM_C702920WA (Ca)
CAALFM_C701670WA (Ca)
YALI0E11847 (Yl)
YALI2E00238g (Yl)
YALI0D22110 (Yl)
YALI0E34771 (Yl)
YALI2F00839g (Yl)
YALI2D01106g (Yl)
YALI0D00869 (Yl)
XP_001395191.1 (An)
NCU08356 (Nc)
YALI0F16401 (Yl)
XP_001396568.2 (An)
XP_001393013.1 (An)
XP_001390467.2 (An)
XP_001393811.1 (An)
KMAR_40009 (Km)
XP_001399172.1 (An)
XP_001389657.1 (An)
XP_001398050.2 (An)
KMAR_20785 (Km)
CAALFM_C602660CA (Ca)
CP014584.1 (K. pastoris)
YDR242W (Sc)
amdS
NCU08803 (Nc)
XP_001394300.1 (An)
XP_001393419.2 (An)
0.10

(B)

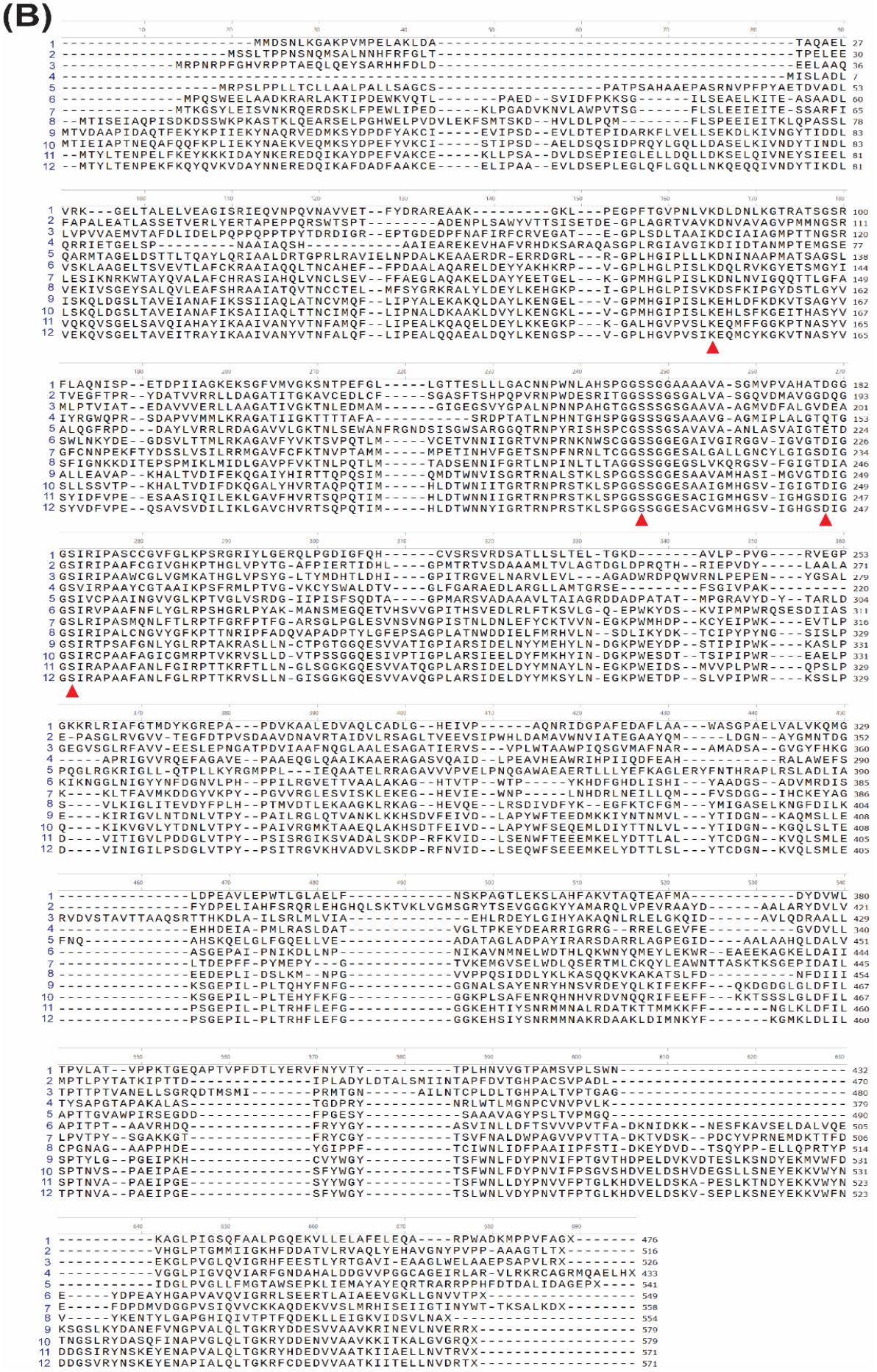


Figure 2. Bioinformatic analysis of acetamidases. (A) Tree showing the relationship based on sequence homology between putative acetamidases from good, poor, and non-metabolising fungi of acetamide. Proteins from good metabolisers are in blue texts, poor metabolisers in green and non-metabolisers in red. The tree was constructed by neighbour-joining method. (B) Alignment of putative *amdS*, putative acetamidases from *K. marxianus* and *K. lactis* and amidases from bacteria. The bacterial amidases were chosen for the alignment because their important active site residues have been characterised. Conserved active site residues are indicated with red triangles. Proteins sequences are listed in the following order: 1. *Parvibaculum lavamentivorans*; 2. *Rhodococcus rhodochrous* (GenBank accession number: BAA03744); 3. *Rhodococcus erythropolis* (GenBank accession number: AY026386)*;* 4. *Bradyrhizobium japonicum* (GenBank accession number: AF012735); 5. *Stenotrophomonas maltophilia* (GenBank acession number: CAC93616); 6. *amdS;* 7. KMAR_20785; 8. KMAR_40009; 9. KMAR_50306; 10. KLLA0_D12628g; 11. KMAR_80308; 12. KLLA0_B08800g.

1.2. Deletion of *KMAR_50306* and *KMAR_80308* caused loss of *K. marxianus* growth on acetamide.

To elucidate which of the genes is responsible for acetamidase activity in *K. marxianus,* we initially disrupted each gene individually in a wildtype strain (KmASR.005) and tested for growth of the deletion strains on acetamide plates. Only the deletion of *KMAR_50306* in strain KmJA.268 caused a moderate growth defect with the other strains growing as well as KmASR005 (figure 3A). Double deletion strains of *KMAR_80306* and the other three genes were constructed by additional deletion of each gene in KmJA.259 (*kmar_80306*). KmJA.272 carrying double deletion of *KMAR_80306* and *KMAR_50306* completely lost the ability to grow on acetamide (figure 3A). The individual integration of either gene or *amdS* for expression under transcriptional control of the native *PDC1* promoter in KmJA.272 revived the growth of the strain (figure 3B). These results clearly show that *KMAR_50306* and *KMAR_80308* are responsible for *K. marxianus* growth on acetamide, their disruption allowed the construction of a *K. marxianus* strain that is unable to grow on acetamide and that expressing the yeast's native or heterologous acetamidases complements this phenotype.

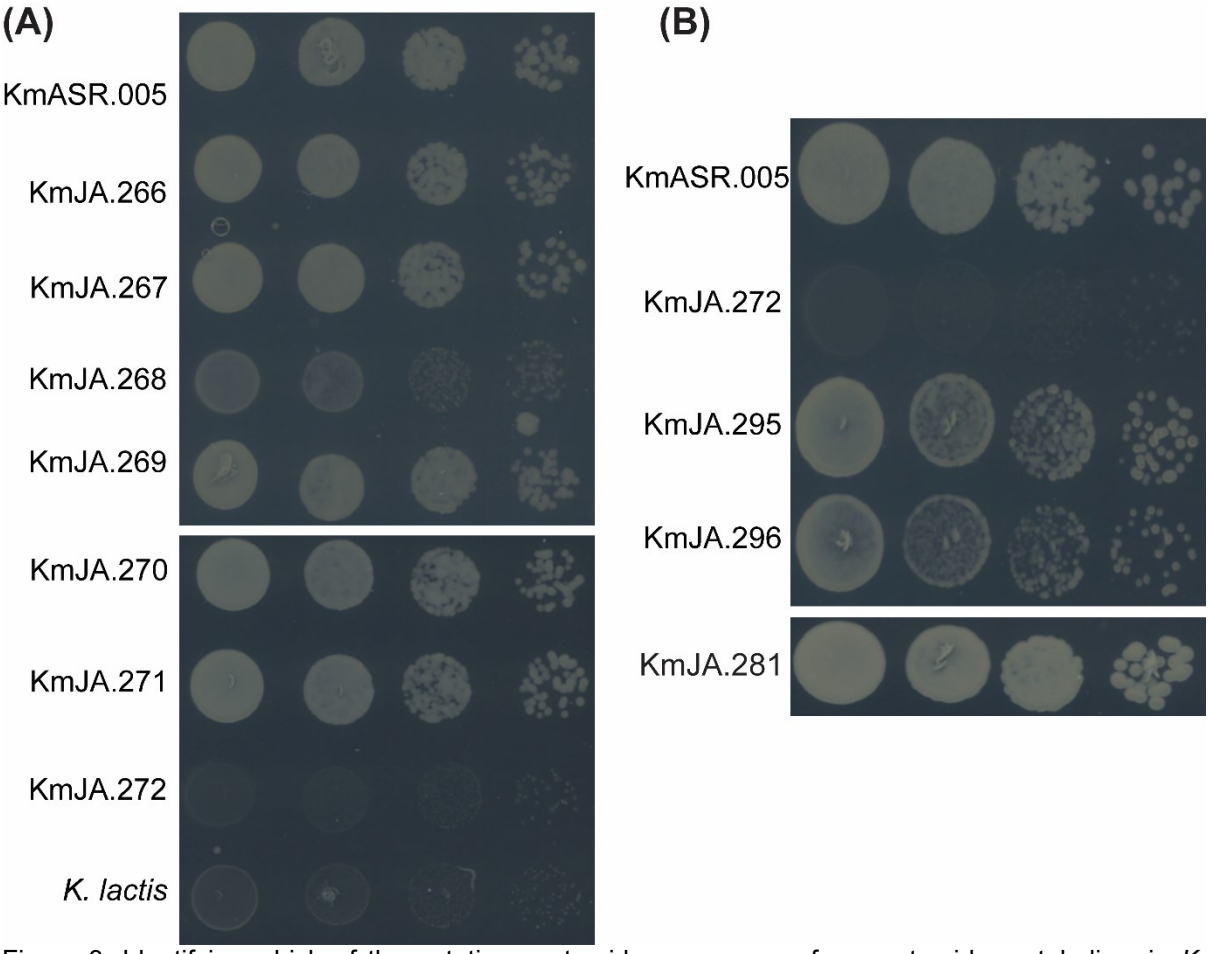


Figure 3. Identifying which of the putative acetamidase genes confers acetamide metabolism in *K. marxianus*. (A) The four putative acetamidase genes of *K. marxianus* were individually disrupted in KmASR.005 (*dln4*) to generate strain KmJA.266 (*kmar_20785*), KmJA.267 (*kmar_40009*), KmJA.268 *(kmar_50306*), KmJA.269 (*kmar_80308*). KmJA.269 was further used to generate strains KmJA.270 (*kmar_80308kmar_20785*), KmJA.271 (*kmar_80308kmar_40009*), KmJA.272 (*kmar_80308kmar_50306*). (B) *KMAR_50306, KMAR_80308* and *amdS* were integrated for expression under transcriptional control of *KmPDC1* promoter in the genome of KmJA.272 to generate strains KmJA.295, KmJA.296 and KmJA.281 respectively.

1.3. Construction and testing of integrative plasmids for genomic integration in *K. marxianus* by *amdS* selection

The result that *amdS* revived KmJA.272 growth on acetamide motivated us to test the ability of the gene to be used as selectable marker for genomic integration in *K. marxianus.* To achieve this, plasmids for genomic integration in *K. marxianus* at loci that we previously evaluated for efficient integrations (I3 and I4) were constructed as illustrated in figure 4A using our publicly available set of parts provided for Golden Gate cloning based on the YTK standard and listed here in Table 2. The yellow fluorescent protein (YFP) reporter gene (*Venus*) from the YTK standard was cloned into the plasmids for expression in *K. marxianus* under transcriptional control of the yeast's native *PGK1* promoter and *PDC1* terminator. YFP-*amdS*

integrative cassettes for both loci were then individually used to transform KmJA.272 on acetamide plates. Integrations were confirmed in transformants by PCR using the primers illustrated as primers A to D and fluorescence detection. Close to one hundred percent integration efficiency was found based on eight randomly picked transformants and all PCR-positive transformants displayed fluorescence (Figure 4B). These results show that *amdS* is applicable as a selectable marker in *K. marxianus* and is now available as a level-I plasmid for constructing additional vectors based on YTK standard by Golden Gate assembly.

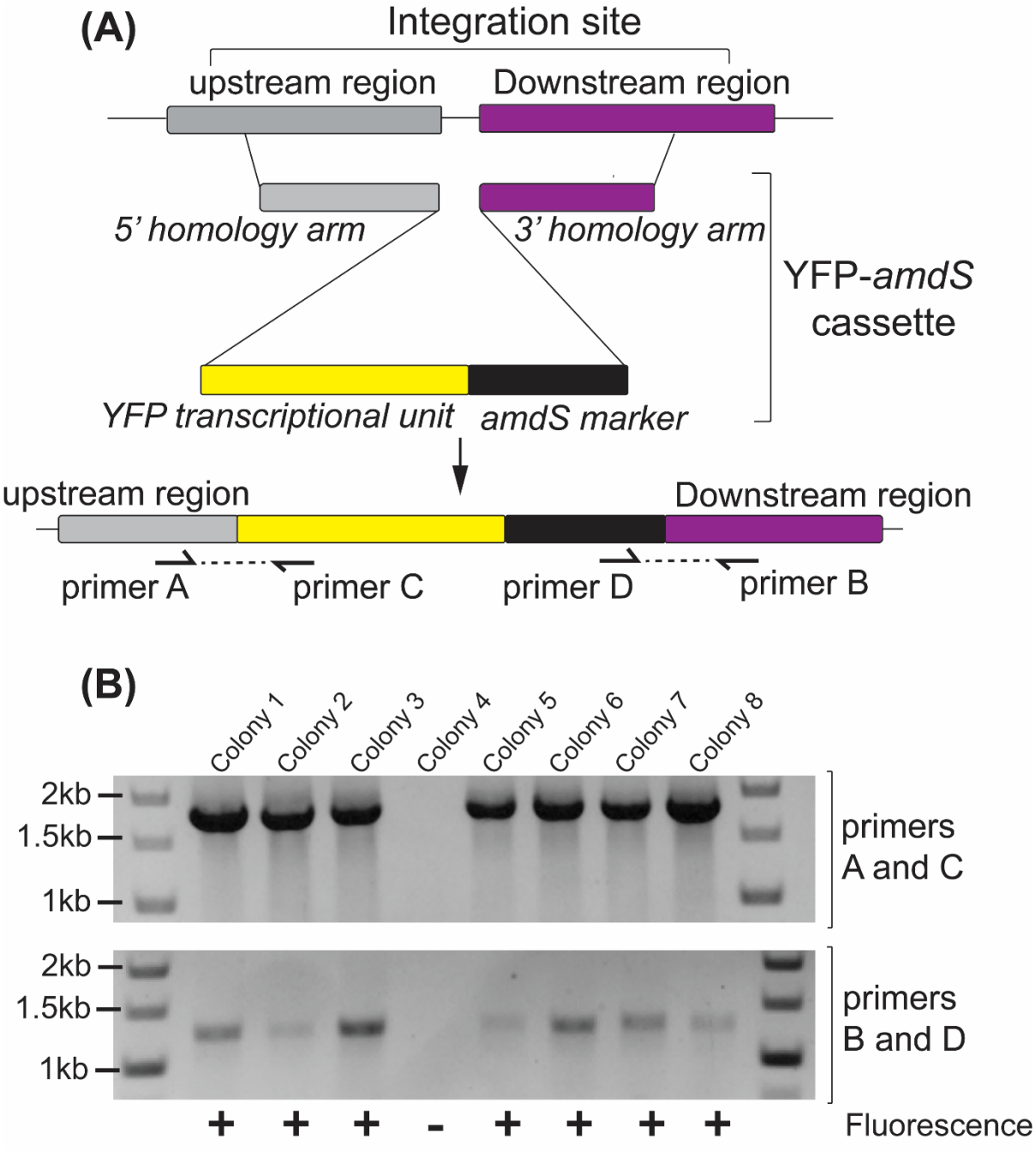


Figure 4. Application of *amdS* as selection marker for genomic integration. (A) YFP reporter cassettes with *amdS* marker were constructed for genomic integration in *K. marxianus* as illustrated. Successful integration is detectable by primer pairs A/C and B/D. (B) Strain KmJA.272 was transformed with the reporter cassette and selected on acetamide plates. Using the primer pairs describe in (A), PCR was undertaken to determine whether integration was successful in eight randomly picked transformants. Fluorescence was also determined in the transformants.

1.4. Expression of *amdS* gene under a strong constitutive promoter confers fluoroacetamide sensitivity of *K. marxianus*.

To test the sensitivities to fluoroacetamide of wildtype *K. marxianus,* KmJA.272, KmJA.318 (KmJA.272 expressing *amdS*), the strains were spotted on plates of ammonium media supplemented with 2.3g/L and 23g/L fluoroacetamide. As controls, the sensitivities of *S. cerevisiae* BY4741 expressing *amdS* from an episomal vector (ScASR.002) or simply harbouring the vector backbone were also tested (ScASR.001). All strains showed poorer growth at higher concentration of fluoroacetamide (figure 5). ScASR.002 grew poorer than ScASR.001 at both concentrations, showing that BY4741 is sensitive to these concentrations of fluoroacetamide when it expresses *amdS*. No difference in growth was observable between KmASR.005 and KmJA.272, showing that the presence of the two acetamidase genes in wildtype *K. marxianus* did not make the yeast more sensitive to fluoroacetamide than when they were disrupted in KmJA.272. On the other hand, KmJA.318 which harbours genomic integration of *amdS* for expression under transcriptional control of the strong constitutive *KmPDC1* promoter from its genome in the absence of the native acetamidases grew poorer than KmJA.272 at both concentrations with clearer difference in growth observable at higher concentration. Put together, these results show that while *K. marxianus* acetamidase genes under native expression on ammonium did not confer fluoroacetamide sensitivity, heterologous expression of an acetamidase gene under transcriptional control of a strong constitutive promoter in *K. marxianus* did. Furthermore, the growth difference between ScASR.001 and ScASR.002 appears to be greater than between KmJA.272 and KmJA.318

at 2.3g/L, suggesting that *K. marxianus* is less sensitive to fluoroacetamide. However, the difference between KmJA.272 and KmJA.318 intensified at 23g/L.

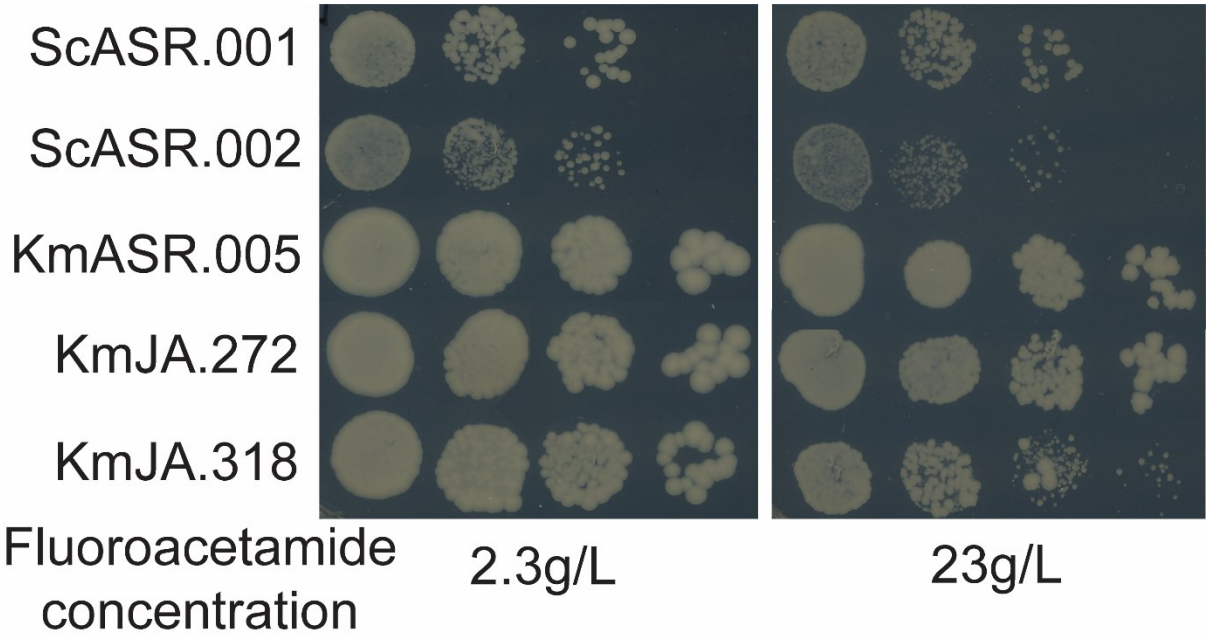


Figure 5. Sensitivity of *K. marxianus* to fluoroacetamide. ScASR.001 and ScASR.002 are *S. cerevisiae* BY4741 strains harbouring pGREG-505 backbone and pGREG with *amdS*, respectively. These strains alongside KmASR.005 (*dln4*), KmJA.272 (*dln4kmar_50306 kmar_80308*) and KmJA.318 (KmJA.272; I3::*amdS*) were streaked or spotted on solid minimal glucose media supplemented with fluoroacetamide at the concentrations indicated. Methionine, uracil and histidine (MUH) were added to media to complement the auxotrophy of BY4741 for these amino acids. Strains were precultured in liquid median for 24 hours before three microlitres of cultures at A 600nm 1, 0.1, 0.01 and 0.001 were spotted.

1.5. Evaluation of *amdS* as a counter-selectable marker in *K. marxianus*.

To evaluate *amdS* as a counter-selectable marker, YFP-*amdS* cassettes with the *amdS* transcriptional unit flanked by direct repeats (YFP-*amdS* recyclable) were constructed as described in the materials and methods section and used to transform KmJA.272 on acetamide plates for integration at the target loci as illustrated in figure 6. Integrative plasmids for I3 and I4 were constructed. Again, almost all randomly picked transformants screened by PCR showed correct integration of the cassettes at I3 and I4 and all PCR-positive colonies displayed fluorescence (not shown). Two strains with correct integrations at I3 (KmJA.355) and I4 (KmJA.356) were cultivated for 24 hours in liquid minimal glucose medium, diluted to A 600nm of 1 and 0.5mL was spread on minimal glucose medium with 0, 2.3 or 23 g/L of fluoroacetamide. Replica plating of well-defined isolates that emerged on the fluoroacetamide plates showed that all isolates maintained their ability to grow on acetamide, showing that none of the isolates were cured of the *amdS* gene (figure 7). Thus the counter-selection was not successful.

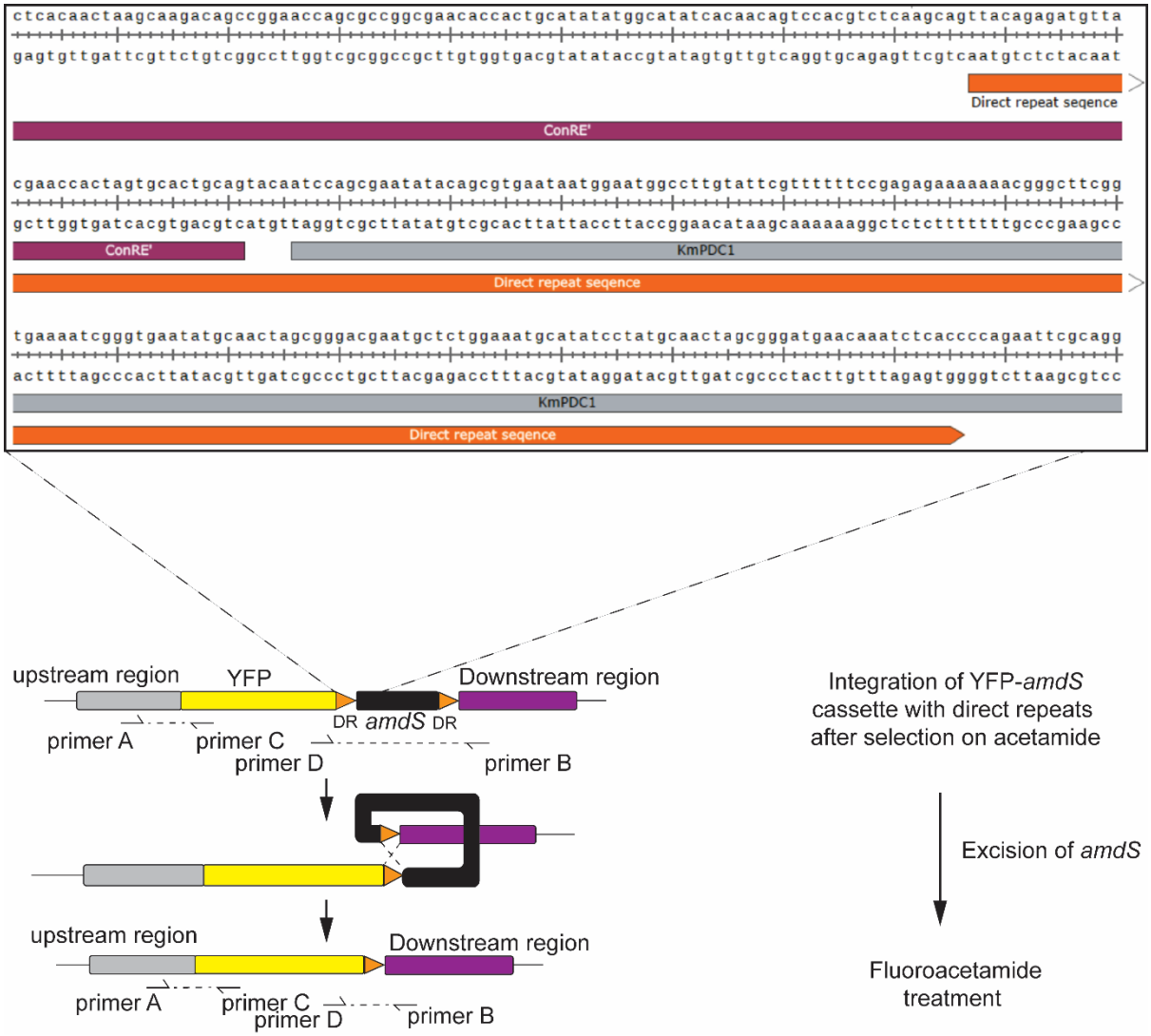


Figure 6. Illustration of *amdS* marker excision mediated by direct repeats for recycling the marker. Once integrated, treatment with fluoroacetamide should select for isolates that have lost the *amdS* gene but with YFP remaining integrated.

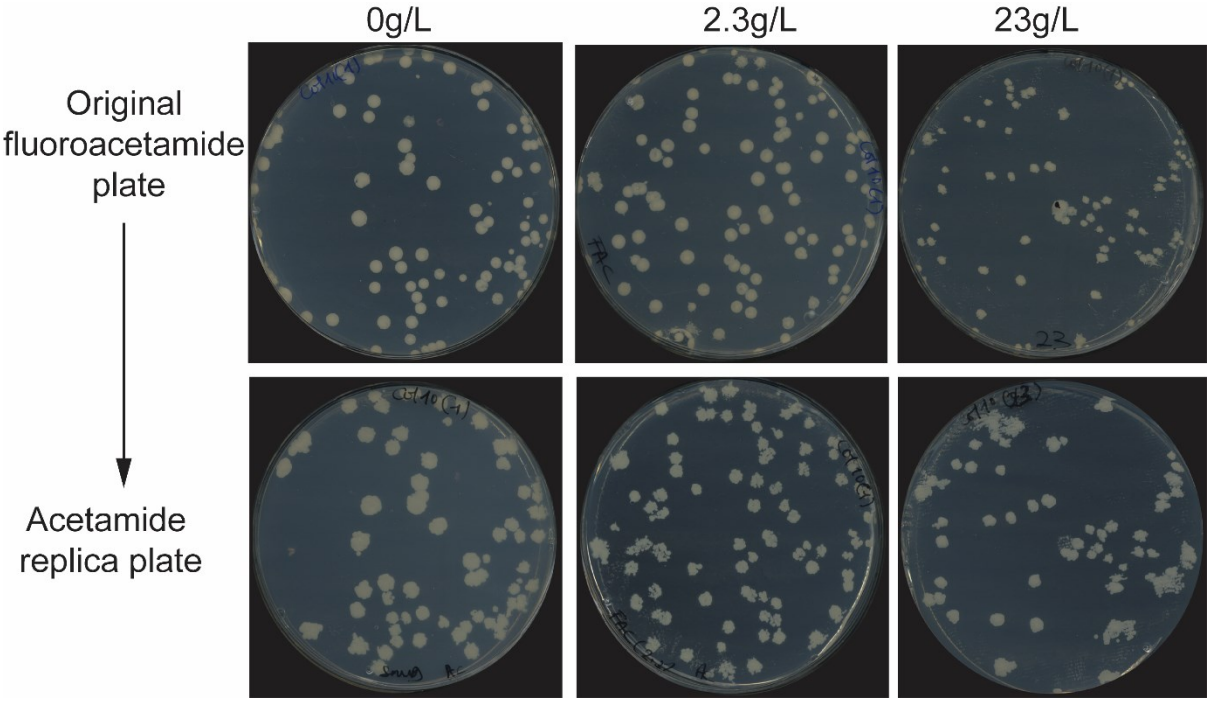


Figure 7. Counterselection of *amdS* on fluoroacetamide. Strain KmJA.355 which has YFP-amdS cassette with direct repeats integrated in its genome was cultured in minimum glucose medium for 18 hours. The culture was then spread on solid minimal glucose medium with the indicated concentrations of fluoroacetamide (original fluoroacetamide plates) to obtain isolates. Isolates obtained on the original fluoroacetamide plates were patched onto acetamide plates by replica plating to determine their acetamide growth phenotypes.

2. Marker-free multi-loci genome editing of *Kluyveromyces marxianus*.

2.1. Design and construction of high copy number CRISPR plasmid for multiplex gene targeting in *K. marxianus*

Efficient CRISPR targeting in *K. marxianus* was previously reported with the aid of plasmid pUCC001 which harbours a pangenomic replication origin, a Cas9 transcriptional unit and ribozyme-flanked gRNAs (Juergens *et al.*, 2018; Rajkumar *et al.*, 2019). Although the pangenomic replication origin makes the plasmid applicable to several yeast species, its copy number when one is interested in specific yeasts may be suboptimal. For this reason, the CRISPR plasmids used in the current study were constructed to contain a *K. marxianus* high copy number replication origin (*ARS2*). It also contained a codon optimised *Spcas9* gene from YTK (Lee et al., 2015) expressed under the control of the *K. marxianus TEF1* promoter and *INU1* terminator. The gRNA expression systems was the same as in pUCC001 in regard to transcription under the control of *ScTDH3* promoter of the self-cleaving ribozyme-flanked gRNA molecules (Juergens *et al.*, 2018). The steps illustrated in figure 8 were followed for designing and constructing CRISPR plasmids with up to 5 targets as described in detail in the materials and methods section and is easily adaptable to any targets of choice.

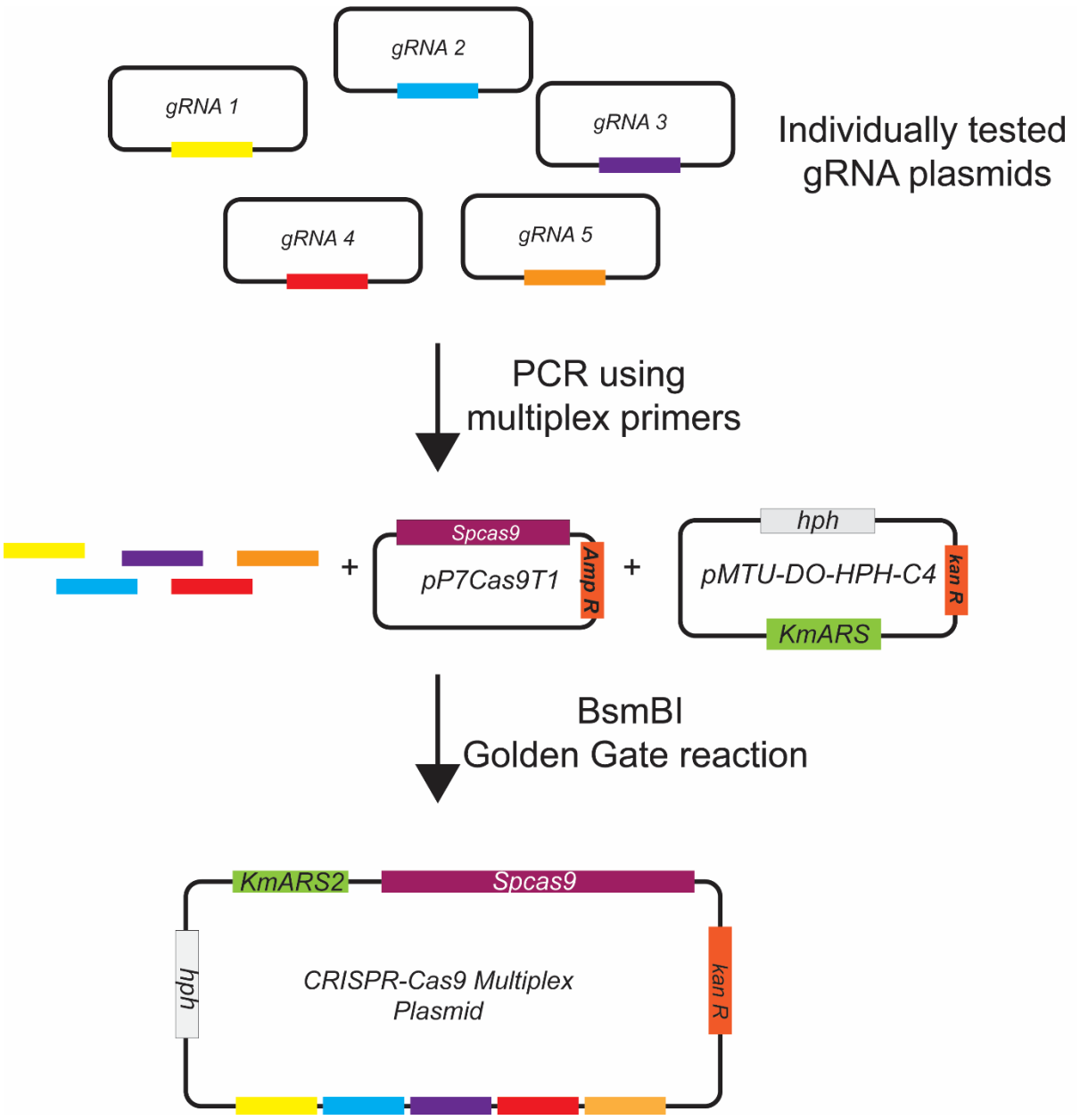


Figure 8. (A) Construction of high copy number CRISPR plasmids for marker-free integration. Transcriptional units of guide RNA (gRNA) targets in pUCC001 were first amplified with specific primer pairs (JA603 to JA612). The primer pair used is determined by the position of the gRNA sequence in the final CRISPR plasmid. The amplification products with 5' and 3' BsmBI Golden Gate overhangs are then purified and put together in a BsmBI/T7 ligase Golden Gate assembly reaction with plasmids pP7Cas9T1 and pMTU-DO-HPH-C4 to generate final CRISPR plasmids with up to five targets, *Streptococcus pyrogenes* Cas9 from Yeast Toolkit for expression in yeasts under transcriptional control of the *KmTEF1* promoter (T7) and *KmINU1* terminator (T1). Plasmid pMTU-DO-HPH-C4 provides a backbone containing bacterial kanamycin-resistance gene, a yeast hygromycin resistance gene and a high copy number replication origin (*KmASRS2*) in the final CRISPR plasmids.

2.2. Repair DNAs rescued wildtype *K. marxianus* from double stranded breakages and increased gene disruption efficiency

We hypothesised that by making HDR available through the provision of repair DNAs that can recombine after DSB generation in addition to NHEJ, it may be possible to increase the number of transformants obtained due to improved cell rescue from DSBs generated by CRISPR plasmids with an expectation that gene disruption efficiency will also increase. This hypothesis was tested by transforming wildtype *K. marxianus* with CRISPR plasmids harbouring targets for DSB generation at the *URA3*, *HIS3* and *LEU2* loci with or without 12.5 ng of repair DNAs/µL of transformation mix (ca. 106fmol/µL). The repair DNAs used here are 190bp DNA fragments that contained 95bp 5' and 3' homology arms. Equimolar concentrations of empty CRISPR plasmids (pJA_U *gRNAempty,* pJA_M2 *gRNAempty,* pJA_M3 *gRNAempty*) were also transformed to give the transformant count of plasmid transformations not leading to DSB generation. These were used to express the transformant count of transformations in which gene(s) were targeted as a percentage of transformation not leading to DSB generation (% relative transformant count). Transformations with repair DNAs showed higher % relative transformant count than transformations without for single gene targeting but no significant difference was observed for multiplex targeting of two or three genes (Figure 9A). No significant difference in disruption efficiency was found for single or double targeting, but triple gene disruptions were only achieved with repair DNAs (figure 9B). To gain insight into how these genes were disrupted when repair DNAs were transformed, PCR was conducted for random transformants in the triple gene targeting experiment. This showed that most of the genes were disrupted by repair DNAs integration by the means of HDR (figure 9C). These results show that the transformation of repair DNAs increased the number of cells rescued from single DSB but not for multiple DSBs in wildtype *K. marxianus*. Nonetheless, triple gene disruption was obtained in cells transformed with repair DNAs and HDR is the dominant repair mechanism in the experimental conditions.

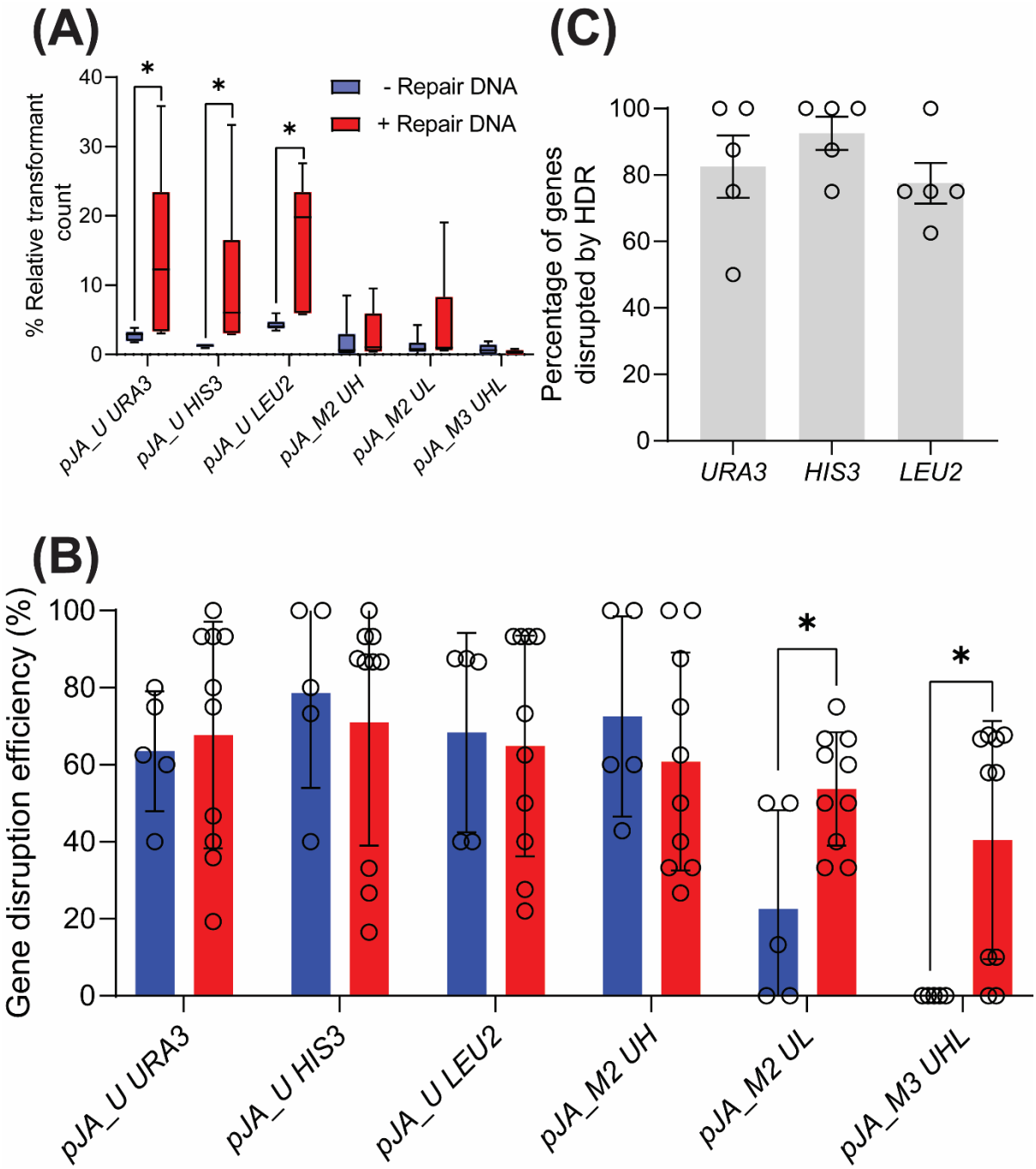


Figure 9. Transformation of repair DNAs rescued *K. marxianus* from double stranded breakages and increased disruption efficiency. (A) Wildtype *K. marxianus* was transformed with CRISPR plasmids targeting *URA3* (U), *HIS3* (H) and *LEU2* (L) with or without their respective repair DNAs and the number of transformants obtained were expressed as a percentage of transformants obtained in transformations with CRISPR plasmids that have no targets (% Relative transformant count). (B) The number of transformants with gene disruptions based on the inability to grow when patched onto synthetic media without uracil, histidine, or leucine (auxotrophy) was expressed as a percentage of total number of transformants tested (gene disruption efficiency (%)). (C) The proportion of genes disrupted by repair DNA integration (i.e DSB repair by homology directed repair (HDR)) was determined in transformants that are auxotrophic for the 3 genes in pJA_M3 *UHL* by PCR. Data shown are the mean ± standard deviation of at least five independent experiments. Statistically significant differences (examined by independent ttest) are denoted by an asterisk (*, $p<0.05$).

2.3. Disruption of *YKU80* increased the efficiency of triple gene disruption

Based on results that HDR is the dominant mechanism for DSB repair when wildtype *K. marxianus* was transformed with repair DNAs, it may be possible to improve gene disruption efficiency achieved through repair DNA integration by inactivating the yeast's non-homologous end joining machinery. The inactivation of key genes involved in NHEJ has been shown to prevent *K. marxianus* from using NHEJ for DSB repairs, thereby forcing the yeast to use HDR and increasing disruption efficiency (Rajkumar *et al.*, 2019). Given that our results so far show that gene disruption by repair DNA integration is possible in wildtype *K.* marxianus, *YKU80*, as one of the key NHEJ genes, was disrupted in wildtype *K. marxianus* by transforming CRISPR plasmid pUCC001-*KmYKU80* and its 190bp repair DNA. To determine whether *yku80* (KmJA.112) yields higher disruption efficiency than wildtype, both strains were transformed with CRISPR plasmids and repair DNAs. Disruption efficiencies were not significantly different between both strains for single and double gene targeting (figure 10). As for triple gene targeting, however, the efficiency was higher with KmJA.112 than wildtype, showing that the triple gene disruption by HDR is better established with *yku80* background.

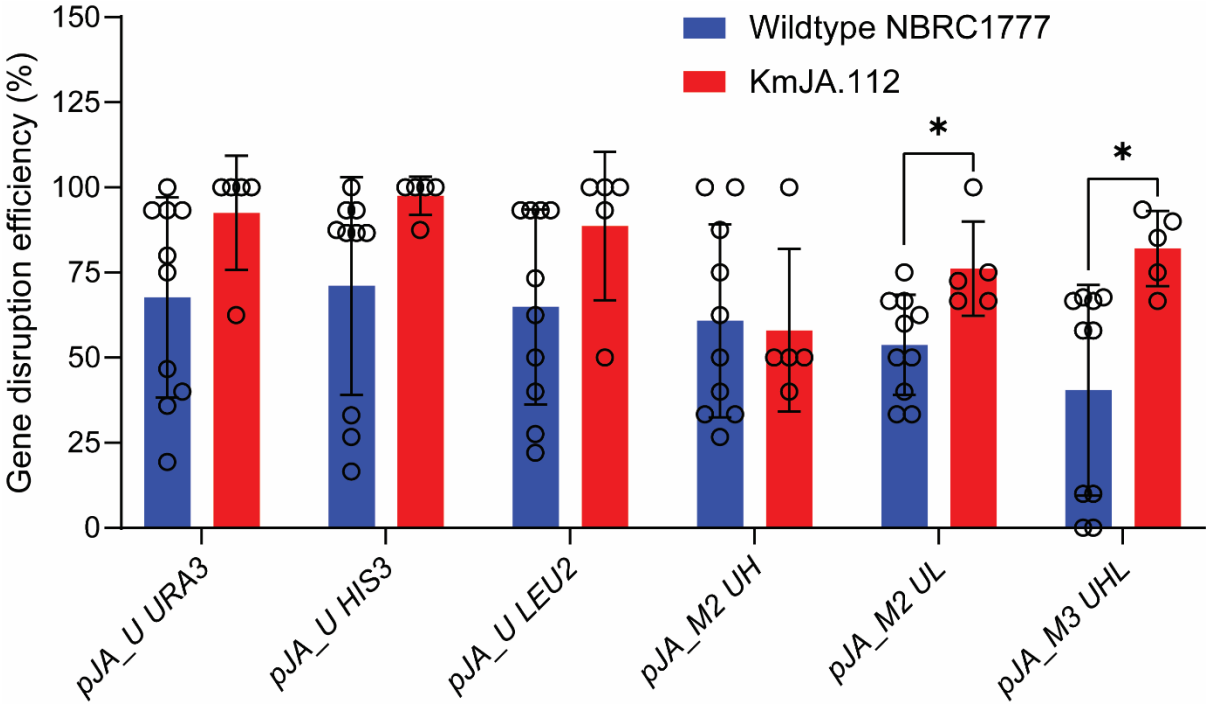


Figure 10. *YKU80* disruption increased the efficiency of triple gene disruption. Wildtype and KmJA.112 (*yku80*) strains of *K. marxianus* were transformed with CRISPR plasmids targeting *URA3* (U), *HIS3* (H) and *LEU2* (L), and their respective repair DNAs. The number of transformants with gene disruptions based on the inability to grow when patched onto synthetic media without uracil, histidine, or leucine (auxotrophy) was expressed as a percentage of total number of transformants tested (gene disruption efficiency). Data shown are the mean ± standard deviation of at least five independent experiments. Statistically significant differences (examined by independent ttest) are denoted by an asterisk (*, $p<0.05$).

2.4. Chemical synchronisation increased the number of transformants rescued form DSB by HDR

Based on our results so far that multiplex targeting drastically reduced the number of transformants obtained presumably because only a few yeast cells are able to repair the multiple DSBs generated, we wished to increase this number by using hydroxyurea to chemically synchronisation the cells into the S/G2 cell cycle phase. Firstly, we optimised the chemical synchronisation procedure by comparing KmJA.112 cultures treated with 100mM hydroxyurea versus untreated. Cultivation in the two conditions showed that hydroxyurea inhibited the growth of the strain (Figure 11A). The percentage of cells in the large-budded stage was higher in treated cultures than in untreated cultures at the second, fourth and at the twenty first hour after supplementation with hydroxyurea (Figure 11B and C), indicating that hydroxyurea treatment arrested the cells at S/G2 phase (Rosebrock, 2017). Next, we transformed untreated versus treated KmJA.112 cells with pJA_M3 *UHL* and repair DNAs. Hydroxyurea supplementation was undertaken after 3 hours preculturing and further cultivated for 2 hours. As shown in Table 5, a disruption efficiency of 100% was obtained with treated cells compared to between 0 and 50% with untreated cells for triple gene disruption. However, treatment did not significantly increase the number of transformants obtained. We speculated that the treatment protocol could be the reason for this as the cultures were already at late exponential growth phase before transformation. Preculturing time was therefore reduced to 1 hour after which hydroxyurea was supplemented and further cultivated for 2 hours. This significantly increased the number of transformants obtained with disruption efficiency between 63 and 100% (Table 5).

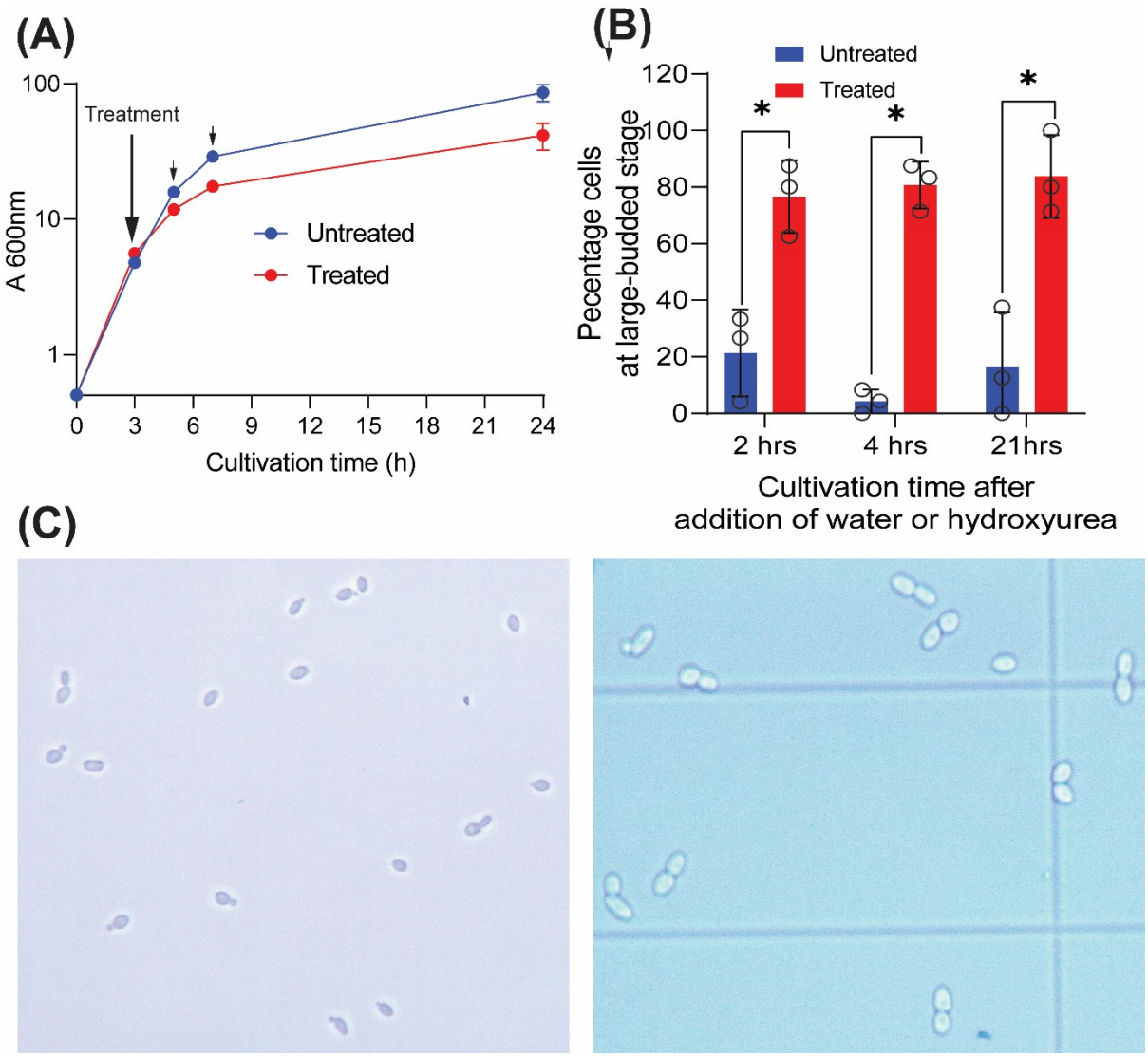


Figure 11. Optimisation of chemical (hydroxyurea) synchronisation for arresting *K. marxianus* at S/G2 cell cycle phase (large-budded stage). KmJA.112 (*yku80*) was precultured overnight in YPD and subsequently reinoculated in fresh YPD at A 600nm 0.5. (A) Culture growth profiles were monitored for 24 hours during which cultures were supplemented with water (untreated) or 100mM hydroxyurea (treated) 3 hours into cultivation. (B) the number of large-budded cells in the culture was determined as a percentage of the total number of cells counted in both cultures. (C) A representative picture of untreated (left panel) and treated (right right) cultures as observed under a light microscope. Data shown are the mean ± standard deviation of three independent experiments. Statistically significant differences (examined by independent ttest) are denoted by an asterisk (*, $p<0.05$).

Table 5. Chemical synchronisation increased the number of transformants rescued during triple deletion of *URA3, HIS3* and *LEU2* in KmJA.112 (*yku80*). The strain was precultured overnight in YPD and subsequently reinoculated in fresh YPD at A 600nm 0.5 for further cultivation. Cultures were supplemented with water (untreated) or hydroxyurea (treated) after 1 or 3 hours of cultivation and then further cultivated for 2 hours. After a total of 3 or 5 hours cultivation in fresh YPD, the cells were transformed with pJA_M3 *UHL* and repair DNAs. The number of transformants with gene disruptions based on the inability to grow when patched onto synthetic media without uracil, histidine, or leucine (auxotrophy) was expressed as a percentage of total number of transformants tested (%*ura3*, %*his3*, %*leu2*, %*ura3his3leu2*). Data shown are from three independent experiments.

3 hours preculturing before hydroxyurea treatment

	Untreated				
	Total number of transformants	*% ura3*	*% his3*	*% leu2*	*% ura3his3leu2*
Experiment 1	2	50	50	50	0
Experiment 2	2	100	50	100	50
Experiment 3	2	50	50	50	50

	Treated				
	Total number of transformants	*% ura3*	*% his3*	*% leu2*	*% ura3his3leu2*
Experiment 1	1	100	100	100	100
Experiment 2	2	100	100	100	100
Experiment 3	2	100	100	100	100

1 hour preculturing before hydroxyurea treatment

	Untreated				
	Total number of transformants	*% ura3*	*% his3*	*% leu2*	*% ura3his3leu2*
Experiment 1	4	75	75	75	75
Experiment 2	1	100	100	100	100
Experiment 3	1	100	100	100	100

	Treated

	Total number of transformants	*% ura3*	*% his3*	*% leu2*	*% ura3his3leu2*
Experiment 1	50	88	75	75	63
Experiment 2	65	100	100	100	100
Experiment 3	69	100	75	88	75

2.5. CRISPR-mediated marker-free integration of expression cassettes

Having optimised the protocol for chemically synchronising KmJA.112 for increasing the number of transformants obtained during triple gene disruption, we further tested the system for integrating expression cassettes using *Venus* which encodes a yellow fluorescent protein (YFP) as a reporter gene. Hydroxyurea-treated KmJA.112 cells were transformed without repair DNAs or with markerless YFP cassettes bearing homology arms of 750 to 1000bp for integration at *URA3, HIS3* and *LEU2* loci as illustrated for *URA3* as an example in figure 12A. No transformants were obtained in transformations without repair DNAs (figure 12B), but the transformation of the markerless YFP cassettes at concentrations of 5.5ng/µL (ca. 1fmol/µL) resulted in zero to one auxotrophic, PCR-positive, and fluorescent transformants for single and triple targeting. Doubling the concentration to 11ng/µL (ca. 2fmol/µL) significantly increased the number of transformants obtained with all of them confirmed to be auxotrophic, PCR-positive, and fluorescent.

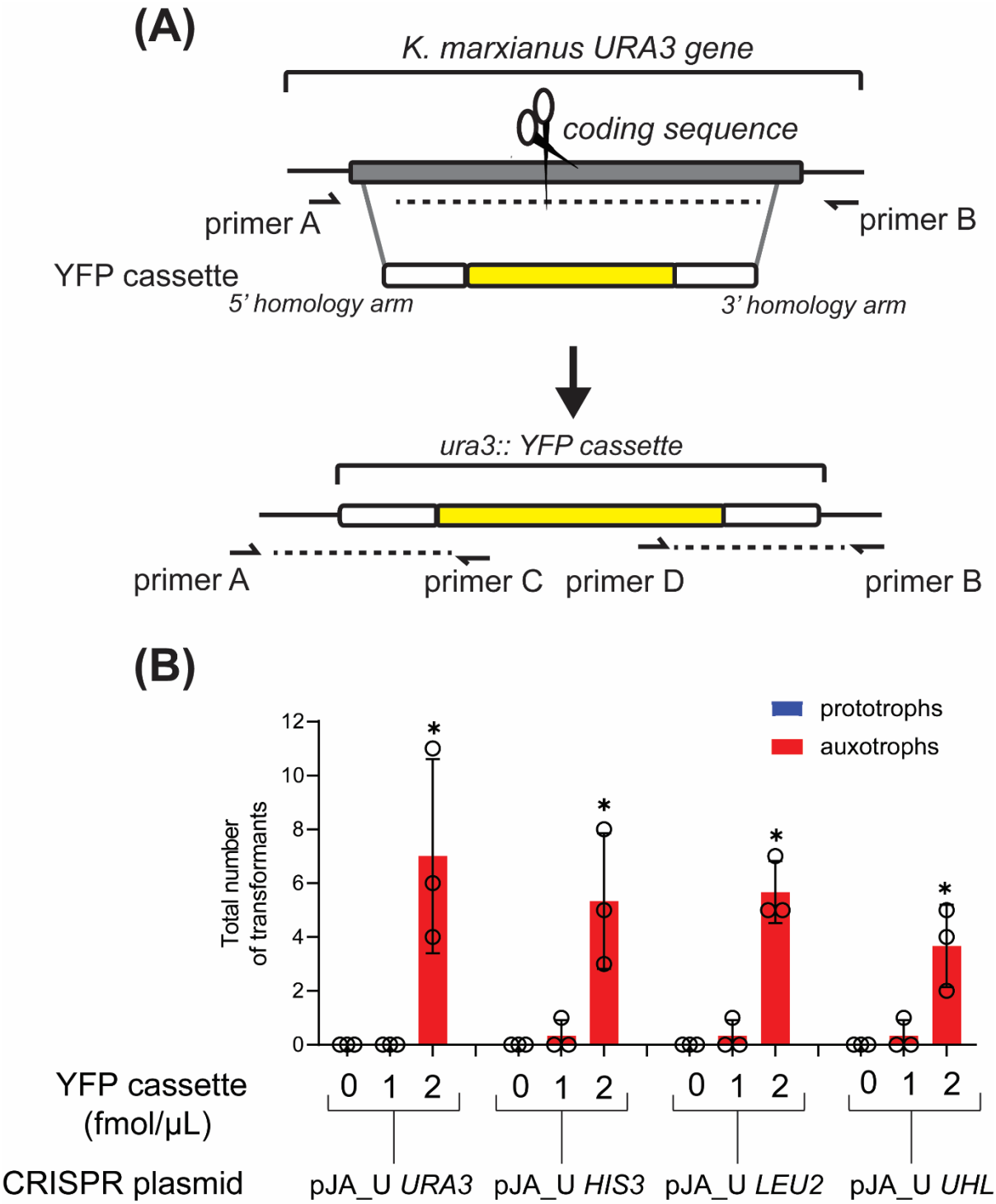


Figure 12. Markerless gene integration at *URA3, HIS3* and *LEU2* loci in *K. marxianus* genome. (A) Marker-free YFP cassettes with 750-1000bp 5' and 3' homology arms for integration at the 3 loci were constructed for achieving genomic integration as illustrated for *URA3* as an example. Successful integration can be detected by PCR using primer pairs A/C and B/D (B) Chemically synchronised KmJA.112 (*yku80*) cells were transformed with CRISPR plasmids targeting *URA3* (U), *HIS3* (H) and *LEU2* (L) without or with their YFP cassettes for the targeted loci at the indicated concentrations. Transformants were tested for auxotrophy by patching onto synthetic media without uracil, histidine, or leucine. All auxotrophic transformants were confirmed to be PCR-positive and fluorescent. Data shown are the mean ± standard deviation of three independent experiments. Statistically significant differences (examined by independent ttest) are denoted by an asterisk (*, $p<0.05$).

2.6. Delta site integration in *K. marxianus*

We wished to extend the CRISPR-mediated integration of expression cassettes to multi-loci cassette integration at repeat sequences distributed across the genome. Delta sites are suitable for this purpose and delta site integration is widely used for multi-copy integration in *S. cerevisiae*. To determine the sequences at these sites, *pol* and *gag* proteins which have their coding sequences flanked by the sites were first located by homology search with the *S. cerevisiae* orthologues of these proteins in *K. marxianus* NBRC1777 genome. Five hits were found on chromosomes I, IV, VI, VII and VIII. An assessment of the flanks showed that the sequences of 5' and 3' delta sites are identical at all hits except for the substitution of adenine 119 for guanine at the 5' delta site on chromosome VII. The delta sites are a 290bp consensus sequence found at the 5' and 3' flanks of the five hits. A Blast search of this sequence against the *K. marxianus* genome revealed that, in addition to the 10 loci, other loci with low sequence identities are also present in the genome and are potential integration sites. Based on the consensus sequence CRISPR plasmids pJA_M4 delta and pJA_M5 delta harbouring four and five identical copies of a functional gRNA sequence, respectively, were constructed for DSB generation at these sites. YFP cassette with homology arms were also designed based on the consensus sequence and constructed for DSB repair by cassette integration through HDR as illustrated in figure 13A. The plasmids were individually used to transform KmJA.112 with and without 11ng/µL (ca. 2fmol/µL) of the YFP cassette. Transformants were obtained with pJA_M4 delta but not pJA_M5 delta and 85% of transformants from the pJA_M4 delta and YFP cassette transformation displayed fluorescence (figure 13B).

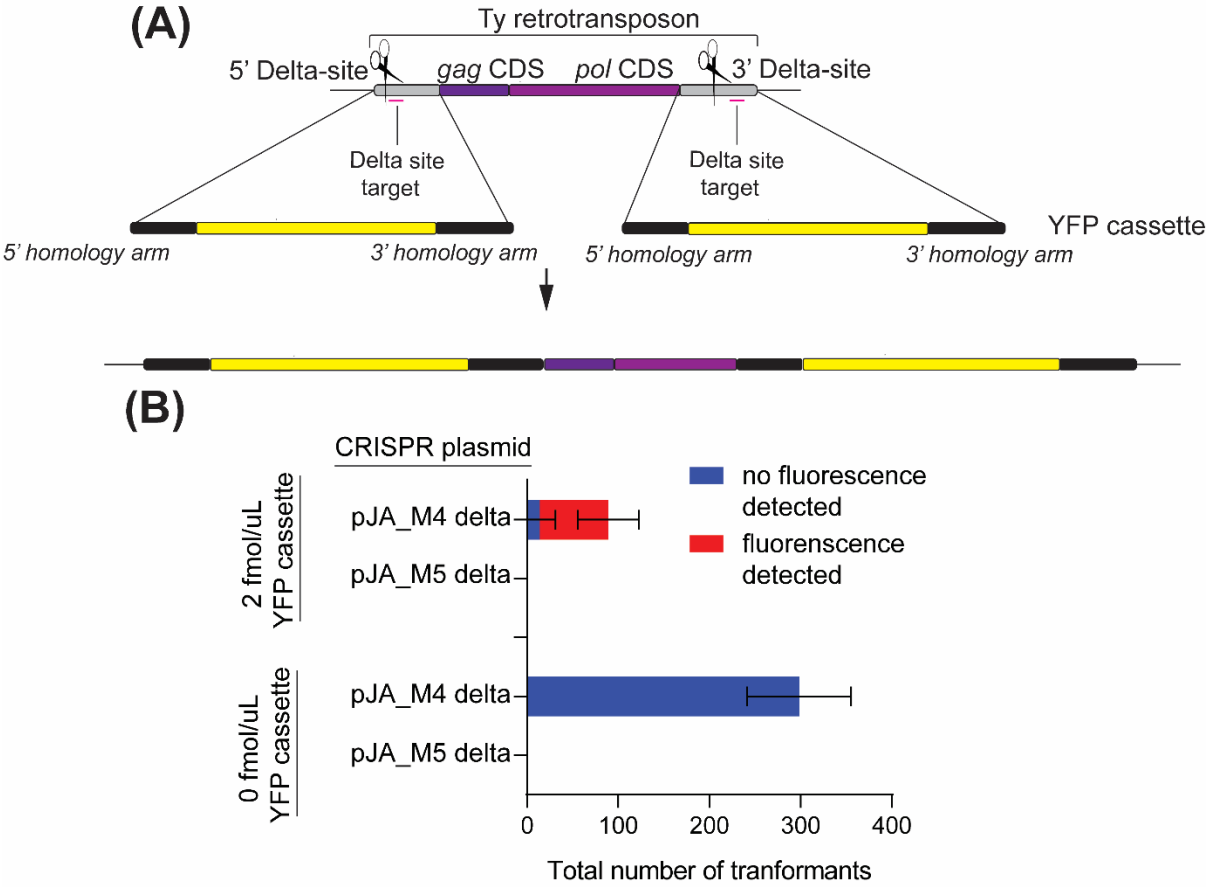


Figure 13. Delta site integration for *K. marxianus.* (A) A CRISPR plasmid is used for DSB generation at a target located within delta site sequences that are distributed across the genome. By transforming the plasmid with the illustrated YFP cassette as repair DNA, multi-loci integration of the cassette will be achieved at delta sites. The YFP cassette has flanks that are homologous to sequences outside the DSB site. (B) KmJA.112 (*yku80*) was transformed with plasmids pJA_M4 delta and pJA_M5 delta bearing 4 and 5 identical copies of a functional gRNA target, respectively. The transformation was done with or without the YFP cassette. The number of transformants with and without successful cassette integration based on fluorescence detection is reported. Data shown are the mean ± standard deviation of three independent experiments.

2.7. Optimisation of delta site integration

Chemical synchronisation by hydroxyurea treatment and heterologous expression of recombinases were employed for optimising CRISPR-mediated delta site integration. To test whether chemical synchronisation can increase the proportion of transformants that have the cassette integrated (integration efficiency), hydroxyurea treated and untreated KmJA.112 cells were transformed with pJA_M4 delta with or without YFP cassette or with the empty CRISPR plasmid (pJA_M4 *gRNAempty*). Hydroxyurea treatment significantly reduced the number of transformants obtained with the empty plasmid, but when delta sites were targeted with plasmid pJA_M4 delta in the absence of the cassette, similar number of transformants were obtained from both treated and untreated cultures (figure 14A). By adding 11ng/µL (ca. 2fmol/µL) of the cassette to the transformation, lower number of transformants was obtained

with treated versus untreated cultures and no difference in integration efficiency was found between the two conditions (figure 14A) nor was there any clear increase in the distribution of fluorescence intensities of transformants from synchronised cells (figure 7B and figure S1).

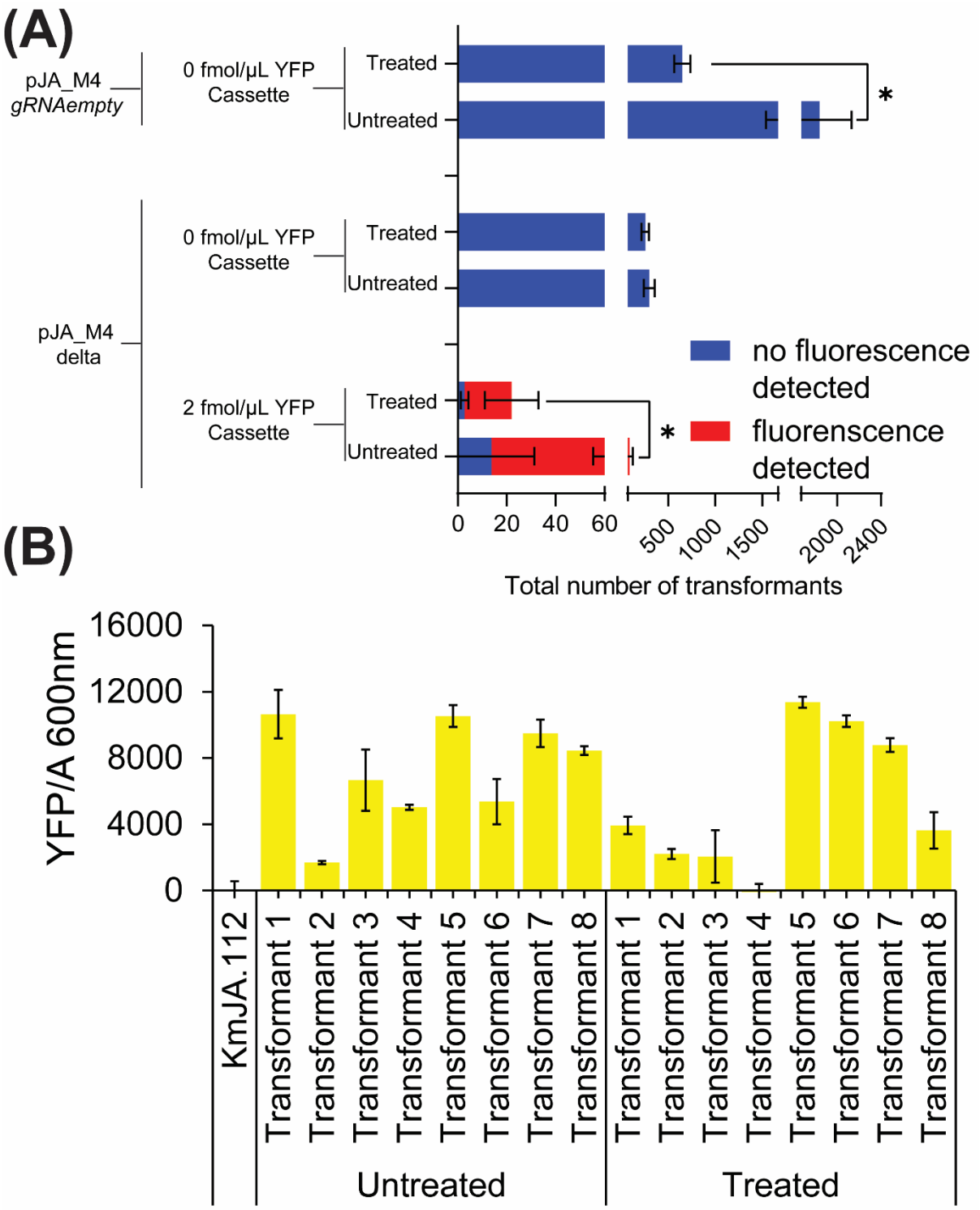


Figure 14. Effect of chemical synchronisation on delta site integration. (A) Cells treated and untreated with hydroxyurea were transformed with pJA_M4 *gRNAempty* or pJA_M4 delta with or without 2fmol/µL of YFP cassette. The number of transformants with and without successful cassette integration based on fluorescence detection is reported. (B) Representative examples of fluorescence intensities (A

600nm-normalised YFP signals) in 8 random transformants obtained from transformations with untreated and treated cultures are displayed. Data shown for total number of transformants are the mean ± standard deviation of three independent experiments and A 600nm-normalised YFP signals were determined in duplicates for each transformant.

KmJA.098, KmJA.099 and KmJA.100 are KmJA.112 derivative strains with integrations of cassettes with *E. coli recA*, *S. cerevisiae RAD52* and the empty cassette at a locus downstream of *HSP104* in the genome, respectively. The strains were transformed with pJA_M4 delta and YFP cassette. Transformations were done on galactose in order to induce the expression of the recombinases under the control of *ScGAL1* promoter in the strains. Although the number of transformants obtained with *ScRAD52* expression was lower than with *EcrecA* or the empty cassette, the three strains showed no difference in the proportion of fluorescent transformants (figure 15A). However, transformants from KmJA.098 and KmJA.099 clearly showed higher distribution of fluorescence intensities that those from KmJA.100 (figure 15B and figure S2), indicating that the heterologous expression of recombinases caused higher copies of the YFP cassette to be integrated.

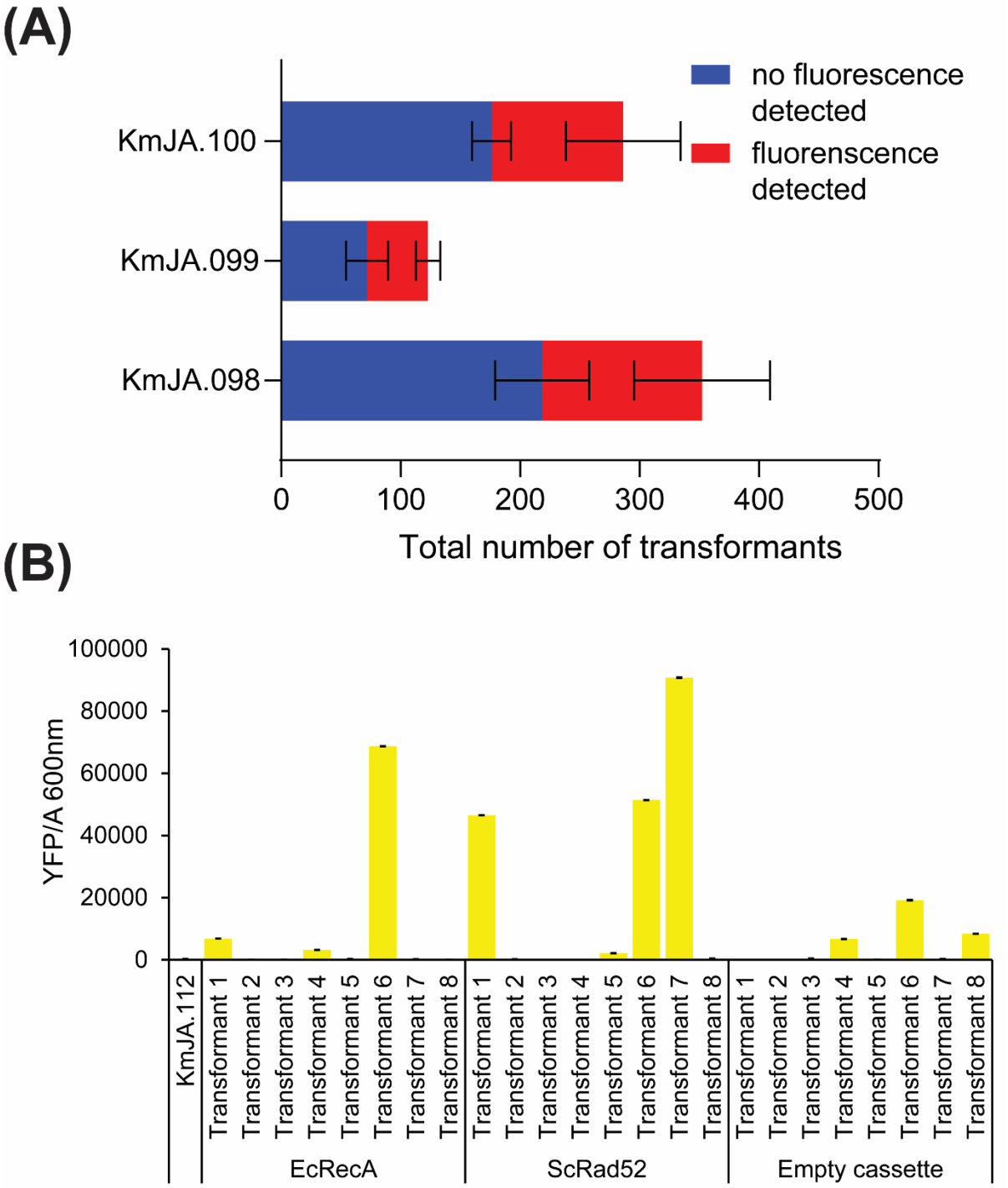


Figure 15. Expression of recombinases increased the distribution of fluorescence intensities of transformants. (A) *E. coli* recombinase gene *recA* and *S. cerevisiae RAD52* gene were integrated into the genome of KmJA.112 (*yku80*) for expression under the control of galactose inducible *ScGAL1* promoter in KmJA.098 and KmJA.099, respectively. KmJA.100 harbouring the integration of the empty cassette at the same locus was also constructed. The three strains were transformed with pJA_M4 delta and YFP cassette. The number of transformants with and without successful cassette integration based on fluorescence detection is reported. (B) Representative examples of fluorescence intensities (A 600nm-normalised YFP signals) in 8 random transformants are displayed. Data shown for total number of transformants are the mean ± standard deviation of three independent experiments and A 600nm-normalised YFP signals were determined in duplicates for each transformant.

2.8. Application of multi-loci integration for expressing chalcone synthase gene

Naringenin is a flavonoid that is produced from coumaroyl-CoA which is in turn produced from coumaric acid as illustrated in figure 16A. Chalcone synthase (Chs) is a known bottleneck enzyme in the naringenin pathway, and the reaction catalysis is famously improved intracellularly through multi-copy expression of genes encoding this enzyme (Sheng *et al.*, 2020). To demonstrate the potential of the delta site integration system presented here, we used it to integrate multiple copies of Chs gene in order to alleviate the bottleneck at this point in the pathway. To that end, we constructed a delta site integrative cassette bearing two copies of the *CHS3* gene from *Arabidopsis thaliana* (*AtCHS3*) and a single copy of *Venus*. pJA_M4 delta and the integrative cassette were used to transform a *K. marxianus* strain that has been extensively modified elsewhere to produce low yield of naringenin (see KmJA.282 later in Chapter 5 of this thesis). Of the six transformants obtained, fluorescence was detected in only transformant 2 (figure 16B), and the cultivation of this strain in minimal glucose medium produced higher yield of naringenin than KmJA.282 (figure 16C).

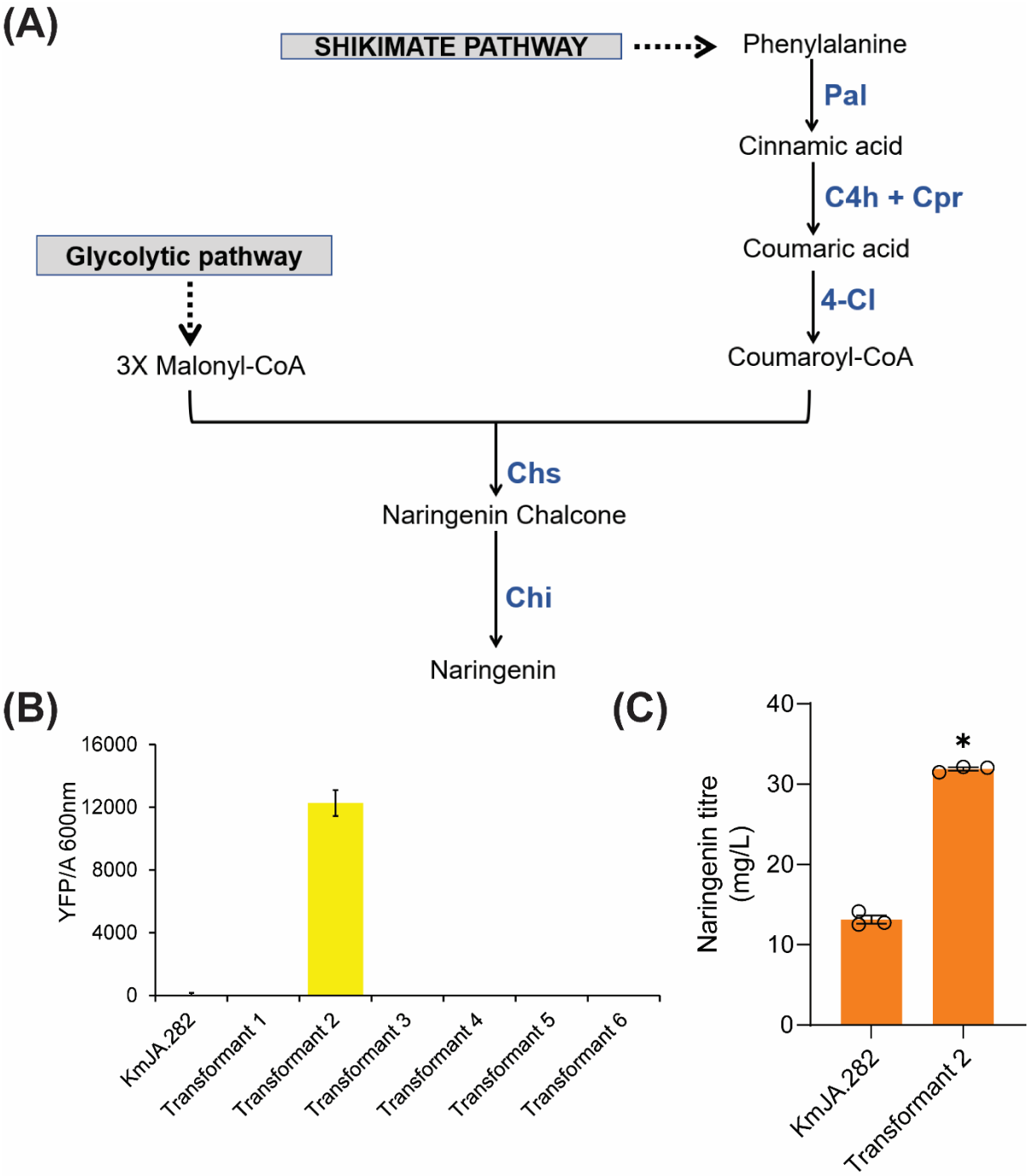


Figure 16. Delta site integration of *Arabidopsis thaliana* chalcone synthase gene. (A) Schematic illustration of the pathway for producing naringenin from phenylalanine. Phenylalanine ammonia lyase (Pal) deaminates phenylalanine into cinnamic acid which is further converted into coumaric acid by cinnamic acid 4-hydroxylase (C4h) with the help of cytochrome P450 reductase (Cpr). Coumaric acid is then ligated with coenzyme A to form coumaroyl-CoA by coumaric acid:CoA ligase (4-Cl). Coumaroyl-CoA is then used by chalcone synthase (Chs) in addition to 3 malonyl-CoA molecules to generate naringenin chalcone. This step is the bottleneck reaction of the pathway and is often improved by multi copy expression of Chs. Chalcone isomerase then isomerises naringenin chalcone into the final naringenin molecule. (B) A *K. marxianus* strain that has been extensively modified to produce naringenin by expressing a *PAL* gene from *Photorhabdus luminescens* and the other genes of the pathway from *A. thaliana* (KmJA.282) was transformed with pJA_M4 delta and a delta site integrative plasmid harbouring 2 copies of *CHS3* gene from *A. thaliana* and a copy of *Venus*. Fluorescence was determined for the six transformants obtained. (C) Transformant 2 from (B) and KmJA.282 were cultivated in minimal glucose medium and their naringenin production titres were determined. Data

shown for naringenin are the mean ± standard deviation of three independent experiments and A 600nm-normalised YFP signals were determined in duplicates for each transformant.

Discussion

K. marxianus yeasts have in recent years become genetically tractable thanks to work by various groups (Juergens *et al.*, 2018; Rajkumar *et al.*, 2019; Bever, Wheeldon and Da Silva, 2022; Rajkumar and Morrissey, 2022). This is a major step forward for their SynBio applications because it fosters the DBTL strategy for strain development at the build stage. In this study, we set out to further increase the SynBio tools available for these yeasts by developing *A. nidulans amdS* gene as selection marker for genomic integration on acetamide. Due to the ability of *K. marxianus* to naturally grow on acetamide, we first investigated the presence of acetamidase genes in the yeast. Four different genes in *K. marxianus* whose protein products potentially display acetamidase activities were identified. Bioinformatic analysis revealed that two of them, KMAR_20785 and KMAR_40009 are close to protein hits from fungi that poorly metabolise acetamide or those completely unable to metabolise this substrate, indicating that these proteins do not have acetamidase activities (figure 2). On the other hand, KMAR_50306 and KMAR_80308 are close to proteins from good/poor metabolisers of acetamide, suggesting that they have acetamidase activities. The predictions were confirmed by deleting the genes encoding the four proteins in wildtype *K. marxianus* in which double deletion of *KMAR_50306* and *KMAR_80308* (KmJA.272) caused complete loss of growth on acetamide (figure 3A). However, individual deletion of the genes did not cause this phenotype, but *kmar_50306* (KmJA.268) showed greater growth defect than *kmar_80308* (KmJA.269). In fact, KmJA.269 did not show any visible growth defect. Interestingly, sequence alignment showed that KMAR_20785, KMAR_40009 and the two hits from *K. lactis*, KLLA0_D12628g and KLLA0_B08800g which are orthologues of KMAR_50306 and KMAR_80308, respectively, bear conserved active site amino acids residues of the amidase signature family (figure 2B). Although the results here did not indicate that KMAR_20785 and KMAR_40009 have activity against acetamide (figure 3A), they may be active against other amides. This possibility also applies to the *K. lactis* proteins. Given that *K. lactis* poorly metabolises acetamide, the presence of those residues in its protein hits contradicts this phenotype, and it would be interesting to test whether these proteins are functional in *K. marxianus* which can be achieved by assessing their ability to revive the growth of KmJA.272 on acetamide.

Similar to *KMAR_50306* and *KMAR_80308*, the genomic integration of *A. nidulans amdS* revived the growth of KmJA.272 on acetamide (figure 3B). This gene was therefore constructed as selective marker according to YTK standard by inserting *KmPDC1pr-amdS-KmINU1ter* into pYTK001 to generate a level-I plasmid of *amdS* that can be combined with

other parts to generate vectors that have *amdS* as the yeast marker. The plasmid was used here to construct pI3-P1YT4 AMDS and pI4-P1YT4 AMDS from which cassettes for integrating single or multiple transcriptional units into *K. marxianus* genome can be derived. We have previously evaluated integrations at these loci which are between *KmSWF1* and *KmARO1* for I3 and downstream of *KmHSP104* for I4 (Rajkumar *et al.*, 2019) and showed here that integration efficiency of close to one hundred percent is obtainable at the loci with *amdS* as the marker (figure 4).

On account that *amdS* has been demonstrated to be recyclable by flanking the gene with direct repeats and counter-selecting on fluoroacetamide in *S. cerevisiae* (Solis-Escalante *et al.*, 2013). This permitted repeated usage of a cassette harbouring *amdS* as selection marker for establishing multiple genomic modifications. Due to that success in *S. cerevisiae* coupled with our observation here that *K. marxianus* strain expressing *amdS* under the control of a *KmPDC1* promoter is sensitive to fluoroacetamide (figure 5), we likewise attempted to recycle the gene by flanking the *amdS* transcriptional unit with direct repeats. However, all the isolates obtained after treatment with fluoroacetamide of up to 10 times the concentration previously used by others to counter-select for *amdS* in yeast (23g/L) maintained their ability to grow on acetamide (figure 6). Thus, our attempt to counter-select *amdS* in *K. marxianus* on fluoroacetamide was unsuccessful. Fluoroacetamide is a fluorine-containing analogue of acetamide that is metabolised into fluoroacetate and further into fluorocitrate which inhibits the TCA cycle enzyme, aconitase (Matsumura and O'Brien, 1963). We discovered here that while fluoroacetamide inhibited the growth of *K. marxianus* integrant of *amdS* which is an indication of fluoroacetamide toxicity, the integrant appears to tolerate the level of toxicity as shown by the persistence of acetamidase activity after fluoroacetamide treatment. Interestingly, this observation is contrary to the effect reported for *S. cerevisiae* where fluoroacetamide was successfully used to counter-select for *amdS* (Solis-Escalante *et al.*, 2013), representing a major limitation in using *amdS* as marker in *K. marxianus*. Indeed, *K. marxianus* expressing *amdS* also appeared to be less sensitive to fluoroacetamide than BY4741 in our work here (figure 5). It is possible that the ability of *K. marxianus* to take up fluoroacetamide is limited compared to *S. cerevisiae.*

With the integration construct used to achieve *amdS* recycling in *S. cerevisiae* similar to that used here in regard to the presence of direct repeats, the superior homologous recombination system of this yeast compared to *K. marxianus* (Löbs, Schwartz and Wheeldon, 2017) could contribute to the successful recycling reported. Although recycling of *amdS* in *K. lactis* by direct repeats was suggested (Ooyen *et al.*, 2006), this is yet to be substantiated. Based on the evolutionary closeness of *K. lactis* and *K. marxianus*, *K. lactis* could also have low sensitivity to fluoroacetamide, which will, like in *K. marxianus,* hamper the counter-selection of *amdS* in

K. lactis. While the *Cre-loxP* system is an alternative marker recycling system that is versatile in its potential to recycle any marker gene, repeated usage makes strains prone to chromosomal rearrangements (Burgess and Kleckner, 1999; Delneri *et al.*, 2000; Solis-Escalante *et al.*, 2015). The advantage of this system is however achievable by placing an inducible lethal gene beside the selectable marker of choice such as *amdS*. The marker and lethal gene can be excised by flanking the entire sequence with direct repeats. Such approach has been used for efficient seamless gene deletions in *S. cerevisiae* with the aid of galactose inducible cell wall protein 1 (Cwp1) gene (Hu *et al.*, 2019) and may be similarly adaptable to *K. marxianus* and other non-saccharomyces yeasts. However, the issue of poorer homologous recombination of these yeasts than in *S. cerevisiae* remains a challenge in that approach.

To further address genomic integration in *K. marxianus,* we sought to develop a CRISPR-mediated marker-free integration approach. Efficient genome editing by CRISPR/Cas9 was previously developed for *K. marxianus* based on pUCC001 (Juergens *et al.*, 2018; Rajkumar and Morrissey, 2022). The copy number of this plasmid in hosts is governed by a pangenomic replication origin, giving it the advantage of being versatile for application in several yeast species. This may, however, be suboptimal for multiplex targeting. The expression of Cas9 and guide RNA transcription are two highly important factors that determine CRISPR targeting efficiency. Cas9 was reported to have low turnover and remain tightly bound to double-stranded DNA after generating DSB (Sternberg *et al.*, 2014). Consequently, a molecule of the Cas9 protein is not readily available for DSB induction at another locus after induction at a locus. This is important when more than one locus is targeted simultaneously. To ensure that multi-loci targeting is not limited by an inadequacy of Cas9 molecules and the gRNA molecules targeting them to specific loci, we used high copy number CRISPR plasmids, achieved by the *ARS2* replication origin of *K. marxianus* (Rajkumar *et al.*, 2019) in this work. We demonstrated that very low number of transformants is obtained when multiple loci are simultaneously targeted for DSB generation in wildtype *K. marxianus*. Effectively, the number of transformants reduced when the number of gene targets is increased as shown by the fact that % relative transformant count of triple targeting was considerably lower than those of single targeting (figure 9A). We hypothesised that this is due to the difficulty in repairing multiple DSBs compared to single DSB. Although others have proposed the replacement of Cas9 with other endonucleases due to certain advantages that they possess over Cas9 such as Cas12a, which does not rely on host machinery for gRNA processing (Swarts and Jinek, 2018; Paul and Montoya, 2020), and MAD7, a royalty-free alternative to Cas9 (Ouedraogo and Tsang, 2020), the success of these other endonucleases for multiplex genome editing, like Cas9,

requires the host to have in place an efficient system for DSB repair by HDR, which is a challenge in many non-saccharomyces yeasts (Löbs, Schwartz and Wheeldon, 2017).

To improve DSB repair here, short repair DNA of 190bp were provided during transformation, making HDR available for repair in addition to NHEJ in wildtype *K. marxianus*. This strategy increased the % relative transformant counts of single targeting but not of multi-loci targeting (figure 9A). Nonetheless, simultaneous disruption of three genes was only successful with this approach and PCR showed that majority of the disruptions were achieved by HDR (figure 9B and C). These results motivated us to focus on improving DSB repair by HDR first by comparing gene disruption efficiencies in a NHEJ-deficient strain to wildtype based on previous reports that repair by HDR is improved in such strains of *K. marxianus* (Nambu-Nishida *et al.*, 2017; Rajkumar *et al.*, 2019; Bever, Wheeldon and Da Silva, 2022). Although higher efficiency was indeed achieved with this strain than with wildtype for triple targeting (figure 10), the number of transformants obtained after transformation remained low. This was increased from a mean of 2 to 61 transformants by chemical synchronisation of cells into S/G2 phase before transformation. While attempts to integrate YFP cassettes at *URA3*, *HIS3* and *LEU2* loci using 95bp or 750 to 1000bp homology flanks was unsuccessful in our hand because we obtained no transformants (not shown), our chemical synchronisation method yielded transformants, albeit at low counts, with 100% integration efficiencies for single and triple integrations (figure 5B).

Multi-loci integration at delta sites distributed across the genome has been reported in *S. cerevsiae* (Shi *et al.*, 2016). This was achieved by targeting DSB generation to the sites and then repairing by HDR to integrate cassettes that bear short flanks that are homologous to sequences immediately adjacent to the DSB. We devised a similar strategy for *K. marxianus* here by testing CRISPR plasmids with four versus five identical copies of a functional delta site targeting gRNA sequence for DSB generation. YFP cassette integration efficiency at delta sites of 85% was achieved with the four copies plasmid while the five copies plasmid yielded no transformants (figure 6). Based on reports that the heterologous expression of recombinases enhanced HDR in mammalian cells (Paulsen *et al.*, 2017; Shao *et al.*, 2017), the fusion of Cas9 to recombination factors enhanced Cas9-mediated gene editing (Tran *et al.*, 2019) and the extensive engineering of the homologous recombination machinery of the non-saccharomyces yeast, *Pichia pastoris*, improved HDR and enhanced multi-loci editing (Cai *et al.*, 2021), we likewise presented here data suggesting that galactose-induced expression of *EcrecA* and *ScRAD52* increased the copy number of YFP cassettes integrated at delta sites (figure 15B).

As proof-of-concept, we used the multi-loci integration system to improve naringenin biosynthesis in a *K. marxianus* strain that overexpresses the native genes encoding enzymes

involved in the shikimate pathway in addition to expressing genes that encode the enzymes of the naringenin pathway. The application of our delta site integration method without enhancing HDR resulted in 16.6% integration efficiency of a *AtCHS3-Venus-AtCHS3* integrative cassette (figure 16B), which is lower than the efficiency achieved in the transformation of KmJA.112 with the YFP-cassette. This low efficiency can be attributed to the larger size of the *AtCHS3-Venus-AtCHS3* cassette compared to the nearly 100% efficiency when smaller cassette of a single copy of *Venus* (YFP-cassette) was integrated. A similar result was reported for delta site integration of *S. cerevisiae* (Shi *et al.*, 2016). In addition, the KmJA.282 background used here is an extensively engineered strain in contrast to KmJA.112. This could contribute to the low integration efficiency, but it may be possible to increase the efficiency by the chemical synchronisation method presented in this work. Nonetheless, we demonstrated that successful delta site integration of *AtCHS3* doubled naringenin titre (figure 16C).

Conclusion

In conclusion, acetamidase activity in wildtype *K. marxianus* is contributed by two of four putative acetamidase genes and the yeast can be rendered unable to grow on acetamide by disrupting these genes. Thus, such a strain is useful for SynBio applications with amidase gene as the selection marker. To that end, we provide in this study the *A. nidulans amdS* gene as YTK level-I selection marker plasmid for constructing yeast vectors that can be selected on acetamide. We discovered that *K. marxianus* integrant of *amdS* grows in the presence of fluoroacetamide, albeit with moderate sensitivity, while maintaining acetamidase activity. This resulted in our failure to successfully counter-select for *amdS* in *K. marxianus*. The CRISPR tools developed in this work can drive the development of *K. marxianus* for SynBio applications in four aspects: (i) they are based on the extensively established YTK standards, meaning that it can be combined with other YTK parts. (ii) they are marker-free and eliminate the need for selectable markers. Furthermore, markerless integrations at the three genes targeted in this study results in uracil, histidine and leucine auxotrophy, creating an opportunity for using *URA3, HIS3* and *LEU2* for subsequent selections as required. (iii) the time required for strain construction which often requires the DBTL approach can be reduced with our multiplexing tool. The fact that the integrative cassettes presented here can be used for expressing multiple genes of biosynthetic pathways contributes to this advantage. (iv) a widely used strategy in metabolic engineering is the provision of multiple copies of biosynthetic genes. This can benefit from the delta site integration tool developed here as we demonstrated with the integration of *AtCHS3* for naringenin production.

Supplementary materials

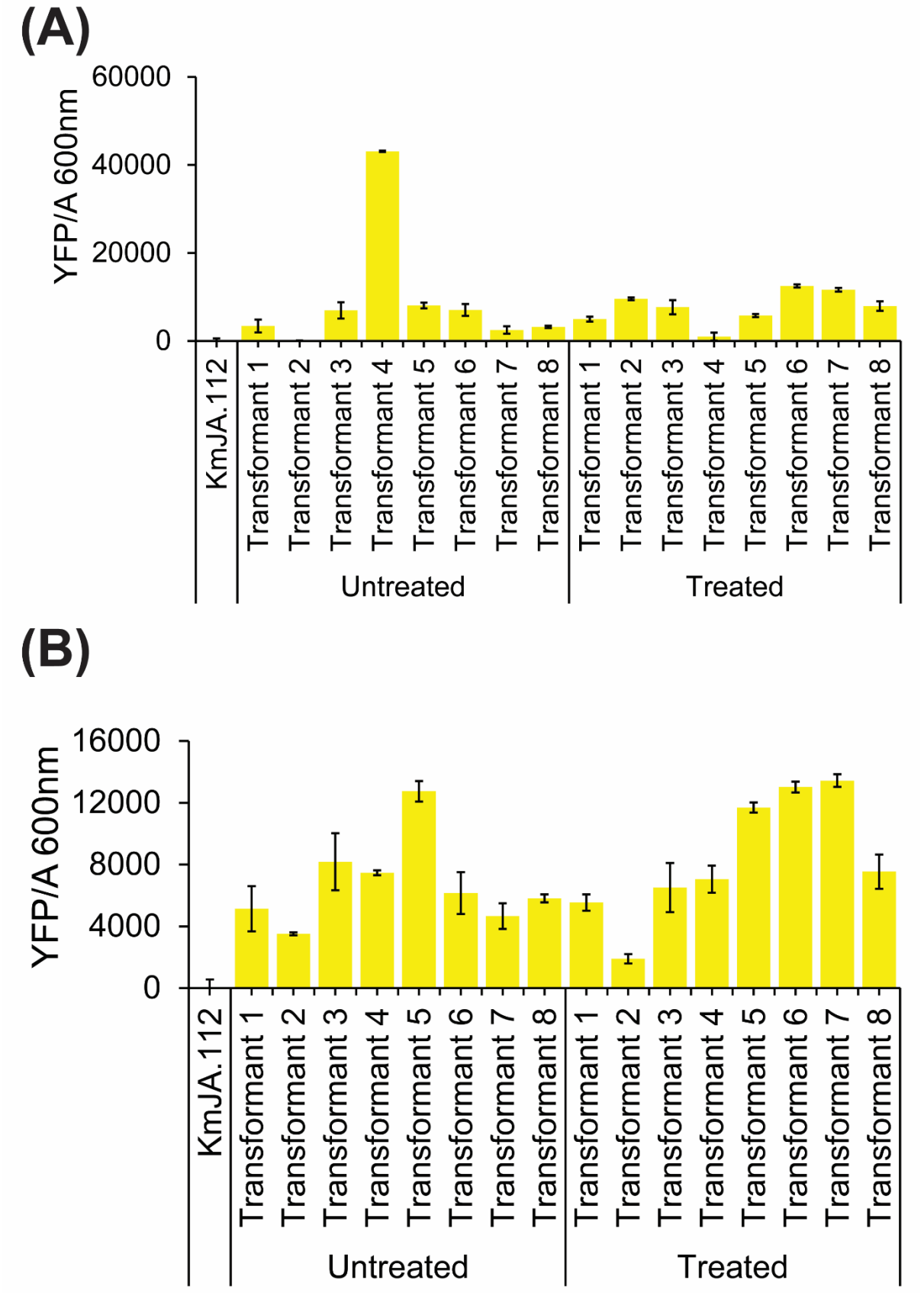


Figure S1. Second (A) and third (B) attempts at testing the effect of chemical synchronisation on delta site integration. Cultures treated and untreated with hydroxyurea were used for co-transforming pJA_M4 delta with or without 2 fmol/µL of YFP cassette. A 600nm-normalised YFP signals of 8 colonies obtained from the transformations are displayed. Data shown are A 600nm-normalised YFP signals determined in duplicates for each colony.

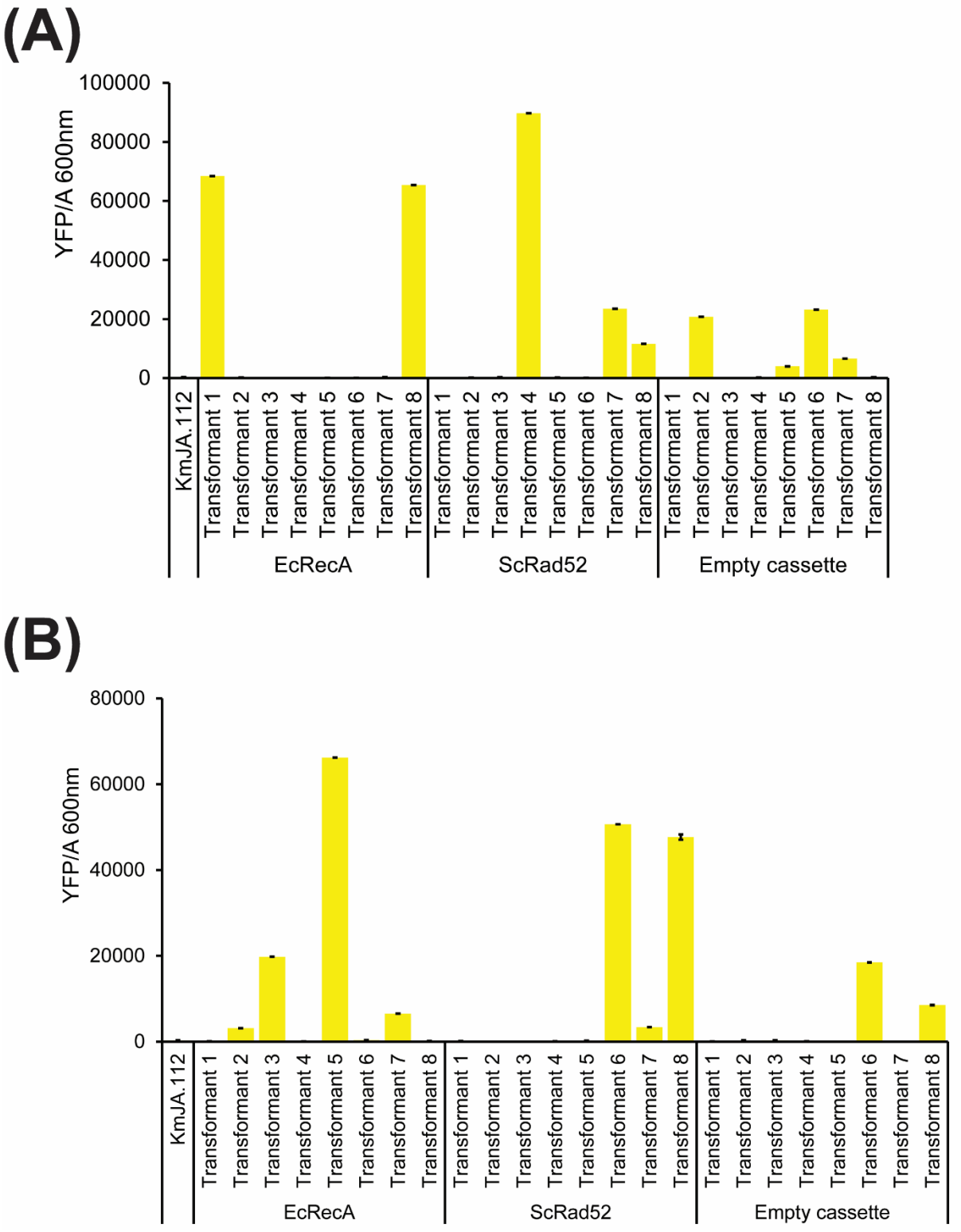


Figure S2. Second (A) and third (B) attempts at testing the effect of recombinase expression on delta site integration. *E. coli* recombinase gene *recA* or *S. cerevisiae RAD52* genes were integrated into the genome of KmJA.112 (*yku80*) for expression under the control of galactose inducible *ScGAL1* promoter. The two strains were the co-transformed with pJA_M4 delta and YFP cassettes. A strain in which an empty cassette was integrated in the same locus was also transformed as control. A 600nm-normalised YFP signals of 8 colonies obtained from the transformations are displayed. Data shown are A 600nm-normalised YFP signals determined in duplicates for each colony.

Chapter 3. Optimisation of coumaric acid production from phenylpyruvate in *Kluyveromyces marxianus*

Joel A. Akinola, Arun S. Rajkumar, John P. Morrissey.

This chapter is in preparation for submission

JAA performed all experiments, interpreted the data and prepared the manuscript

ASR constructed the KmASR.184 platform strain and assisted in the design of strain configurations

JPM provided supervision and edited the manuscript

Abstract

Coumaric acid is an important aromatic molecule from which many other valuable aromatics that have applications in the food, pharmaceutical, nutraceutical and agrochemical industries can be biosynthesised. *Kluyveromyces marxianus* is an industrially attractive yeast because it is thermotolerant, grows fast and is able to metabolise diverse substrates, allowing the potential to reduce manufacturing cost. In this work, synthetic biology tools were used to engineer coumaric acid production in this yeast. As the biological precursors of coumaric acid, aromatic amino acids (AAAs) biosynthesis was first enhanced by overexpressing important genes in the pentose phosphate and shikimate pathways. Multiple copies of coumaric acid biosynthetic genes were then integrated into the genome of the strain with enhanced AAA biosynthesis and endogenous phenylpyruvate decarboxylase activity was attenuated in order to reduce the degradation of AAAs in the Ehrlich pathway, leading to the production of coumaric acid. To further optimise production, we explored the final step of AAA biosynthesis which requires the donation of an amino group from glutamate by overexpressing glutamate biosynthetic genes in a coumaric acid producer. This strategy resulted in 68% increase in coumaric acid yield. Due to the respiratory physiology of *K. marxianus* in that glucose can be metabolised via the TCA cycle in this yeast, we explored and found that aeration in batch bioreactor contributes to coumaric acid production. In the end, a maximum yield of 42mg/g glucose is reported here for coumaric acid production in aerated batch bioreactors. While this production is lower than has been reported in *S. cerevisiae*, our study provides *K. marxianus* platforms that can be further engineered for producing valuable aromatic compounds and brings us closer to availing of the physiological properties of this yeast for cost-effective production of aromatics.

Introduction

Extraction from plants is commonplace for commercial production of plant natural products (PNPs) that cannot be chemically synthesised due to their complex structure. However, limited natural plant sources hinders production and the physicochemical similarity between thousands of plant molecules makes downstream processing difficult and results in low yield, making supply by this means unable to meet product demand and overall commercial feasibility difficult to achieve. These reasons coupled with demand for renewable and sustainable sources of these products has resulted in on-going attempts to develop microbial cell factories for commercial production of PNPs (Liu *et al.*, 2020). Plant aromatic molecules are of high interest due to their potential applications in diverse industries where they are used

as food additives, fragrances, cosmetics, health supplements and active pharmaceutical ingredients. Phenylpropanoids are a large and structurally diverse group of aromatic plant secondary metabolites produced from phenylalanine and tyrosine in response to biotic and abiotic stressors and are crucial for plant growth and development (Fraser and Chapple, 2011). This group of molecules are so-called due to the presence of a phenyl (6 carbon) ring and three carbon aliphatic chain of coumaric acid, the central metabolite from which they are produced. The production of phenylpropanoids in plants requires initial biosynthesis of coumaric acid through non-oxidative deamination of aromatic amino acids (AAAs) by ammonia lyases and the ligation of coumaric acid with coenzyme A to produce 4-coumaroyl-CoA, which is subsequently directed towards various routes by relevant polyketide synthases to produce specific classes of phenylpropanoids (Yu *et al.*, 2019). The structural diversity of phenylpropanoids in plants arises as a result of regio-specific enzymes that add and remove moieties to and from existing molecules to form derivatives through condensation, aromatization, glycosylation, prenylation, cyclization, hydroxylation, acylation, methylation and sulfation reactions (Noel, Austin and Bomati, 2005). The structural classes of phenylpropanoids include flavonoids, monolignols, phenolic acids, stilbenes and coumarins, which include generally bioactive molecules some of which some are associated with health benefits in animals and humans, and potential nutraceutical and pharmaceutical significance (Neelam, Khatkar and Sharma, 2020).

With aromatic amino acids (AAAs) phenylalanine and tyrosine as precursors for the biosynthesis of aromatics, efforts to develop microbial cell factories for producing aromatic compounds feature increasing the intracellular abundance of AAAs (Sheng *et al.*, 2020).Yeasts represent suitable microorganisms for producing plant aromatics due to their superior functional expression, compared to in bacteria, of eukaryotic proteins such as membrane-bound cytochrome P450 oxidases required in important plant aromatic biosynthetic pathways(Liu *et al.*, 2020). While *S. cerevisiae* has been successfully engineered for aromatics production by various authors, the need for optimum precursor supply, recent advances in genetic tractability and improving understanding of their physiology are fostering the exploration of non-saccharomyces yeasts as cell factories for production(Cao *et al.*, 2020; Liu *et al.*, 2020). Accordingly, the requirement of malonyl-CoA for the biosynthesis of complex aromatics has resulted in researchers turning attention to *Yarrowia lipolytica,* an oleaginous yeast known for its superior ability to biosynthesise fatty acids which are derived from malonyl CoA. The heterologous expression of the naringenin pathway in the oleaginous yeast *Y. lipolytica* resulted in higher production of naringenin, a complex aromatic polyketide, from glucose compared to *S. cerevisiae* (Lv *et al.*, 2019; Palmer *et al.*, 2020). The potential of this yeast encouraged its development into platform strains with increased AAA precursor

biosynthesis for further engineering to produce aromatics of choice (Gu, Ma, Zhu, Ding, *et al.*, 2020). *K. marxianus* attracts interest for industrial applications due to its thermotolerance and fast growth (Bilal *et al.*, 2022). The ability of this respiro-fermentative yeast to metabolise diverse carbon sources strengthens its candidacy as microbial cell factory for a circular economy where wastes from several industries can serve as production feedstocks. This has motivated the engineering of *K. marxianus* for de novo production of 2-phenylethanol, a rose-like smelling compound (Kim, Lee and Oh, 2014), and, similar to *Y. lipolytica, K. marxianus* was developed into platform for aromatics biosynthesis(Rajkumar and Morrissey, 2020). With those works laying the foundation for producing aromatics in *K. marxianus*, we sought to engineer the yeast to produce heterologous aromatic molecules. The centrality of coumaric acid in the phenylpropanoid pathway means that its biosynthesis is a required step in the production of phenylpropanoid molecules in microbial cells factories. The deamination of phenylalanine and tyrosine is catalysed by two distinct groups of enzymes called phenylalanine ammonia lyases (Pals) and tyrosine ammonia lyases (Tals), with Tals directly producing coumaric acid by deaminating tyrosine, Pals first deaminates phenylalanine into cinnamic acid. Cinnamic acid is then hydroxylated into coumaric acid by cinnamic acid 4-hydroxylase (C4h), the cytochrome P450 enzyme that requires cytochrome P450 reductase (Cpr)-mediated electron acceptance from NADPH.

In this current work, we engineered *K. marxianus* for *de novo* production of coumaric acid as a platform for producing phenylpropanoids. Several ammonia lyase genes were tested to identify the one that allows highest coumaric acid production in a platform *K. marxianus* strain engineered for increased intracellular phenylalanine and tyrosine biosynthesis. Furthermore, flux was diverted away from 2-phenylethanol towards coumaric acid biosynthesis, and the coumaric acid yield was increased by improving the amination of phenylpyruvate into phenylalanine.

Materials and Methods

1. Strains and growth conditions

Kluyveromyces marxianus strain NBRC1777 from the Biological Resource Centre, NITE (NBRC; Tokyo Japan) was used as background in this study. All yeast strains used in this study are listed in Table 1. Yeasts were routinely cultured at 30°C in YPD broth (2% yeast extract, 1% peptone and 2% glucose) or in appropriately supplemented minimal medium (5g/L $(NH_4)_2SO_4$, 3g/L KH_2PO_4, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 15g/L glucose, and filter-sterilized vitamins and trace elements) brought to pH 5.5-6 with KOH (Verduyn *et al.*, 1992). For yeast growth on solid medium, 2% agar supplementation was applied. Nutritional reagents and agar were purchased from Formedium, Hunstanton, UK. Bacterial transformation was with *E. coli* DH5α

grown in LB broth (1% NaCl, 1% peptone, 0.5% yeast extract) or LB supplemented with 1.5% agar and appropriate antibiotic (50mg/L chloramphenicol, 100mg/L ampicillin or 50mg/L kanamycin; Fisher Scientific (Dublin, Ireland)).

Table 1. Strains used in this study.

Strain	Genotype	Source
NBRC1777	Wildtype	NITE Biological Resource Center
KmASR.039	*dln4aro4aro7ura3his3*	(Rajkumar and Morrissey, 2020)
KmASR.088	KmASR.039, *aro3Δ:: p*I6-Chor/*A2 [URA3]*	This work
KmASR.184	KmASR.088 *lac4Δ::*pI1-PPP [*HIS*]	This work
KmJA.010	KmASR.184 *leu2*	This work
KmJA.073	KmJA.010 *aro10::*pJA_L2.22 *[ScLEU2]*	This work
KmJA.074	KmJA.010 *aro10::*pJA_L2.23 *[ScLEU2]*	This work
KmJA.075	KmJA.010 *aro10::*pJA_L2.24 *[ScLEU2]*	This work
KmJA.076	KmJA.010 *aro10::*pJA_L2.28 *[ScLEU2]*	This work
KmJA.077	KmJA.010 *aro10::*pJA_L2.36 *[ScLEU2]*	This work
KmJA.007	KmASR184 *l4Δ::*pJA_L3.29 *[HphMX]*	This work
KmJA.078	KmJA.007 *aro10::*pJA_L2.22 *[ScLEU2]*	This work
KmJA.079	KmJA.007 *aro10::*pJA_L2.23 *[ScLEU2]*	This work
KmJA.080	KmJA.007 *aro10::*pJA_L2.24 *[ScLEU2]*	This work
KmJA.081	KmJA.007 *aro10::*pJA_L2.28 *[ScLEU2]*	This work
KmJA.082	KmJA.007 *aro10::*pJA_L2.36 *[ScLEU2]*	This work
KmJA.011	KmJA.010 *swf1aro1::*pJA_3.38 *[ScLEU2]*	This work
KmJA.013	KmJA.011 *aro10*	This work
KmJA.014	KmJA.011 *pdc5*	This work
KmJA.009	KmJA.010 *pdc5*	This work
KmJA.015	KmJA.009 *aro10*	
KmJA.016	KmJA.015 JA_3.39 *[ScLEU2]*	This work
KmJA.017	KmJA.015 JA_3.38 *[ScLEU2]*	This work
KmJA.091	KmJA.016 JA_L2.59 *[HphMX]*	This work
KmJA.092	KmJA.017 JA_L3.29 *[HphMX]*	This work
KmJA.088	KmJA.016 JA_L3.24 *[HphMX]*	This work
KmJA.090	KmJA.017 JA_L3.24 *[HphMX]*	This work
KmJA.119	KmJA.017 *aro9*	This work
KmJA.171	KmJA.119 L3.24 *[HphMX]*	This work
KmJA.180	KmJA.119 L3.60 *[HphMX]*	This work
KmJA.188	KmJA.119 L3.64 *[HphMX]*	This work
KmJA.189	KmJA.119 L3.65 *[HphMX]*	This work
KmJA.182	KmJA.119 L3.62 *[HphMX]*	This work
KmJA.228	KmJA.119 L3.68 *[HphMX]*	This work
KmJA.229	KmJA.119 L3.69 *[HphMX]*	This work
KmJA.249	KmJA.119 L3.76 *[HphMX]*	This work
KmJA.250	KmJA.119 L3.77 *[HphMX]*	This work

2. Plasmid construction

2.1. Construction of plasmids for gene expression

All expression plasmids used in this study were integrative and constructed by Golden Gate assembly based on the hierarchical Yeast Toolkit standard (Lee *et al.*, 2015; Rajkumar *et al.*, 2019; Rajkumar and Morrissey, 2022) and are listed in Table 2 . Briefly, genes were sourced by Q5 High-Fidelity Polymerase (M4092L, New England Biolabs (NEB), Ipswich, UK) PCR amplification using primers bearing BsaI and BsmBI sites at their 5' and 3' ends or as gene blocks (Twist Biosciences, San Francisco, U.S.A.). All non-yeast native genes expressed in this study were codon optimised for *S. cerevisiae* and are listed on Table S1. Genes was then cloned by a BsmBI (R0739L, NEB)/T7 DNA ligase (M0318L, NEB) Golden Gate reaction into pYTK001 to construct level-I gene plasmids (listed in Table S2). When a gene was to be individually expressed, the level-I plasmid of that gene was put together with level-I promoter and terminator plasmids and an integrative dropout plasmid in a BsaI (R0535L, NEB)/T7 DNA ligase Golden Gate reaction to construct a level-II plasmid of the gene. The integrative dropout plasmids used were constructed in this study or from previous work (Rajkumar *et al.*, 2019), and are listed in Table S2.

Ten level-II red florescence protein (RFP) dropout plasmids required for constructing level-III plasmids with up to six concatenated transcriptional units (TUs) were constructed in this work by a BsaI/T7 DNA ligase reaction comprising of a left connector part plasmid (one of pYTK002 to pYTK008), pYTK095 for AmpR-ColE1, a right connector part plasmid (one of pYTK067 to pYTK073) and the bacteria expressible PCR product of red florescence protein TU amplified from pYTK083 with primers JA69 and JA70. For constructing the level-III plasmids with multiple concatenated TUs, the genes to be expressed from the plasmid were first individually cloned alongside the desired yeast promoters and terminators into level-II plasmids using BsaI/ T7 DNA ligase reactions to construct a set of level-II plasmids that individually harboured transcriptional units for each gene of interest. The level-II dropout plasmid in which a TU is placed is specified by the predetermined position for that TU in the level-III plasmid to be constructed. Once a set of level-II TU plasmids were constructed and confirmed, they were assembled and inserted into appropriate dropout plasmids by BsmBI/ T7 DNA ligase reaction to generate the level-III biosynthetic plasmid. Golden Gate reaction products were transformed into *E. coli* DH5α and selected on LB supplemented with chloramphenicol for level-I plasmids, ampicillin for level-II plasmids and kanamycin for level-III plasmids. The primers used in this study are listed in Table S3.

Table 2. Expression plasmids used in this study.

Name	Description	Source
pI1-PPP	pI1-MTU-DO-HIS with insert *KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmTDH3pr-KmRPE1-mPDC1t KmTSA1pr-KmRKI1-ScPGK1t*	(Rajkumar and Morrissey, 2020)
pI6-Chor/A2-URA	pI6-MTU-DO-URA with insert *KmPDC1pr-KmARO4^{fbr}-ScADH1t KmENO1pr-ARO1-KMXK_A03020t KmPGK1pr-ARO2-KmLAC4t KmTEF1pr-KmARO7^{fbr}-KmPDC1t*	This work
pJA_NM L2. 22	pI7/*ARO10* MTU-DO-LEU with insert *KmPDC1-AtPAL1-ScPGK1t*	This work
pJA_NM L2. 23	pI7/*ARO10* MTU-DO-LEU with insert *KmPDC1-RgPTAL1-ScPGK1t*	This work
pJA_NM L2. 24	pI7/*ARO10* MTU-DO-LEU with insert *KmPDC1-PlPAL-ScPGK1t*	This work
pJA_NM L2. 28	pI7/*ARO10* MTU-DO-LEU with insert *KmPDC1-SmPAL-ScPGK1t*	This work
pJA_NM L2. 36	pI7/*ARO10* MTU-DO-LEU with insert *KmPDC1-RcTAL-ScPGK1t*	This work
pJA_NM L3.29	pI4 MTU-DO-HPH with insert *KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work
pJA_NM 3.38	pI3 MTU-DO-LEU with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work
pJA_NM 3.39	pI3 MTU-DO-LEU with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t*	This work
pJA_NM L3.24	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work

pJA_NM L2.59	pI4 MTU-DO-HPH with insert *KmENO1pr-AtC4H-KmPGK1t*	This work
pJA_NM L3.60	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmPDC1pr-KmGDH1-ScPGK1t*	This work
pJA_NM L3.64	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmPDC1pr-ScGDH1-ScPGK1t*	This work
pJA_NM L3.65	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmPDC1pr-ScGDH3-ScPGK1t*	This work
pJA_NM L3.62	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmPDC1pr-KmGLN1-ScPGK1t*	This work
pJA_NM L3.68	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmPDC1pr-KmGLT1-ScPGK1t KmPDC1pr-KmGLN1-ScPGK1t*	This work
pJA_NM L3.69	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmPDC1pr-KmGLT1-ScPGK1t KmPDC1pr-KmGLN1-ScPGK1t KmPDC1pr-ScGDH1-ScPGK1t*	This work
pJA_NM L3.76	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmPDC1pr-KmGLT1-ScPGK1t*	This work
pJA_NM L3.77	pI4 MTU-DO-HPH with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmPDC1pr-KmGLT1-ScPGK1t KmPDC1pr-ScGDH1-ScPGK1t*	This work

2.2. Construction of CRISPR plasmids for introduction of double stranded breakages

To replace the pangenomic replication origin of CRISPR plasmid pUCC001(Varela *et al.*, 2019) with a high copy number origin specific to *K. marxianus* (*ARS2*), plasmid pJA_U_*gRNAempty* was constructed as illustrated in figure S1. It was then possible to introduce a protospacer DNA insert for specific targets into pJA_U_*gRNAempty* in the same manner as pUCC001 according to our recent protocol (Rajkumar and Morrissey, 2022). Briefly, complementary oligonucleotides of the gRNA protospacer sequences predicted by the SgRNAcas9 software (Xie *et al.*, 2014; Varela *et al.*, 2019) were designed and synthesised with 5′ and 3′ overhangs (5′-CGTC-3′ and 5′-AAAC-3′) on the sense and antisense strands, respectively. The complementary oligonucleotides were denatured, annealed, and phosphorylated with 1 µL of T4 polynucleotide kinase (10 U/µL, NEB) in a total reaction volume of 10 µL. Fifty femto mole of the annealed protospacer insert was then put together with 100 ng of pJA_U_*gRNAempty* in a BsaI/T7 ligase Golden Gate reaction. Bacterial transformants of reaction products were selected on kanamycin supplemented LB medium and colonies harbouring plasmids with correctly inserted gRNA protospacer DNA were confirmed by PCR using primer JA603 and the antisense oligonucleotide of each gRNA protospacer sequence. The CRISPR plasmids constructed in this study are listed in Table S2 and the protospacer sequences for CRIPSR targeting in Table S4.

2.3. Bacterial transformation and confirmation of plasmids

E. coli DH5α was transformed with Golden Gate reaction products and selected on LB medium supplemented with appropriate antibiotics. The presence of GFP dropout cassette in pYTK001 before part insertion and in integrative dropout plasmids, and RFP dropout cassette in the level II plasmids allowed green/white and red/white colony screening. The plasmids were confirmed by PCR using OneTaq Quick-Load Polymerase (M0486L, NEB) and sequencing for level I gene plasmids or NotI (FD0596, Fisher Scientific) digestion for level II and III plasmids.

3. Yeast strain construction

3.1. Transformations and confirmation of constructed strains

A variation of the lithium acetate/PEG transformation method was used for introduction of plasmids into *K. marxianus*(Rajkumar and Morrissey, 2022). All gene expressions were established by genomic integrations. For integrative plasmids, two micrograms of plasmid DNA were digested with SgsI (FD1894, Fisher Scientific), used to transform yeasts, and selected on appropriate selective medium. Successful integrations were confirmed by

screening yeast colonies by PCR using two sets of primers: one amplifying the product between a region upstream of an integration site and the promoter of the integrated fragment (or the first promoter of integrated fragments harbouring multiple concatenated TUs) while the second set of primers amplified the product between a region downstream the integration site and the selective marker gene sequence. For gene disruption by deletion, yeasts were co-transformed with DNA repair fragment (see section 3.2) and 150ng of the CRISPR plasmids described in section 2.2 and selected on YPD agar supplemented with 400mg/L hygromycin. The deletions were confirmed by colony PCR using diagnostic primers complementary to internal regions of the gene but outside the deleted region. The CRISPR plasmid was removed from a confirmed deletion strain by passaging on YPD before further establishing further modifications.

3.2. Gene disruption

Gene disruption was achieved by clean deletion of gene segments by previously described method(Rajkumar and Morrissey, 2022). Briefly, DSBs were generated in genes using the CRISPR plasmid described in section 2.2 and were repaired by homology directed repair with DNA repair fragments with flanks that are homologous to the 5' and 3' ends of the gene. Gene repair fragments were generated by synthesizing forward and reverse overlapping primers. The primers were annealed and extended in 50 µL PCR reactions using Q5 High-Fidelity Polymerase. Yeasts were then co-transformed with 20 µL of the 50 µL reaction mix and 150 ng of CRISPR plasmids and selected on YPD agar supplemented with 400 mg/L hygromycin. Gene deletions were confirmed by colony PCR using diagnostic primers complementary to internal regions of the gene but outside the deleted region. The CRISPR plasmid was removed from a confirmed deletion strain by passaging on YPD before establishing further modifications.

3.3. Construction of platform strain with increased production of aromatic amino acids.

To construct a platform strain with increased phenylalanine and tyrosine biosynthesis, the I6-Chor/A2-URA cassette which harbours genes of the shikimate pathway for overexpression in *K. marxianus* was first integrated at *ARO3* locus of strain KmASR.039 (*dln4aro4aro7ura3his3*) to yield strain KmASR.088. Next, I1-PPP-HIS was integrated into KmASR.088 genome at *LAC4* locus for overexpressing the native *TKL1*, *TAL1*, *RPE1* and *RKI1* of the pentose phosphate pathway. The resulting strain, KmASR.184, produced 2-phenylethanol and tyrosol as proxy metabolites for phenylalanine and tyrosine, respectively, compared to wildtype *K. marxianus* in which neither metabolite was not detected (Figure 1A).

3.4. Construction of cinnamic and coumaric acid producing strains.

The genealogy of the cinnamic and coumaric acid producing strains is illustrated in figure 1B. For testing ammonia lyase genes for cinnamic and coumaric acid production in *K. marxianus*, *LEU2* was disrupted in KmASR.184 to generate strain KmJA.010. Integrative expression cassettes of each gene were individually integrated at *ARO10* locus of KmJA.010, consequently disrupting *ARO10* and establishing the cassette in the genome for expressing each gene under the control of the *KmPDC1* promoter and *ScPGK1* terminator to generate strains KmJA.073, KmJA.074, KmJA.075, KmJA.076, KmJA.077. KmJA.007 was derived from KmJA.010 by integrating *AtC4H* and *AtCPR1* downstream of *HSP104* for expression of both genes under the control of the *KmENO1* promoter and *KmPGK1* terminator. To identify the best coumaric acid module, the ammonia lyase genes were integrated into KmJA.007 genome similar to in KmJA.10 to generate strains KmJA.078, KmJA.079, KmJA.080, KmJA.081, KmJA.082.

A multi-gene expression cassette of *PlPAL, AtC4H* and *AtCPR1* was integrated between *SWF1* and *ARO1* of KmJA.010 to generate strain KmJA.011 and the putative *K. marxianus* phenylpyruvate decarboxylase genes were subsequently deleted in the strain to create single and double deletion variants of *PDC5* and *ARO10* for testing how the deletions affect 2-phenylethanol and coumaric acid production. Strain KmJA.015 was constructed by deleting *PDC5* in KmJA.010 to generate KmJA.009 followed by additional deletion of *ARO10.* To optimise the coumaric acid module through gene dosage increase, cassettes with varying copies of *PlPAL, AtC4H* and *AtCPR1* were integrated into KmJA.015. For optimising the transamination of phenylpyruvate into phenylalanine/coumaric acid, the putative *ARO9* gene was deleted in KmJA.017 to generate strain KmJA.119. That strain was then further engineered by integrating cassettes harbouring only single copies of *PlPAL, AtC4H* and *AtCPR1* or in addition to glutamate biosynthesis genes.

4. Cultivation and sampling of production strains

For determining the yields of aromatic molecules produced by strains, they were precultured overnight in 10mL of minimal medium. For comparing flavonoid production when amino donors are used as sole nitrogen sources, glutamate, glutamine, leucine or methionine was added at a nitrogen concentration of 0.5g/L to minimal medium without $(NH_4)_2SO_4$ with its absence compensated for by adding equimolar concentration of K_2SO_4. Cultures were started at A 600nm of 0.1 in 50mL of fresh minimal medium in 250mL Erlenmeyer flasks and incubated at 30°C in a shaker at 200 rpm for 5 days.

For strains cultivated in bioreactors, these were performed in batch mode at minimal medium working volumes of 2.5L in 5L bioreactors (Applikon, The Netherlands) supplemented with

0.2g/L antifoam (Sigma-Aldrich). For bioreactor cultivation with glutamate as sole nitrogen source, 6.6g/L of monosodium glutamate was added to media. All bioreactor cultivations were started from precultures at exponential growth phase with pH uncontrolled and bioreactors sparged with pressurised air with dynamic agitation speed to ensure optimum aeration. On account of the respiro-fermentative physiology of *K. marxianus* (Morrissey *et al.*, 2015), aeration is an important factor in its cultivation. To identify suitable rate of aeration for bioreactor cultivation, metabolite production by strain KmJA.090 during cultivation at 1L/min versus 2L/min aeration were determined (figure S1). The strain showed shorter lag phase time, higher growth rate and increased coumaric acid yield by 45% at 2L/min aeration than 1L/min (figure S3B and C). This was accompanied by higher alcoholic fermentation, which directs glucose towards ethanol/acetate production and can potentially reduce the yield of coumaric acid, at lower aeration. Thus, bioreactor cultivations in this work were routinely conducted at 2L/min aeration rate.

Extracellular metabolites concentrations were determined in culture supernatants. For aromatics, supernatant was collected by mixing 700µL of cultures with 96% ethanol at 1:1 ratio, vortexed and pelleted by centrifugation using a benchtop centrifuge at 4^{o}C for 5 minutes at maximum speed. Glucose, ethanol, acetate, and glycerol concentrations were determined in undiluted culture. Supernatants were stored at -20^{o}C until analysis. For determining cell dry weights, 10mL of culture was added into a pre-weighed 50mL tube, washed with 10mL of ethanol to dissolve and remove aromatic metabolites from the biomass. The biomass was recovered in the tube by centrifugation at room temperature for 5 minutes at maximum speed and desiccated at 50^{o}C. The cell dry weight was then determined as the total mass of dried biomass and tube minus the initial mass of the tube.

5. Analytical methods

The concentrations of extracellular aromatics were determined by an Agilent 1260 Infinity HPLC Infinity system with Zorbax Eclipse Plus C18 column (4.6 x 150mm, 3.5µm; Agilent, Little Island, Ireland) operated at 40^{o}C with 20mM KH_2PO_4 brought to pH2 with phosphoric acid and supplemented with 1% acetonitrile/acetonitrile as the mobile phase at a flow rate of 0.8 mL/min as previously described(Koopman *et al.*, 2012). For elution, acetonitrile was increased from 0-10% for the first 6 minutes, then to 40% between 6 to 23 minutes, and then switched to 100% KH_2PO_4 for 23 to 27 minutes. Aromatic metabolites were detected by a VWD100 multi-wavelength detector with the absorbance of Ehrlich pathway metabolites measured at 200nm and 280nm for cinnamate and coumarate. Data were analysed using OpenLab CDS (Agilent). The elution time for the metabolites of interest in this study are as follows: 2-phenylethanol, 17.2 minutes; tyrosol,10.22 minutes; para-hydroxyphenylacetate,

11.17 minutes, phenylacetate, 17.86 minutes; phenylpyruvate, 14.43 minutes; cinnamic acid, 21.38; coumaric acid 14.1. All HPLC standards for metabolites were purchased from Sigma-Aldrich or Fisher Scientific. The concentrations of glucose and extracellular acetate, glycerol and ethanol were determined with an Agilent 1200 HPLC system (Agilent Technologies, CA, USA) with a REZEX 8µL 8% H+ organic acid column (300 ×7.8mm) (Phenomenex, CA, USA) at 65ºC using 0.01 N H_2SO_4 as the mobile phase at a flow rate of 0.6 mL/min and were detected by a refractive index detector. The concentration of residual glucose in media and extracellular concentrations of metabolites expressed as a proportion of glucose consumed (in mg/g) by strains are listed in Table S5.

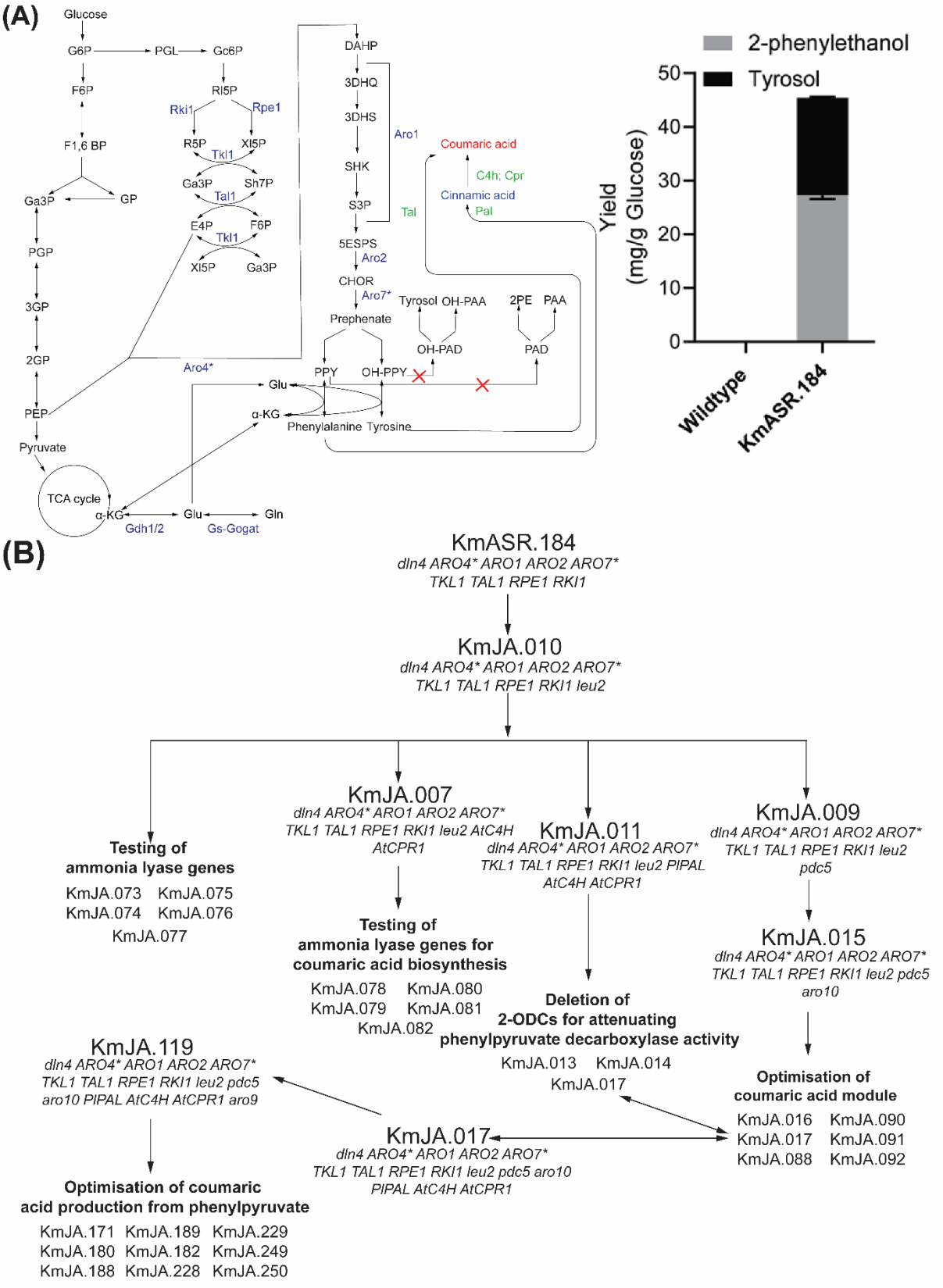


Figure 1. (A) Left panel: To increase phenylalanine and tyrosine production in *K. marxianus* the shikimate pathway was engineered by overexpressing the native *ARO1*, *ARO2*, and feedback resistant *ARO4* and *ARO7*, and the non-oxidative branch of the pentose phosphate pathway by overexpressing the native *RPE1, RKI1, TKL1* and *TAL1* to generate strain KmASR.184; Right panel: KmJA.184 produces 2-phenylethanol (2PE) and tyrosol on minimal glucose medium compared to wildtype *K.*

marxianus in which production was not detected. This indicates higher intracellular abundance of phenylalanine and tyrosine in KmASR184. This strain is an ideal platform for testing the phenylalanine and tyrosine routes for coumaric acid production. Yeast native enzymes are represented in blue texts, enzymes for catalysing plant reactions that are further established in KmASR.184 are represented in green, and reactions with red "X" represent the attenuated phenylpyruvate decarboxylase activity that can degrade aromatic amino acids. Metabolite abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F-1,6BP, fructose 1, 6-bisphosphate; Ga3P, glyceraldehyde 3-phosphate; DHAP, dihydroacetone phosphate; 1,3-BPG, 1, 3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; α-KG, α-ketoglutarate; Glu, Glutamate; Gln, glutamine; 6PGL, 6-phosphogluconolactone; 6PGc, 6-phosphogluconate; Rl5P, ribulose 5-phosphate; Xl5P, xylulose 5-phosphate; R5P, ribose 5-phosphate; Sh7P, sedoheptulose 7-phosphate; E4P, erythrose 4-phosphate; DAHP, 3-deoxy-D-arabino-heptulosonate-7-phosphate; 3DHQ, 3-dehydroquinate; 3DHS, 3-dehydroshikimate; SHK, shikimate; S3P, shikimate 3-phosphate; 5ESPS, 4-enolpyruvoyl-shikimate 3-phosphate; CHOR, chorismate; PPY, phenylpyruvate; OH-PPY, hydroxyphenylpyruvate; PAD, phenylacetaldehyde; OH-PAD, hydroxy-phenylacetaldehyde; 2PE, 2-phenylethanol; PAA, phenylacetic-acid; OH-PAA, hydroxy-phenylacetic-acid. Enzyme abbreviations: Rpe1, D-ribulose-5-phosphate 3-epimerase; Rki1, ribose-5-phosphate ketol isomerase; Tkl1, transketolase; Tal1, transaldolase; Aro4*, feedback resistant DAHP synthase; Aro1, Pentafunctional arom protein; Aro2, chorismate synthase; Aro7*, feedback resistant chorismate mutase; Pal, phenylalanine ammonia lyase; Tal, tyrosine ammonia lyase; C4h, cinnamic-acid 4-hydroxylase; Cpr, cytochrome P450 reductase; Gdh1/2, glutamate dehydrogenases; Gs-Gogat, glutamine synthetase-glutamate synthase (B) The genealogy of strains constructed from KmASR.184 in this study.

Results

1. Identifying a suitable module for coumaric acid production.

KmASR.184 is a *K. marxianus* strain that is engineered to increase flux through the pentose phosphate and shikimate pathways by overexpressing *KmTKL1*, *KmTAL1*, *KmRPE1*, *KmRKI1*, *KmARO4*^{fbr}, *KmARO1*, *KmARO2*, *KmARO7*^{fbr}. This increase in flux towards the biosynthesis of AAAs is illustrated by the production of 2-phenylethanol (27 mg/g glucose) and tyrosol (14mg/g glucose) (Figure 1A, right panel). As it is only engineered to increase flux towards prephenate, and *KmTYR1* and *KmPHA2* remain under native transcriptional regulation, KmASR.184 is ideal for comparing the phenylalanine and tyrosine routes for coumaric acid production. In order to have a selectable marker for later engineering steps, the *LEU2* gene was disrupted in KmASR.184 to create KmJA.010, which was then used for further strain constructions (Figure 1B). To identify a suitable module for establishing a functional starting point for phenylpropanoids production in *K. marxianus*, ammonia lyase genes from plants, bacteria and fungi were expressed under the control of the *KmPDC1* promoter and the *ScPGK1* terminator. To reduce flux leakage from the pathway, the ammonia lyase genes being tested were integrated into the *KmARO10* locus of KmJA.010. *ARO10* which encodes a phenylpyruvate decarboxylase enzyme that degrades AAAs in yeasts. Strains KmJA.073 and KmJA.075, expressing *Arabidopsis thaliana PAL1* (*AtPAL1*) and *Photorhabdus luminescens PAL* (*PlPAL*), respectively, produced cinnamic acid with KmJA.075 producing the higher yield of 1.3mg/g glucose (Figure 2A). Strains KmJA.076 (*Streptomyces maritamus PAL*, *encP*) and KmJA.077 (*Rhodobacter capsulatus TAL*, *RcTAL*) produced coumaric acid yields of 0.23mg/g and 0.1mg/g, respectively, with no cinnamic acid detected in both strains. KmJA.074

(*Rhodotorula glutinis PTAL*, *RgPTAL*) failed to produce neither cinnamic acid nor coumaric acid. Next, the five genes were similarly expressed in KmJA.007, which expresses *A. thaliana* cinnamic acid 4-hydroxylase (*AtC4H*) and cytochrome P450-reductase (*AtCPR1*) genes, which encode the enzymes necessary for converting cinnamic acid to coumaric acid (Figure 2B). In this case, the five strains produced coumaric acid, with KmJA.080 (*PlPA*L) being the highest yielding strain producing 6.5mg/g coumaric acid. KmJA.078 (*AtPAL1*) and KmJA.079 (*RgPTAL*) produced ~2mg/g coumaric acid. The result with KmJA.079 is interesting because *RgPTAL* has been reported to deaminate both tyrosine and phenylalanine (Zhu *et al.*, 2013; MacDonald *et al.*, 2016) but our data here indicate that in *K. marxianus,* RgPtal has Pal but no detectable Tal activity (since coumaric acid was not detected in KmJA.074 but was in KmJA.079). The fact that cinnamic acid was not detected in KmJA.078, KmJA.079 or KmJA.080 indicates that AtC4h and AtCpr1 efficiently converted cinnamic acid into coumaric acid. Based on the fact that the levels of coumaric acid detected in strains expressing *AtPAL1, RgPTAL*, *encP* and *RcTAL* in addition to *AtC4H* and *AtCPR1* were substantially lower than that achieved with *PlPAL* (Figure 2B), it was concluded that the best coumaric acid module in this host was the combined expression of *PlPAL*, *AtC4H* and *AtCPR1*.

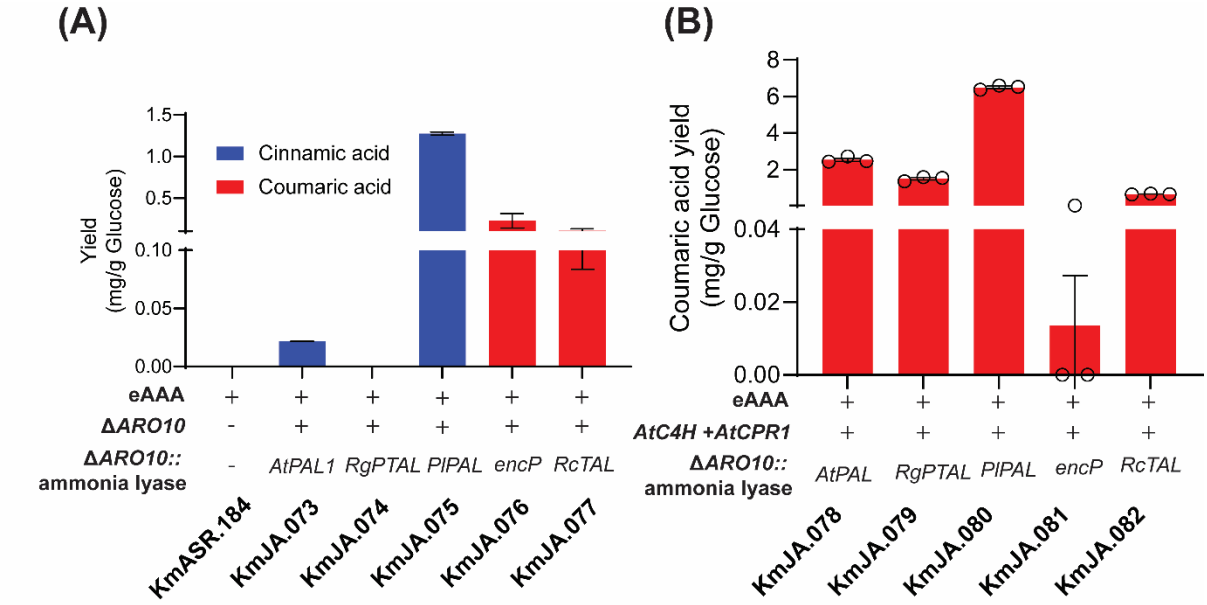


Figure 2. Heterologous expression of plant, bacterial or fungal ammonia lyases to identify a suitable coumaric acid module in *K. marxianus*. (A) The phenylpyruvate decarboxylase gene (*ARO10*) was disrupted by integrating phenylalanine and tyrosine ammonia lyase (*PAL* and *TAL*) genes from *Arabidopsis thaliana, Rhodotorulla glutinis, Photorhabdus luminescens Streptomyces maritamus* or *Rhodobacter capsulatus* at the gene locus of KmASR.184 for expression under the control of *KmPDC1* constitutive promoter to identify the strain with the highest cinnamic acid yield; (B) *A. thaliana* cinnamic acid 4-hydroxylase (*AtC4H*) and cytochrome P450 reductase (*AtCPR1*) genes were additionally integrated into the genome to identify the ammonia lyase that produces the highest coumaric acid yield. Plus (+) or minus (-) sign denote that a described modification in a strain is present or absent, respectively. All strains were cultivated in flasks of minimal glucose medium. Data shown are the mean ± standard deviation of three replicates.

2. Attenuating phenylpyruvate decarboxylase activity

The final reaction in the biosynthesis of phenylalanine is the conversion of the 2-oxo acid phenylpyruvate to phenylalanine by transaminase reaction. Phenylpyruvate and other 2-oxo acids are also substrates for decarboxylases that remove a carboxyl group to form an aldehyde, that is then oxidised or reduced into a fusel alcohol or acid, depending on the cellular redox status (Hazelwood *et al.*, 2008). This is the process by which 2-phenylethanol is formed and, as well as being part of the Ehrlich amino acid degradation pathway, has been postulated to serve as a redox sink and a means by which *S. cerevisiae* detoxifies 2-oxo acids (Romagnoli *et al.*, 2012). To maximise flux towards phenylalanine biosynthesis, we wanted to eliminate 2-oxo acid decarboxylase (2-ODC) activity in coumaric acid producers. In *S. cerevisiae*, *PDC1*, *PDC5* and *ARO10* encode 2-ODC and the orthologues in *K. marxianus* have been identified (Hassing *et al.*, 2019). We assessed the impact of deleting either or both *PDC5* and *ARO10* in strain KmJA.011, which was equivalent to KmJA.080, except that *PlPA*L was integrated at a locus between *SWF1* and *ARO1* rather than the *ARO10* locus. *PDC1* was not deleted as Pdc1 is the main contributor of pyruvate decarboxylase in *K. marxianus* (Choo *et al.*, 2018). Moreover, the *S. cerevisiae* Pdc1 orthologue was reported to show no phenylpyruvate decarboxylase activity unlike Pdc5 and Aro10 (Romagnoli *et al.*, 2012). First, higher production of 2-phenylethanol was detected for KmJA.011 than KmASR.184 (figure 2A). However, the deletion of either *PDC5* (KmJA.013) or *ARO10* (KmJA.014) or both genes (KmJA.017) abolished the production of 2-phenylethanol (2-PE) in KmJA.011, indicating that both genes are required for optimum phenylpyruvate decarboxylase activity (Figure 3A). As predicted, attenuating the flux towards 2-PE substantially increased the yield of coumaric acid in the three strains (Figure 3B).

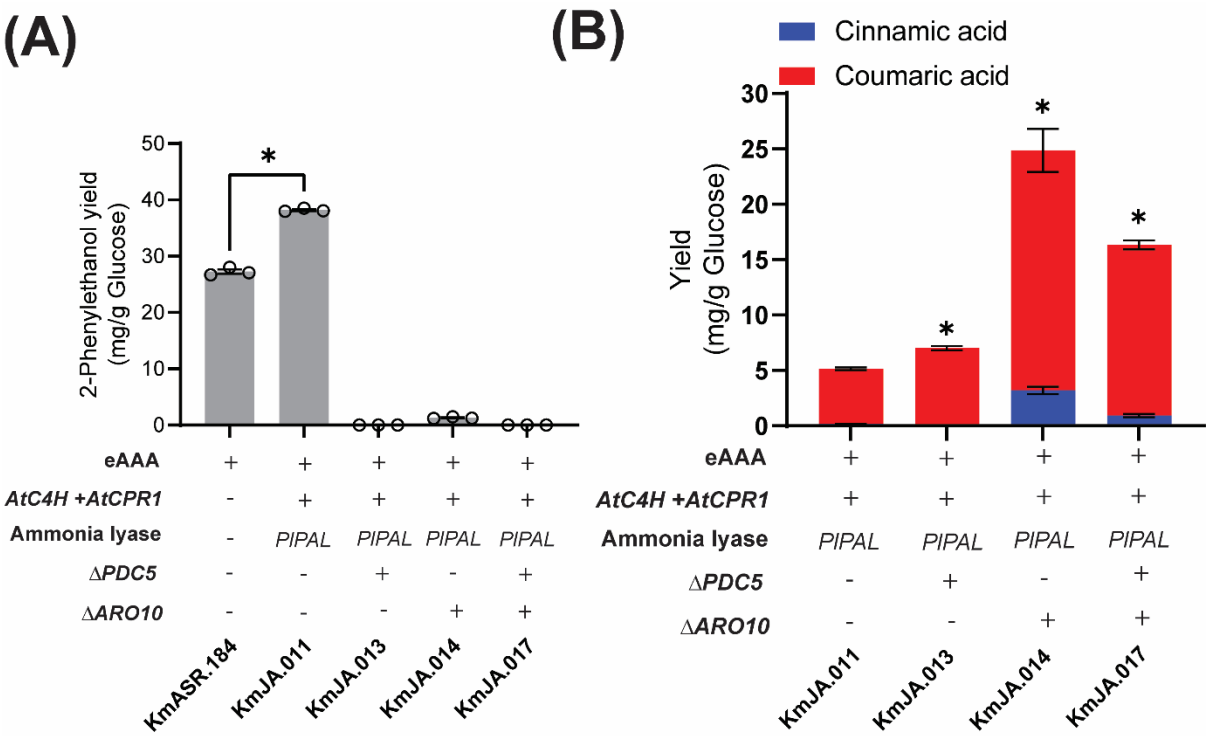


Figure 3. Attenuation of phenylpyruvate decarboxylase activity by deleting putative *K. marxianus* 2ODC genes. *Photorhabdus luminescens PAL* (*PlPAL*), *A.thaliana C4H* and *CPR1* (*AtC4H* and *AtCPR1*) were integrated into the genome of KmASR.184 without disrupting any of its 2ODC genes to generate KmJA.011. *PDC5* and *ARO10* were subsequently deleted in KmJA.011 to derive KmJA.013, KmJA.14 and KmJA.017. All strains were cultivated in minimal glucose medium flasks and the yield on glucose of 2-phenylethanol (A) and extracellular cinnamic/coumaric acid (B) were determined. Plus (+) or minus (-) sign denote that a described modification in a strain is present or absent, respectively. Data shown are the mean ± standard deviation of three replicates. Statistically significant differences (examined by independent ttest) in extracellular metabolite yield by strains relative to control strains, KmASR.184 and KmJA.011 are denoted by an asterisk (*, $p<0.05$).

3. Optimising the coumaric acid module

We wished to investigate the effect of varying the copy number of the genes that encode the coumaric acid module (*PlPAL, AtC4H* and *AtCPR1*). Therefore, we constructed KmJA.015 (KmASR.184; *leu2pdc5aro10*) to serve as the parent strain for this purpose. The bioconversion of cinnamic acid to coumaric acid by C4h is a hydroxylation reaction, requiring electron transfer from NAD(P)H to C4h by cytochrome P450 reductase activity (Cpr). Since yeasts have native Cpr encoded by *NPE1* which may be sufficient for AtC4h catalysis, we first investigated whether heterologous expression of Cpr was necessary by constructing strains KmJA.016 and KmJA.017, which differed only in the presence of *AtCPR1* in the latter. While KmJA.016 accumulated 5mg/g of cinnamic acid and no coumaric acid was detected, KmJA.017 produced 15mg/g of coumaric acid with low yield of cinnamic acid (Figure 4A). This shows the need for heterologous Cpr expression. As endogenous enzymes might compete for CPR activity, we next constructed strains KmJA.091 and KmJA.092, which have two copies of *AtC4H* in order to increase the abundance of the enzyme molecule available to compete, but this did not increase in coumaric acid production (Figure 4A). We further assessed these

strains by looking at growth parameters (Figure 4B) and glycerol production (Figure 4C). Strains lacking any copy of *AtCPR1* failed to fully consume glucose, produced lower biomass and secreted higher yields of glycerol than strains expressing *AtCPR1*. These results are indicative of redox imbalance and stress. The comparison of strains expressing one (KmJA.017) or two (KmJA.092) copies of *AtC4H* was interesting as the latter produced lower yield of coumaric acid but higher biomass and no glycerol. We also assessed the effect of increasing the copy number of *PlPAL* and found that an additional copy doubled the yield of coumaric acid (Figure 4D, strain KmJA.088). Finally, increasing the copy number of *AtCPR1* to two in alignment with the higher copy number of *PlPAL* and *AtC4H* further increased the yield of coumaric acid (Figure 4D, strain KmJA.090).

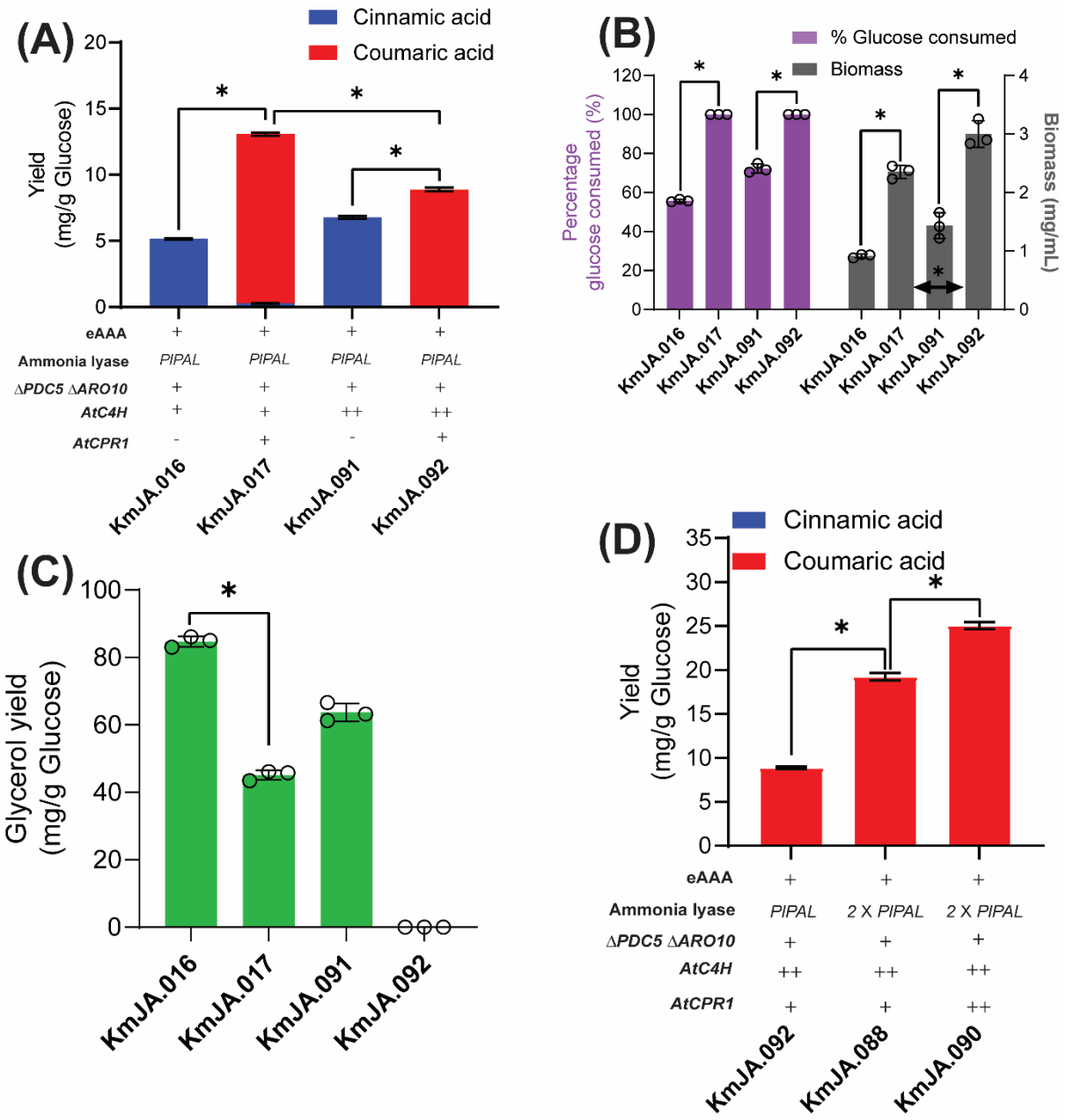


Figure 4. Optimising the coumaric acid module. (A) The yield on glucose of cinnamic/coumaric acid were determined in strains that have a single copy of *PlPAL* and varying copy numbers of *AtC4H* and *AtCPR1* integrated into their genomes. The percentage glucose consumed and biomass yield (B), and glycerol yield (C) were also determined. (D) *PlPAL* copy number was increased from one in KmJA.092 to two in KmJA.088 to test whether this strategy increased coumaric acid production. Finally, *AtCPR1* copies were increased to two, resulting in the integration of two copies of the entire coumaric acid module in KmJA.090. This strain produced the highest coumaric acid yield. All strains were cultivated in minimal glucose medium flasks and the yields on glucose of extracellular metabolites were determined. Plus (+) and minus (-) sign denote that a described modification is present or absent in a strain. Data shown are the mean ± standard deviation of three replicates. Statistically significant difference (examined by independent ttest) in extracellular metabolite yield between strains that are linked by a black line is denoted by an asterisk (*, $p<0.05$).

4. Glutamic acid is a preferred amino donor for aminating phenylpyruvate to phenylalanine in *K. marxianus*

Having established the optimum coumaric acid module with phenylalanine as precursor in strain KmJA.090, we re-visited the phenylalanine biosynthetic pathway to determine whether further optimisation was possible. The final step in the biosynthesis of phenylalanine involves the transamination of phenylpyruvate to phenylalanine, which is a reversible reaction catalysed by aromatic amino transferases (figure 5A). This reaction requires an amino donor and previous work in *S. cerevisiae* found that different amino acids displayed varying capacities to act as amino donors (Urrestarazu *et al.*, 1998). To assess whether the nitrogen source used in growth medium would affect the yield of coumaric acid KmJA.090 was cultivated in minimal medium with glutamate, glutamine, leucine, methionine, or ammonium sulphate as sole nitrogen sources. Concentrations were adjusted such that there was 0.5g of nitrogen/L in all cultivations to enable direct comparison. At the end of the cultivation, glucose was completely depleted in all the cultures but the use of amino acids as nitrogen source led to a higher final biomass formation than ammonium sulphate (Figure 5B). This may be explained by the fact that these amino acids also provide carbon in addition to glucose in the medium. Thus, to allow for direct comparison, the biomass-normalised yield of coumaric acid on glucose was determined. This value was similar for ammonium and glutamate but lower for glutamine, leucine and methionine suggesting that glutamate may be more suitable as an amino donor than the other amino acids for coumaric acid production.

We next assessed the performance of KmJA.090 when grown in a bioreactor with controlled aeration (Figure 5C) with either ammonium sulphate (left panel) or glutamate (right panel) as sole nitrogen sources. When growing on glutamate, the strain completely consumed the glucose by 18hrs compared to ammonium in which 25% (3.8g/L) of the starting glucose remained at stationary growth phase. This was accompanied by higher biomass formation in glutamate than ammonium (figure S2). Ethanol and acetate were produced in both nitrogen sources with higher ethanol yields in ammonium than glutamate. During growth on glutamate,

the ethanol and acetate produced during glucose consumption was later completely metabolised whereas the picture was more complex on ammonium. In this case, ethanol was consumed but later accumulated whereas acetate was not consumed. There was less glycerol accumulated in glutamate suggesting less redox imbalance. Importantly, difference in coumaric acid production was striking with yield on glutamate 31% higher than ammonium sulphate (42mg/g versus 32mg/g).

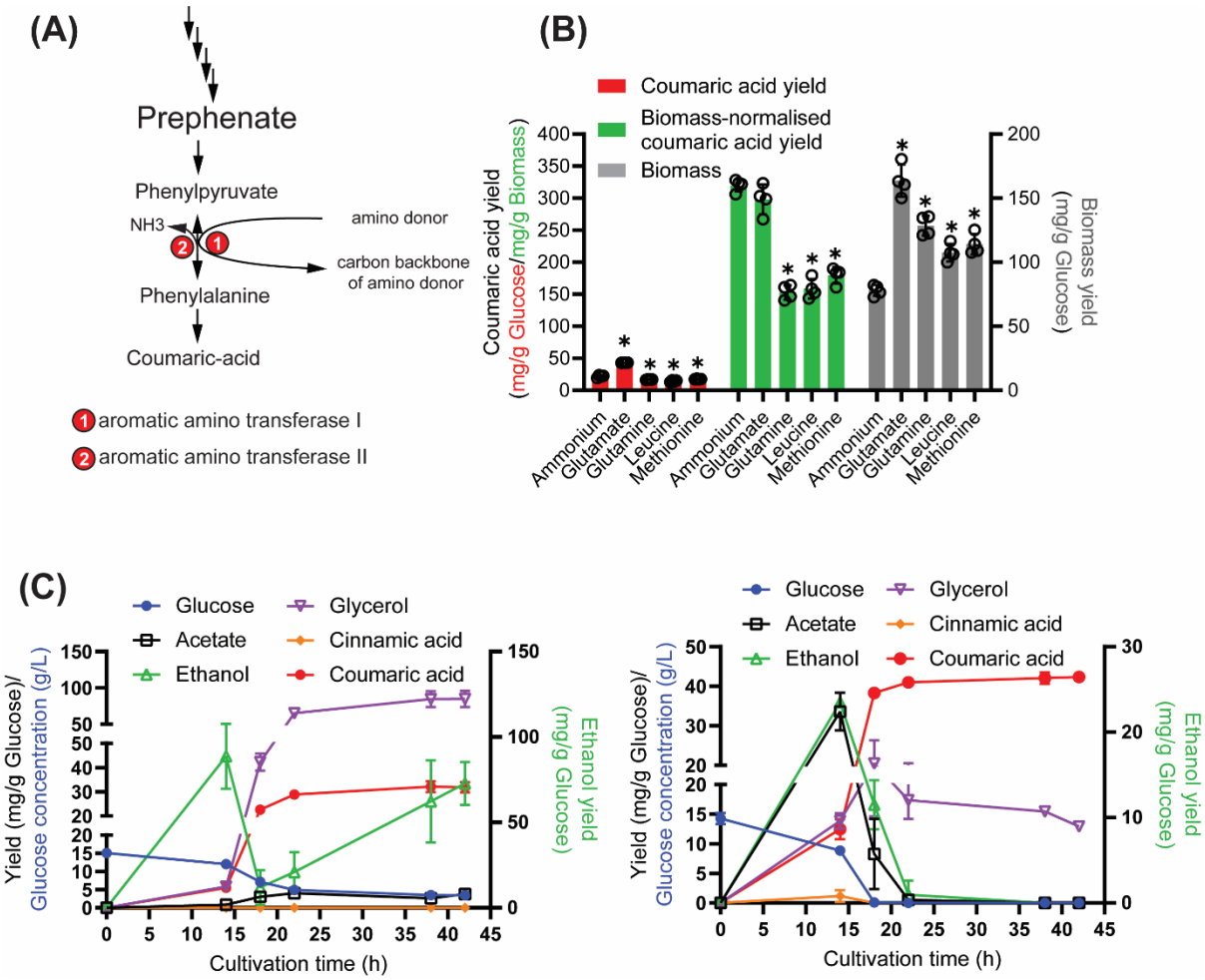


Figure 5. Cultivation of coumaric acid producer in amino donors. (A) Prephenate produced in the shikimate pathway is converted into the aromatic 2-oxo acid, phenylpyruvate from which phenylalanine is produced through the action of aromatic amino transferase I. This requires an amino donor to provide the amino group for transfer unto phenylpyruvate. (B) Coumaric acid and biomass yields on glucose were determined in KmJA.090 cultivated in minimal glucose medium with ammonium sulphate or known amino donors (glutamate, glutamine, leucine or methionine) as sole nitrogen sources. All nitrogen sources were added at final nitrogen concentrations of 0.5gN/L. Data shown are the mean ± standard deviation of duplicates from two independent flask experiments. Statistically significant differences in yield (examined by independent ttest) relative to control are denoted by an asterisk (*, $p<0.01$). (C) Metabolite yield and glucose consumption were determined in KmJA.090 cultivated in minimal glucose medium in batch bioreactor aerated at 2L/min with ammonium sulphate (left panel) or glutamate (right panel) as sole nitrogen sources. Both nitrogen sources were added at final nitrogen concentrations of 0.5 g/L. Data shown are the mean ± standard deviation of three bioreactor replicates.

5. Optimising coumaric acid production from phenylpyruvate on ammonium

Although glutamate appears to be a better nitrogen source than ammonium for coumaric acid production, this may not be feasible in a commercial process, so we explored additional ways to improve yield on ammonium. First, we disrupted the *K. marxianus* orthologue of *ARO9*, the transaminase that favours the conversion of phenylalanine to phenylpyruvate in *S. cerevisiae* (Figure 6A) and found that the *aro9* deletion strain (KmJA.171) produced 13% more coumaric acid than its parent KmJA.090 (Figure 6A). Next, we looked at the pathway by which ammonium is assimilated in the cell (Figure 6B). This involves its incorporation into glutamate by NADP+-glutamate dehydrogenase (Gdh1/3) and into glutamine by the concerted catalysis of glutamine synthetase (Gln1) and glutamate synthase (Glt1) (Gs- Gogat). Our hypothesis was that by modulating the glutamate biosynthetic pathway through the overexpression of the genes that encode the involved enzymes, nitrogen would be more effectively incorporated into glutamate, thereby increasing the available pool of glutamate to support the biosynthesis of phenylalanine and coumaric acid. To test this hypothesis, genes encoding variants of these enzymes were expressed in KmJA.171. The only strain that showed an increase in coumaric acid yield was KmJA.188, which overexpressed *S. cerevisiae GDH1* and had a 68% increase in yield relative to KmJA.171. Overexpressing the *K. marxianus* orthologue *KmGDH1* led to a small decrease in yield whereas expressing the *S. cerevisiae* paralogue *GDH3* severely reduced the yield. All the efforts to manipulate the Gs-Gogat part of the pathway also reduced yields indicating that this is not a productive approach to improving coumaric acid production in this particular strain.

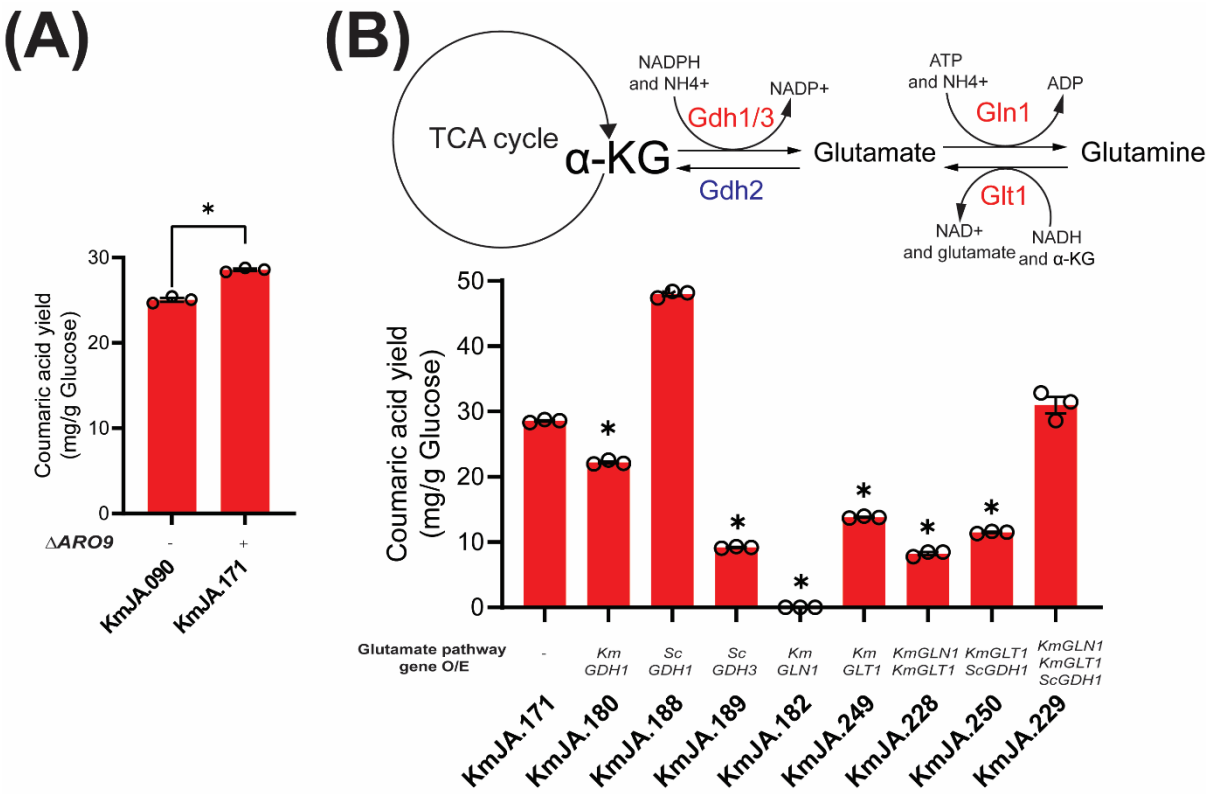


Figure 6. Deleting *ARO9* and overexpressing glutamate biosynthetic genes in coumaric acid producer. (A) Coumaric acid yield by *ARO9*-deleted version of KmJA.090 (KmJA.171) and KmJA.090 itself during cultivation in minimal glucose medium. (B) Top panel: an illustration of glutamate/glutamine biosynthesis from α-ketoglutarate (α-KG): NADP+-dependent glutamate dehydrogenase (Ghd1/3) forms glutamate by aminating α-KG using intracellular ammonium and glutamine synthetase (Gln1) forms glutamine through ATP-dependent amination of glutamate, also using intracellular ammonium. NAD+-dependent glutamate synthase (Glt1) can convert glutamine back into glutamate, transferring an amino group unto α-KG in the process to form an extra glutamate molecule. (B) Coumaric acid yield of strains with similar modifications as KmJA.171 in addition to the integration of the genes of the glutamate biosynthetic pathway for expression under the control of a constitutive promoter when cultivated in flasks of minimal glucose medium. Data shown are the mean ± standard deviation of three replicates. Statistically significant differences in yield (examined by independent ttest) relative to KmJA.171 are denoted by an asterisk (*, $p<0.05$).

6. Cultivation of coumaric acid producer in batch bioreactor

In flask cultivation with ammonium as the nitrogen source, the yield of coumaric acid achieved by KmJA.188 was almost twice that of our previous best strain KmJA.090 so we decided to assess performance in a bioreactor. Unexpectedly, the yield of coumaric acid was lower in the bioreactor than in flasks (37 versus 48mg/g) (Figure 7A, left panel). We speculated that the requirement of Gdh for NADPH may cause redox imbalance and the strong aeration in the bioreactor means that there is little alcoholic fermentation occurring, limiting the reduction of NADP+ back into NADPH by aldehyde dehydrogenases. To test this, we reduced the aeration from 2L/min to 1L/min, which did cause ethanol production to go from a maximum yield of 14mg/g to 126 mg/g without any real difference in glycerol yield (Figure 7A). This was accompanied by 13% higher coumaric acid yield and higher biomass was produced when aerated at 1L/min compared to 2L/min by stationary phase (Figure 7B). At this lower rate of

aeration, the yield of coumaric acid by strain KmJA.188 remained 31% higher than that of KmJA.090 aerated at 2L/min (Figure 5C, left panel). Therefore, the improved yield evident in flasks due to *ScGDH1* overexpression was transferred to bioreactor scale despite production being lower in bioreactors than in flasks.

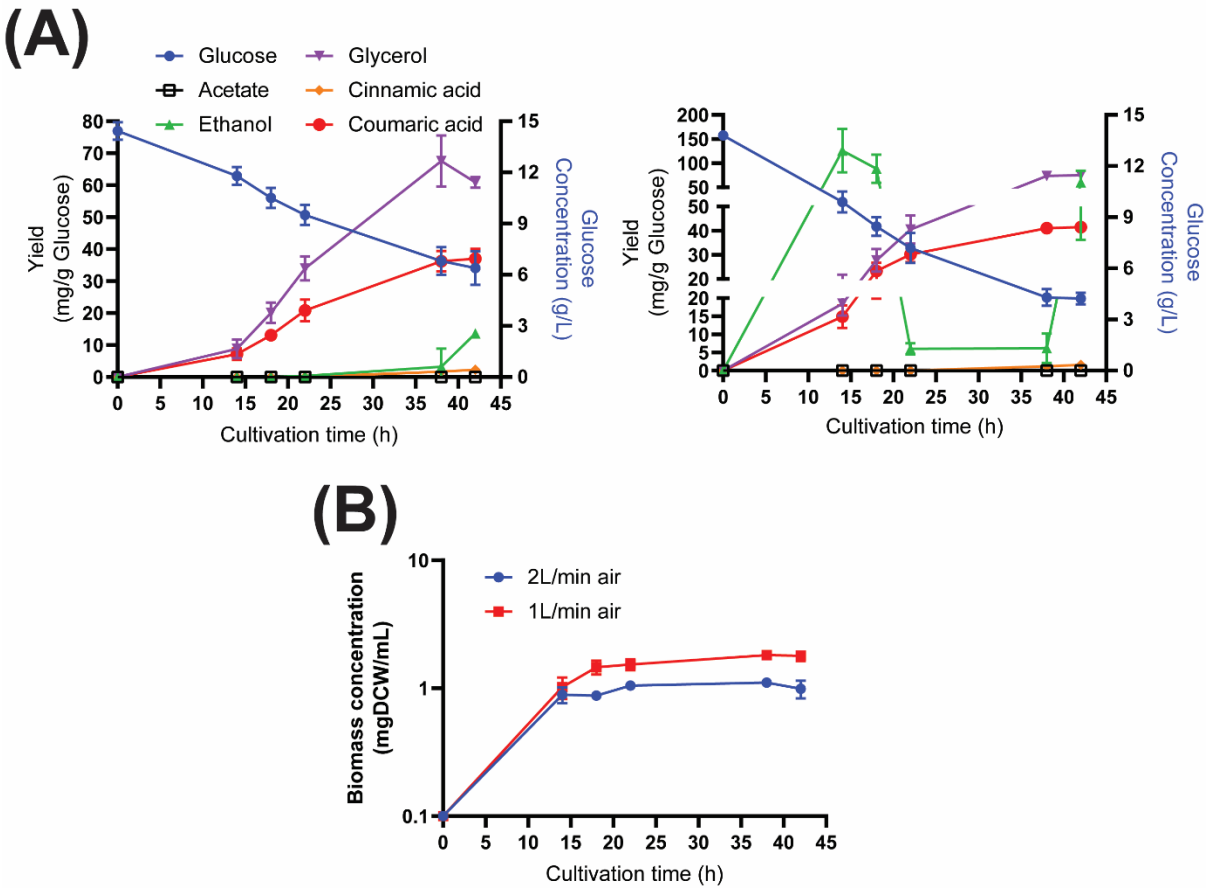


Figure 7. Aeration limits glucose consumption and coumaric acid production in coumaric acid producer that is overexpressing *ScGDH1*. (A) Strain KmJA.188 which expresses *ScGDH1* and produces the highest coumaric acid yield in flask was cultivated in minimal glucose medium with ammonium sulphate in batch bioreactor aerated at 2L/min (left panel) and at 1L/min (right panel). The growth profile of the strain in each condition was also determined (B). Data shown are the mean ± standard deviation of three bioreactor replicates.

Discussion

The main attraction of *K. marxianus* for industrial application lies in its potential to metabolise and grow on diverse feedstocks. In that light, *K. marxianus* is continuously being developed for industrial productions that are potentially cost-effective and compatible with the circular economy paradigm (Bilal *et al.*, 2022). With recent efforts to understand and improve *K. marxianus* sugar metabolism (Hua *et al.*, 2019; Donzella *et al.*, 2021) laying the foundation for developing the yeast for cheap and sustainable bioprocesses, we successfully engineered *K. marxianus* for the first time to produce coumaric acid, the branch-point molecule from which phenylpropanoids are produced, in this work. Coumaric acid production was systematically engineered into a *K. marxianus* platform that was initially engineered for increased

phenylalanine and tyrosine production through the overexpression of feedback resistant *KmARO4* and *KmARO7* and the native *ARO1, ARO2* to increase flux through the shikimate pathway. Increased availability of the E4P precursor was additionally engineered by overexpressing the native *TKL1, TAL1, RPE1* and *RKI1*. These modifications culminated a platform strain that produces 2-phenylethanol at 27mg/g of glucose and 18mg/g of tyrosol (Figure 1A), equivalent to 3.4mM and 2mM respectively. Firstly, it was demonstrated in this platform strain that the phenylalanine route is superior to the tyrosine route for coumaric acid production, leading to the identification of *PlPAL* as the ammonia lyase of choice, and in addition to *AtC4H* and *AtCPR1*, as the best coumaric acid module in this work (figure 2).

The deletion of putative native 2-oxo acid decarboxylase genes (*KmPDC5* and *KmARO10*) whose products potentially have phenylpyruvate decarboxylase activities in *K. marxianus* diverted flux away from phenylpyruvate towards coumaric acid, leading to a 3.2-fold increase in coumaric acid yield in the double deletion strain compared to when neither gene was deleted (Figure 3B). Interestingly, while 2-phenylethanol production was detected in KmJA.013 (*pdc5*), only a moderate but significant increase in coumaric acid yield was detected compared to KmJA.014 (*aro10*) which showed a 5-fold increase. In an effort to find 2-oxo acid decarboxylase enzymes with pyruvate decarboxylase but no phenylpyruvate decarboxylase activity for producing heterologous aromatics in *S. cerevisiae,* Hassing and colleagues tested several orthologues from different yeasts including *KmPDC5* and *KmARO10* but was unable detected either activity for *KmPDC5* (Hassing *et al.*, 2021). Their result indicated that *KmPDC5* product might not be a 2ODC. Bioinformatic analysis showed that the protein displays higher similarity with Thi3 proteins from other yeasts than 2-oxo acid decarboxylases, suggesting that *KmPDC5* could in fact be a *THI3* whose product is a transcription factor that regulates thiamine biosynthetic genes (data presented in Chapter 4). If that is the case, the deletion of *KmPDC5* in this work could have resulted in loss of 2-phenylethanol production due to its possible regulation of phenylpyruvate decarboxylase activity in *K. marxianus.*

The expression of *PlPAL* and *AtC4H* without heterologous CPR in KmJA.016 resulted in the inability of the strain to convert cinnamic acid to coumaric acid (Figure 4A), demonstrating that the yeast's endogenous Cpr activity is insufficient for electron transfer to AtC4h for cinnamic acid hydroxylation. This is probably due to competition for endogenous CPR activity by important and indispensable yeast native cellular functions. However, increase in *AtC4H* dosage in KmJA.091 increased cinnamic acid production, suggesting a possibility that the abundance of AtC4h resulted in some cinnamic acid hydroxylation. This will give room for more cinnamic acid production because the inhibitory effect of cinnamic acid on Pal is reduced. This effect has been shown in plants to regulate the entry point of the phenylpropanoid pathway (Blount *et al.*, 2000; Yin *et al.*, 2012) and may also be at play in

yeast hosts. The expression of *AtCPR1* (in KmJA.017 and KmJA.092) did not only result in the conversion of cinnamic acid into coumaric acid but also increased glucose consumption, biomass formation and reduced glycerol production (Figure 4B and C). With the hydroxylation of cinnamic acid by C4h relying on Cpr activity, it can be inferred that improved C4h hydroxylation of cinnamic acid by Cpr activity reduced glycerol production (figure 4C). This can be explained by the fact that the hydroxylation requires electron transfer from NAD(P)H to C4h by Cpr to form NAD(P)+ in the process. This will reduce glycerol production by glyceraldehyde 3-phosphate dehydrogenase (Gapdh) as the need for NADPH re-oxidation is alleviated by Cpr activity. The alleviation will benefit coumaric acid production because the diversion of glucose by Gapdh toward glycerol production, which can reduce cinnamic/coumaric acid yield on glucose, is reduced.

KmJA.090 represents a strain containing similar modifications as have been established by others in *S. cerevisiae.* Although higher coumaric acid production of 1.93g/L (95mg/g glucose) was achieve elsewhere in *S. cerevisiae* strain with *PDC5* and *ARO10* deletion and overexpressing feedback resistant *ARO4* and *ARO7, E. coli* shikimate kinase II (*aroL*) and *Flavobacterium johnsoniaeu TAL* in complex medium (Rodriguez *et al.*, 2015), subsequent attempt by Lui and colleagues to reproduce the result with the same modifications in minimal medium resulted in 22.3mg/L (1mg/g glucose) of coumaric acid (Liu *et al.*, 2019), which is comparable to the yield reported for KmJA.082 expressing *RcTAL* in our work here (Figure 2). Pal was used in the more recent attempt to produce coumaric acid in *S. cerevisiae* in which a yield of 151mg/g was achieved in flask (Liu *et al.*, 2019). While this yield is 6-fold higher than in KmJA.090 in this current study, it is not directly comparable because the *S. cerevisiae* coumaric acid producer was cultivated in slow glucose release medium.

To further optimise coumaric acid production, we hypothesised that the amination of phenylpyruvate into phenylalanine could be a limiting step due to inadequate intracellular amino donor for the reaction. Indeed, KmJA.090 produced higher coumaric acid yield in glutamate than in other amino donors tested (Figure 5B), suggesting that increase in intracellular glutamate could improve phenylpyruvate amination into phenylalanine. Moreover, the link between glutamate biosynthesis, ammonium assimilation and the TCA cycle (Figure 6B) is an important factor to consider especially in respiro-fermentative yeasts like *K. marxianus*. Such physiology can potentially predispose the yeast to lower intracellular glutamate availability for amino transfer due to competition for α-ketoglutarate by the TCA cycle whose flux could be higher than in fermentative yeasts such as *S. cerevisiae.* Unlike *S. cerevisiae* which has two NADP(+)-dependent Gdh isoforms (*GDH1* and *GDH3*), *K. marxianus* has one with similar percentage of sequence homology to ScGdh1 and 3. The marked disparity in the effect of *KmGDH1* and *ScGDH1* overexpression on coumaric acid yield (figure

6B) suggests that the two gene products may have different post-transcriptional/translational regulations in the *K. marxianus* host. In *S. cerevisiae* where the Gdh reaction is the main contributor to glutamate production (Campero-Basaldua *et al.*, 2017), ScGhd1 displays greater rate of α-ketoglutarate utilisation than ScGdh3 (DeLuna *et al.*, 2001; Tang, Sieg and Trotter, 2011) and *GDH1* is transcribed at all growth phases (Lee *et al.*, 2012; Campero-Basaldua *et al.*, 2017), the reaction is monitored by post-translational regulation of Gdh1/Gdh3 ratio through growth phase-specific degradative ubiquitination of Gdh (Mara *et al.*, 2018). Such regulation and the kinetic difference between the two *S. cerevisiae* isoforms are proposed to monitor α-ketoglutarate utilisation by Gdh, ensuring that the TCA cycle is not compromised. Gdh1 is responsible for α-ketoglutarate utilisation on glucose, a condition where ATP is mainly generated via alcoholic fermentation and flux through the TCA cycle is generally low, while Gdh3p is the main isoform for α-ketoglutarate utilisation on ethanol, a condition where ATP mainly comes from oxidative phosphorylation/TCA cycle. The respiratory nature of *K. marxianus* could warrant a stricter regulation of α-ketoglutarate utilisation by its single Gdh1 in order to maintain flux through the TCA cycle. It is possible that this regulation is specific to KmGdh1 but not ScGdh1. In line with this is the fact that the overexpression of *ScGDH1* increased coumaric acid production but *KmGDH1* did not in this work. Interestingly, the overexpression of *KmGLT1* and *KmGLN1* individually or together showed negative effect on coumaric acid production (figure 6B), suggesting that the modifications did not increase intracellular glutamate. A possible explanation for this is that the catalysis of glutamine to glutamate by Glt1 was poorer than that of glutamate to glutamine by Gln1. This is supported by the fact that *KmGLN1* overexpression by itself in KmJA.182 resulted in complete loss of coumaric acid and reduced production in KmJA.228 which additionally overexpressed *KmGLT1*. Thus, *GLT1* overexpression in our work might be insufficient for converting the glutamine pool into glutamate.

The cultivation of KmJA.090 in aerated bioreactors showed that the strain produced higher yield of coumaric acid at 2L/min versus 1L/min (figure S3B and C). The effect of aeration on coumaric acid production can be explained in light of the link between central carbon metabolism and AAA biosynthesis. *K. marxianus* metabolises pyruvate via the TCA cycle in aerobiosis, diverting pyruvate away from metabolism by pyruvate decarboxylase and reducing alcoholic fermentation (Sakihama *et al.*, 2019). Increased aeration can benefit coumaric acid production in two ways: first, reduced alcoholic fermentation implies reduced loss of glucose to unwanted side reactions. Second, the metabolism of pyruvate via the TCA cycle can increase α-ketoglutarate. Since α-ketoglutarate is a precursor of glutamate, increased aeration can provide higher intracellular glutamate for aminating phenylpyruvate into phenylalanine, thereby benefiting coumaric acid production. Alcoholic fermentation is however important for

reducing NAD(P)+ back into NAD(P)H in yeasts in order to drive glucose assimilation via glycolysis. Indeed, increase in aeration for KmJA.090 from 1 to 2L/min evidently reduced flux towards alcoholic fermentation, as indicated by reduced ethanol/acetate production, and increased coumaric acid production by 1.5-fold, but also resulted in incomplete glucose consumption accompanied by slightly higher glycerol production (figure S3C), suggesting that the strain suffered from redox imbalance.

With *ARO9* deletion and *ScGDH1* overexpression in KmJA.188 increasing phenylalanine (coumaric acid) production compared to KmJA.090, these modifications also considerably reduced alcoholic fermentation (ethanol/acetate production) compared to KmJA.090 (left panel of figure 7A versus bottom panel of figure S3C). This indicates that those additional modifications created a thermodynamic “pull” on the TCA cycle through increased catalysis of α-ketoglutarate to glutamate by ScGdh1, thereby potentially increasing the metabolism of pyruvate via the TCA cycle. However, the strain also displayed poorer glucose consumption and biomass formation than KmJA.090 (figure 7B versus bottom panel of figure S3B). The poorer growth parameters observed with this modification could be associated with the requirement of glutamate dehydrogenase for NADPH for which a major source is from alcoholic fermentation which is limited here by high aeration (in addition to pentose phosphate pathway). In line with that notion is the fact that KmJA.188 produced higher phenylalanine (coumaric acid) in flasks where aeration is poor than in the highly aerated bioreactors and that reducing bioreactor aeration from 2 to 1L/min increased production from 37mg/g to 42mg/g (figure 7A).

Supplementary materials

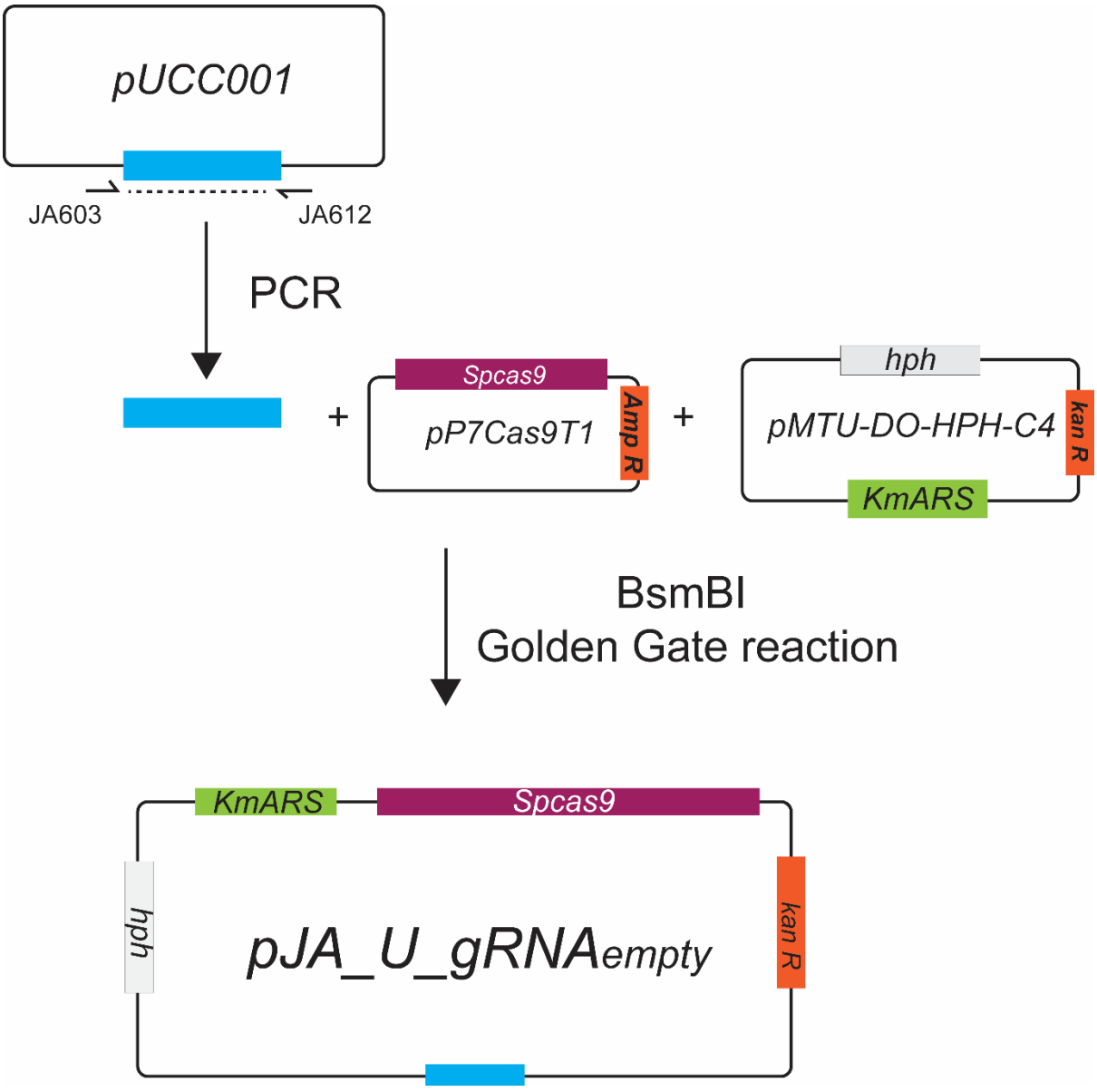


figure S1. Construction of high copy number CRISPR plasmid, pJA_U_gRNAempty. The empty single guide RNA transcriptional unit of pUCC001 was first amplified with primers JA603 and JA612 using Q5 High-Fidelity Polymerase (New England Biolabs, Ipswich, UK). The amplification product with 5' and 3' BsmBI Golden Gate overhangs was then purified and put together in a Golden Gate reaction with plasmid pP7Cas9T1 and pMTU-DO-HPH-C4 to generate the empty CRISPR plasmid pJA_U_*gRNAempty* from which *Streptococcus pyrogenes* Cas9 from the yeast toolkit[1] can be expressed under the control of the KmTEF1 promoter (T7) and KmINU1 terminator (T1). Plasmid pMTU-DO-HPH-C4 provided a backbone containing bacterial kanamycin-resistance gene, a yeast hygromycin-resistance gene and a high copy number replication origin in pJA_U_*gRNAempty*.

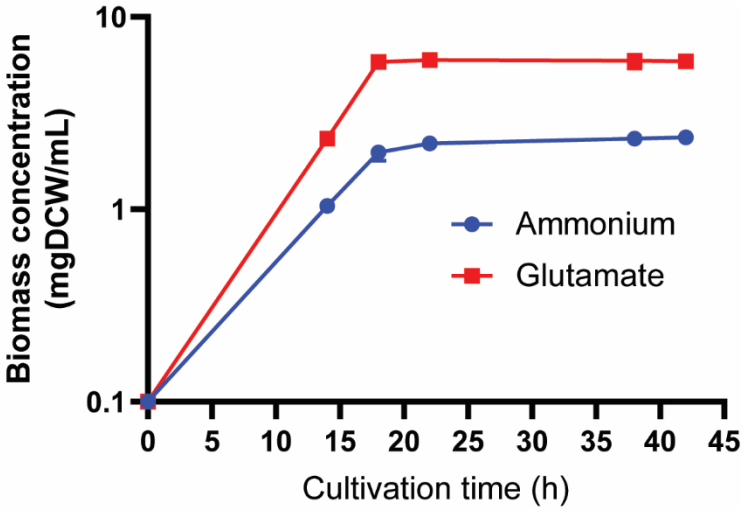


figure S2. Biomass formation of coumaric acid producer (KmJA.090) in ammonium sulphate and glutamate. Time interval biomass concentrations were determined for KmJA.090 cultivated in minimal glucose medium in batch bioreactor with ammonium sulphate or glutamate as sole nitrogen sources. Both nitrogen sources were added at final nitrogen concentrations of 0.5gN/L. Data shown are the mean ± standard deviation of three bioreactor replicates.

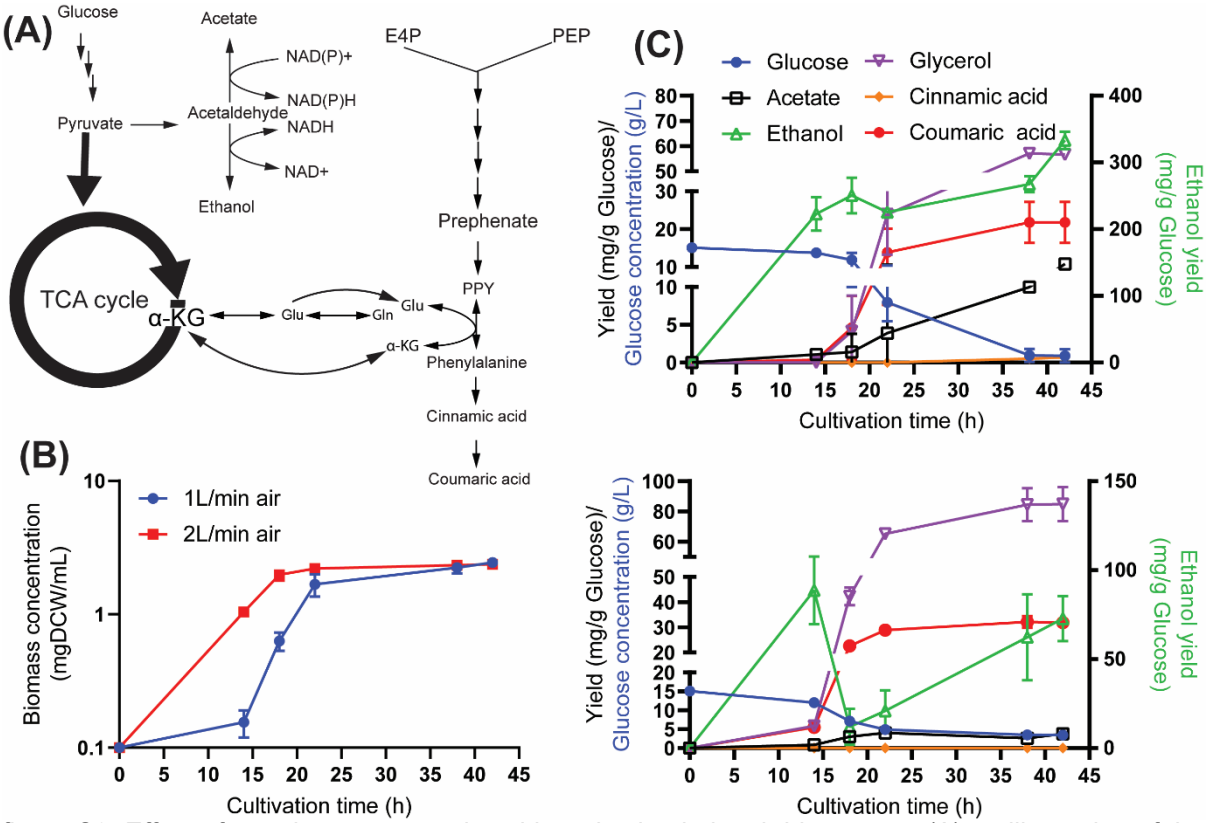


figure S3. Effect of aeration on coumaric acid production in batch bioreactor. (A) an illustration of the link between the TCA cycle, a major route for carbon metabolism in *K. marxianus*, and biosynthesis of coumaric acid precursor, phenylalanine: α-ketoglutarate from the TCA cycle is converted into glutamate which can serve as amino donor for aromatic amino transferase I reaction to produce phenylalanine from phenylpyruvate (PPY). Therefore, oxygen limitation due to insufficient aeration, which limits flux through the TCA cycle may also limit intracellular glutamate abundance, thereby capping phenylalanine biosynthesis. Glucose metabolism via fermentation (the pyruvate decarboxylase route) also increases in oxygen limitation, directing glucose toward ethanol and acetate production and potentially limiting the yield of coumaric acid on glucose. KmJA.090 was cultivated in minimal glucose medium batch

bioreactors to study the effect of aeration on its (B) growth (biomass production) profile; and (C) metabolites yield on glucose (top panel: 1L/min; bottom panel: 2L/min). Data shown are the mean ± standard deviation from three bioreactor replicates.

Table S1. Table S1. Non-yeast native genes expressed in this study.

Gene name	Source organism	GenBank accession number
AtPAL1	*Arabidopsis thaliana*	NM_129260
RgPTAL	*Rhodotorula glutinis*	KF770992
AtC4H	*Arabidopsis thaliana*	AY065145
PlPAL	*Photorhabdus luminescens*	BX57186
RcTAL	*Rhodobacter capsulatus*	CP001312
AtCPR1	*Arabidopsis thaliana*	NM_118585
encP	*Streptomyces maritimus*	AF254925

Table S2. Plasmids used in this study.

Name	**Description**	**Source**
Plasmids used as templates from which genes were amplified		
pUDE172	PCR template for *AtC4H* and *AtCPR1*	(Koopman et al., 2012)
pUDI069	PCR template for *AtTAL*	(Koopman et al., 2012)
MGV14-optpluPAL.pRS72N	PCR template for *PlPAL*	(Gottardi et al., 2017)
pUCC001	PCR template for gRNA expression system of pJA_U_*gRNAempty*	(Varela *et al.*, 2019)

Level I plasmids		
JA_NM L1.1	Level I plasmid of *Arabidopsis thaliana PAL1* codon optimised for *S. cerevisiae*	This work
JA_NM L1. 2	Level I plasmid of *Rhodotorula glutinis PTAL* codon optimised for *S. cerevisiae*	This work
JA_NM L1.6	Level I plasmid of *Arabidopsis thaliana C4H* codon optimised for *S. cerevisiae*	This work
JA_NM L1.7	Level I plasmid of *Arabidopsis thaliana TAL* codon optimised for *S. cerevisiae*	This work
JA_NM L1.9	Level I plasmid *Photorhabdus luminescens PAL* codon optimised for *S. cerevisiae*	This work
JA_NM L1.10	Level I plasmid of *Arabidopsis thaliana CPR1* codon optimised for *S. cerevisiae*	This work

JA_NM L1.11	Level I plasmid of *Streptomyces maritimus* EncP codon optimised for *S. cerevisiae*	This work
JA_NM L1.29	Level I plasmid of *K. marxianus GDH1*	This work
JA_NM L1.31	Level I plasmid of *K. marxianus GLN1*	This work
JA_NM L1.32	Level I plasmid of *K. marxianus GLT1*	This work
JA_NM L1.33	Level I plasmid of *S. cerevisiae GDH1*	This work
JA_NM L1.34	Level I plasmid of *S. cerevisiae GDH3*	This work

Level-II Dropout plasmids		
JA_LII RFP-DO 1	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 1 of level III (multigene) plasmids	This work
JA_LII RFP-DO 2	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 2 of level III (multigene) plasmids	This work
JA_LII RFP-DO 3	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 3 of level III (multigene) plasmids	This work
JA_LII RFP-DO 4	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 4 of level III (multigene) plasmids	This work
JA_LII RFP-DO 5	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 5 of level III (multigene) plasmids	This work
JA_LII RFP-DO 6	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 6 of level III (multigene) plasmids	This work
JA_LII RFP-DO 7	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 2 of level III (multigene) plasmids	This work
JA_LII RFP-DO 8	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 3 of level III (multigene) plasmids	This work
JA_LII RFP-DO 9	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 4 of level III (multigene) plasmids	This work
JA_LII RFP-DO 10	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 5 of level III (multigene) plasmids	This work

Level-II plasmids		
JA_NM L2. 18	JA_LII RFP-DO 1 with insert *KmPDC1pr-PlPAL-ScPGK1t*	This work
JA_NM L2. 20	JA_LII RFP-DO 2 with insert *KmENO1pr-AtC4H-KmPGK1t*	This work
JA_NM L2. 21	JA_LII RFP-DO 8 with insert *KmENO1pr-AtCPR1-KmPGK1t*	This work

JA_NM L2. 55	JA_LII RFP-DO 7 with insert *KmENO1pr-AtC4H-KmPGK1t*	This work
JA_NM L2. 74	JA_LII RFP-DO 9 with insert *KmPDC1pr-KmGDH1-ScPGK1t*	This work
JA_NM L2. 79	JA_LII RFP-DO 9 with insert *KmPDC1pr-ScGDH1-ScPGK1t*	This work
JA_NM L2. 80	JA_LII RFP-DO 9 with insert *KmPDC1pr-ScGDH3-ScPGK1t*	This work
JA_NM L2. 76	JA_LII RFP-DO 9 with insert *KmPDC1pr-ScGLN1-ScPGK1t*	This work
JA_NM L2. 84	JA_LII RFP-DO 4 with insert *KmPDC1pr-KmGLT1-ScPGK1t*	This work
JA_NM L2. 78	JA_LII RFP-DO 10 with insert *KmPDC1pr-KmGLN1-ScPGK1t*	This work
JA_NM L2. 85	JA_LII RFP-DO 5 with insert *KmPDC1pr-KmGLN1-ScPGK1t*	This work
JA_NM L2. 86	JA_LII RFP-DO 6 with insert *KmPDC1pr-ScGDH1-ScPGK1t*	This work

MTU-DO plasmids		
pI3 MTU-DO-LEU	Integrative plasmid with GFP dropout targeting locus between *KmSWF1* and *KmARO1*; *ScLEU2* marker	Addgene ID. 160176
pI4 MTU-DO-HPH	Integrative plasmid with GFP dropout targeting locus downstream of *KmHSP104*; HphMX marker	Addgene ID. 160216
pI6-MTU-DO-URA	Integrative plasmid with GFP dropout targeting the ARO3 locus; *ScLEU2* marker	Addgene ID. 160030
pI7 MTU-DO-LEU	Integrative plasmid with GFP dropout targeting and disrupting the *KmARO10* gene locus; *ScLEU2* marker	This work

Plasmids for construction of empty CRIPR plasmid pJA_U_*gRNA$_{empty}$* and CRIPSR plasmids used for gene deletions		
pUCC001	*AaTEF1pr-SpCas9-ScPHO5t ScTDH3pr-HH-gRNA$_{empty}$-HDV-ScCYC1t hphMX*	Addgene ID. 124451
pMTU-DO-HPH-C4	High-copy expression vector for *K. marxianus*, hygromycin selection	Addgene ID. 160343
pKmK.P7	Level I plasmid of *K. marxianus TEF1pr*	Addgene ID. 125040
pKmK.T1	Level I plasmid of *K. marxianus INU1t*	Addgene ID. 125053
pYTK036	Level I plasmid of *SpCas9*	Addgene ID. 65143
pTU-DO-URA	plasmid with GFP dropout with yeast *ScURA3* and bacterial *ampR* markers	This work
pP7Cas9T1	pTU-DO-URA with insert *KmTEF1pr-SpCas9-INU1t*	This work
pJA_U_*gRNAempty*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{empty}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_U_*KmPDC5*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmPDC5}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_U_*KmARO10*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmARO10}$-HDV-ScCYC1t hphMX- KanR*	This work
pJA_U_*KmARO9*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmAro9}$-HDV-ScCYC1t hphMX- KanR*	This work

Table S3. Primers used in this study.

Name	**Sequence**	**Purpose**	**Source**
JA603	CGTCTCACCAAAAGGGAGTTAGAATCATTTTGAATAAAAAA	Forward primers for amplifying empty single guide RNA transcriptional unit from plasmid pUCC001; also used for confirming correct protospacer sequence insertion with the antisense oligonucleotide of each gRNA protospacer sequence.	This work

JA612	CGTCTCATGCTCCAAAACCTTCTCAAGCAAGG	Reverse primers for amplifying empty single guide RNA transcriptional unit from plasmid pUCC001	This work
ASR_K001F	TTTGCTGGCCTTTTGCTC	TTTGCTGGCCTTTTGCTC	Rajkumar et al., 2019
ASR_K002R	CATCTGGATTTGTTCAGAACG	Colony PCR reverse primer, priming downstream of the BsmBI cloning site in YTK001; used for confirmation of plasmid assembly and sequencing of level I plasmid inserts	Rajkumar et al., 2019
ASR_K001R	ATTGGTAACTGTCAGACCAAGTTTA	Colony PCR reverse primer, priming at the 5' end of AmpR; used for confirmation of level II plasmid assembly	Rajkumar et al., 2019
ASR_K007R	CAGTCATCGGTATGATCTG	Colony PCR reverse primer, priming in ConRE; used for confirmation of I7/ARO10 MTU-DO plasmid assembly	Rajkumar et al., 2019
JA198	CTTGAATAGTCATTTGAGCAGCACCATCACCTTCACACA AGATCAACTTTGGAACATAGTCTGCTGGGACAGCAGATT GGTTGATAATGTGAGATTTCTAGCAAAGGCAA	Forwards primer of overlapping primer set for producing *KmARO10* repair fragment	This work
JA198	CTTGAATAGTCATTTGAGCAGCACCATCACCTTCACACA AGATCAACTTTGGAACATAGTCTGCTGGGACAGCAGATT GGTTGATAATGTGAGATTTCTAGCAAAGGCAA	Forwards primer of overlapping primer set for producing *KmARO10* repair fragment	This work
JA199	TCACTCTTGGTCGTTATGTGTTCGAGAGATTGCTCAACT GTGGTTCCAAGACCATTTTCGGTGTTCCAGGTGACTTCA ACTTGCCTTTGCTAGAAATCTCACATTATCAA	Reverse primer of overlapping primer set for producing KmARO10 repair fragment	This work
JA100	GATGTTCCTATTCATGAGATTCTTCCTCAGCTTATAGCAA AAATTGACGCATCAAAGCTCTCTAATCTGTACGCCGTCA CAGTTCCCGATATGGTGGGGAAGCTGGACAA	Forwards primer of overlapping primer set for producing *KmPDC5* repair fragment	This work
JA101	AACCTTGAGTGAGAATGGTACTCAATTCTTGGAAAGTGA ATTGTAGCGAACCATCACCTACGATCAAAATGATCCGAT GCTTGTCCAGCTTCCCCACCATATCGGGAACT	Reverse primer of overlapping primer set for producing *KmPDC5* repair fragment	This work
JA061	ATCAAAGTTTTGAACGCTATTTCAGAAGCACGCCCTTCT ATCAAGTTTAATTTTGAACATCATTTGATTGGTGGTGCTG CAATTGATGCTACTGGTGGCATCATCTAGAT	Forwards primer of overlapping primer set for producing *KmLEU2* repair fragment	Work in chapter 2
JA062	AGCGGAATCGATCAATTGATGTTGGACAGTCAATTGAGG GAACTCATTCTTGATGGTTTCTTCTACTGTTTTTCTCCAC AATCTAGATGATGCCACCAGTAGCATCAATT	Reverse primer of overlapping primer set for producing KmLEU2 repair fragment	Work in chapter 2
JA629	AACCACGCCTTGACATGGTCGCCCTCAAACGCATGGGA CCATTTCTCAATTATATTGGCGGTCATTGCCTGGGAAAT CCCGTGAAAACTTTTGAC	Forwards primer of overlapping primer set for producing *KmARO9* repair fragment	This work
JA630	TATTTGAACAAGCCTGCTTATGATAATTGGGACATCATG ATGGCCAACGGTTCTTCTGACTCTTTGTCAAAAGTTTTCA CGGGATTTCCCAGGCA	Reverse primer of overlapping primer set for producing *KmARO9* repair fragment	This work
JA204	AGTGTAACCATTGTTGTTGAA	Forward diagnostic primer to amplify and confirm deletion at KmARO10 locus	This work
JA205	CCAGTAGTTCTAGACGATAAATC	Reverse diagnostic primer to amplify and confirm deletion at *KmARO10* locus	This work
JA102	GTATGCTGATAGGTACAACTTGG	Forward diagnostic primer to amplify and confirm deletion at *KmPDC5* locus	This work
JA103	CTAAAGTCTTACTCTTTTGCTTTGAGG	Reverse diagnostic primer to amplify and confirm deletion at *KmPDC5* locus	This work

ASR_J016CF	TCAAAGAATATCGTTGTCTTACCAG	Forward diagnostic primer to amplify and confirm deletion at *KmLEU2* locus	Rajkumar et al., 2019
ASR_J016CR	TGCCAAAATCTCCTTTACAGC	Reverse diagnostic primer to amplify and confirm deletion at KmLEU2 locus	Rajkumar et al., 2019
JA277	GCATCGTCTCATCGGTCTCATATGTCAGAGCCAGAATTTCAACAAGCTTACGAAGAAGTTGTCTCC	Forward primer for amplifying *S. cerevisiae GDH1* from genomic DNA	This work
JA278	ATGCCGTCTCAGGTCTCAGGATTTAAAATACATCACCTTGGTCAAACATAGCATCAGAAACCTTGATGAAACTTGCGAT	Reverse primer for amplifying *S. cerevisiae GDH1* from genomic DNA	This work
JA279	GCATCGTCTCATCGGTCTCATATGTCATCCGAACCATACGAACC	Forward primer for amplifying *K. marxianus GDH1* from genomic DNA	This work
JA290	ATGCCGTCTCAGGTCTCAGGATTTAGAACACGTCACCTTGATCTAACATG	Reverse primer for amplifying *K. marxianus GDH1* from genomic DNA	This work
JA283	GCATCGTCTCATCGGTCTCATATGACAAGCGAACCAGAGTT	Forward primer for amplifying *S. cerevsiae GDH3* from genomic DNA	This work
JA284	CACGTCTCATTTCAGAGATGATGCACCCCT	Reverse primer for eliminating an internal BsmBI site from *S. cerevsiae GDH3*; used with JA283	This work
JA285	TTCGTCTCAGAAACGGGCATTACTTCTG	Forwards primer for eliminating an internal BsmBI site from *S. cerevsiae GDH3*; used with JA286	This work
JA286	ATGCCGTCTCAGGTCTCAGGATCTAAAAAACATCTCCCTGGTCAAGCATTGC	Reverse primer for amplifying *S. cerevsiae GDH3* from genomic DNA	This work
JA287	GCATCGTCTCATCGGTCTCATATGTCCGATATCGTTGAAAAGACC	Forward primer for amplifying *K. marxianus GLN1* from genomic DNA	This work
JA288	CACGTCTCAAAACGTTGGTGTGACAACC	Reverse primer for eliminating an internal BsmBI site from *K. marxianus GLN1*; used with JA287	This work
JA289	TTCGTCTCAGTTTCTACCAAGGAAATGAGAGCT	Forwards primer for eliminating an internal BsmBI site from *K. marxianus GLN1*; used with JA290	This work
JA290	ATGCCGTCTCAGGTCTCAGGATTTACAAAGATTCTCTTTCGTATTCCTTAGTC	Reverse primer for amplifying *K. marxianus GLN1* from genomic DNA	This work
JA069	GAACGTGAGACCAGTCCCTATCAGT	Forward primers for amplifying bacteria expressible red fluorescent protein transcriptional unit from pYTK083	This work
JA070	CCAGCTGAGACCGACTTATAAACGCAG	Reverse primers for amplifying bacteria expressible red fluorescent protein transcriptional unit from pYTK083	This work
ASR_I007LF_MTU	GCATCGTCTCATCGGTCTCACCCTTTCAGGCGCGCCACTCGAAGTTAAATCCATTATTCAGA	Forward primer for amplifying the left homology arm of I7/*KmARO10* integration site (chromosome II: 1213310..1214060	This work
ASR_I007LR_MTU	ATGCAGGTCTCACGTTCGTCTCATCAGTCTAGATGCGAATTCTATTCCTTCTGAGTTGGATGCTAT	Reverse primer for amplifying the left homology arm of I7/*KmARO10* integration site (chromosome II: 1213310..1214060)	This work
ASR_I007RFbis	GCATCGTCTCATCGGTCTCAGAGTTGTTAATTATATTTGATTCTTATAGATTTAGGTAC	Forward primer for amplifying the right homology arm of I7/*KmARO10* integration site (chromosome II: 1215903..1216649)	This work
ASR_I007RRbis	ATGCCGTCTCAGGTCTCATCGGCTGAGGCGCGCCAATTCCTGGCGTCACATT	Reverse primer for amplifying the right homology arm of I7/*KmARO10* integration site (chromosome II: 1215903..1216649)	This work

ASR_I7_US_F	GATTTTCACAACGAAATCCAACAT	Forward primer for checking integration at I7/*ARO10* integration site; primes 95bp upstream of the site.	This work
ASR_I7_DS_R	ACACCGACACTTGTCCAA	Reverse primer for checking integration at I7/*ARO10* integration site; primes 103bp upstream of the site.	This work
ASR_I1US_F	CATTAGAACCTTTTTCAACACTC	Forward primer for checking integration at I1/LAC4 integration site; primes 92bp upstream of the site.	Rajkumar and Morrissey 2020
ASR_I1DS_R	CTTAGTGGTTGTGAAGGTTT	Reverse primer for checking integration at I1/LAC4 integration site; primes 90bp downstream of the site.	Rajkumar and Morrissey 2020
ASR_I3_US_F	ATTCCAGACACAGGTGGAA	Forward primer for checking integration at I3 integration site; primes 113bp upstream of the site.	Rajkumar and Morrissey 2020
ASR_I3_DS_R	AAAGGAAGGATTCGGCGTTA	Reverse primer for checking integration at I3 integration site; primes 114bp downstream of the site.	Rajkumar and Morrissey 2020
ASR_I5_US_F	AGTAGTGAGTGACAGACAC	Forward primer for checking integration at I4 integration site; primes 62bp upstream of the site.	Rajkumar and Morrissey 2020
ASR_I5_DS_R	GCAGTTTCTGTGCAGAAAA	Reverse primer for checking integration at I4 integration site; primes 102bp downstream of the site.	Rajkumar and Morrissey 2020
ASR_I6_US_F	AGCCTGTCTACGTACAAC	Forward primer for checking integration at I6/ARO3 integration site; primes 107bp upstream of the site.	Rajkumar and Morrissey 2020
ASR_I6_DS_R	CGTTCCAAGGAAACCACTATA	Reverse primer for checking integration at I6/ARO3 integration site; primes 98bp upstream of the site.	Rajkumar and Morrissey 2020

Table S4. CRIPSR targets for gene deletions.

Gene	**Target sequence**
KmPDC5	GAATCCGGTACATCCGCGAT
KmARO10	GTCCAAAGAAGACTTAGCCA
KmARO9	TTGGATTCTGACCAGTTGGG

KmLEU2	TTGAAACCTGAGTACGCCAA

Table S5. Metabolite yields in this work.

Flask cultivation

Strain	Residual glucose (g/L)	Acetate (mg/g)	Ethanol (mg/g)	Glycerol (mg/g)	2-Phenylethanol (mg/g)	Tyrosol (mg/g)	Cinnamic acid (mg/g)	Coumaric acid (mg/g)
KmJA.073	0	43.1 ± 4.2	0	3.1 ± 1.4	1.4 ± 0.1	3.4 ± 0.2	0.3 ± 0.02	0
KmJA.074	0	36.3 ± 4.9	3.7 ± 2.61	0	16 ± 0.8	12.4 ± 0.02	0	0
KmJA.075	0	2.5 ± 0.1	132.7 ± 0.6	26 ± 1.1	25.6 ± 0.2	14.4 ± 0.2	1.3 ±0.02	0
KmJA.076	1 ± 0.1	2.2 ± 0.7	145.8 ± 19	51 ± 1.6	19 ± 0.2	15.6 ± 0.4	0	0.2 ± 0.1
KmJA.077	0	5.9 ± 0.003	200.9 ± 70.9	28.5 ± 0.8	17.2 ± 0.2	15.4 ± 1.8	0	0.1 ± 0.06
KmJA.078	0	26.6 ± 1.3	162.8 ± 3.6	31.2 ± 1.3	1.2 ± 0.1	3.3 ± 0.2	0	2.5 ± 0.2
KmJA.079	0	7.3 ± 3.5	146.3 ± 11.8	9.4 ± 3.6	17.7 ± 0.6	12.3 ± 0.24	0	1.5 ± 0.1
KmJA.080	0	9.2 ± 5.7	71.1 ± 6.6	3.8 ± 1.9	1.12 ± 0.01	4.13 ± 2.2	0	6.5 ± 0.11
KmJA.081	0	17 ± 4	151.7 ± 7.7	13.1 ± 1.8	1.5 ± 0.002	3.8 ± 0.03	0	0.013 ± 0.02
KmJA.082	0	20.5 ± 1	190.5 ± 5.7	22.5 ± 0.6	1.5 ± 0.03	3.6 ± 0.05	0	0.67 ± 0.02
KmJA.011	0	0	0	15.3 ± 0.7	38.18 ± 0.3	22.8 ± 0.02	0.16 ± 0.02	5 ± 0.13
KmJA.013	4.3 ± 0.1	0	0	85.4 ± 0.5	0	0	0.14 ± 0.01	7 ± 0.2
KmJA.014	3.8 ± 0.9	3 ± 0.3	25 ± 20.1	67.3 ± 2.5	1.3 ± 0.2	4.1 ± 0.4	3.2 ± 0.3	21.7 ± 1.9
KmJA.017	0	3.9 ± 4.5	1.2 ± 2	62.6 ± 6.6	0	0	0.9 ± 0.2	15.4 ± 0.4
KmJA.016	7 ± 0.1	7.2 ± 2.2	13.9 ± 3.5	84.7 ± 1.6	0	0	5.2 ± 0.04	0
KmJA.088	0	6.1 ± 0.4	5.9 ± 0.9	53.1 ± 1.2	0	0	0	19.2 ± 0.4
KmJA.90	0	0	21.4 ± 4.52	0	0	0	0	25 ± 0.4
KmJA.091	4.4 ± 0.4	2 ± 0.5	56.7 ± 8.7	63.7 ± 2.7	0	0	6.8 ± 0.1	0
KmJA.092	0	0	0	0	0	0	0	8.9 ± 0.14
KmJA.171	0	0	10.3 ± 1.8	20.9 ± 1	0	0	0	28.5 ± 0.2
KmJA.180	0.27 ± 0.2	9.2 ± 0.02	44 ± 1.8	48.1 ± 0.4	0	0	0	22.19
KmJA.182	0	0	0	0	0	0	0	0
KmJA.188	7.3 ± 0.2	13.9 ± 1.5	3.1 ± 1.3	94.6 ± 0.8	0	0	1.6 ± 0.1	48 ± 0.5
KmJA.189	0.17 ± 0.1	8.8 ± 1.6	87 ± 0.8	53.9 ± 0.9	0	0	0.2 ± 0.04	9.2 ± 0.1
KmJA.228	0	14.8 ± 1	141 ± 8.1	21.2 ± 2.4	0	0	1.4 ± 0.1	8.3 ± 0.4
KmJA.229	4.6 ± 0.8	8.6 ± 0.7	33.7 ± 7.3	66 ± 0.5	0	0	1.2 ± 0.1	30.1 ± 2.2
KmJA.249	0.02 ± 0.02	2.4 ± 1.4	124.8 ± 5.2	44 ± 0.9	0	0	0.3 ± 0.1	13.8 ± 0.2
KmJA.250	2.1 ± 0.4	0.4 ± 0.7	8.6 ± 3	93.5 ± 4	0	0	3 ± 0.5	11.5 ± 0.2

KmJA.090 in amino acids as nitrogen source

Nitrogen source	Residual glucose (g/L)	Acetate (mg/g)	Ethanol (mg/g)	Glycerol (mg/g)	2-Phenylethanol (mg/g)	Tyrosol (mg/g)	Cinnamic acid (mg/g)	Coumaric acid (mg/g)
Glutamate	0	0	0	0	0	0	8.2 ± 0.5	43 ± 0.3
Glutamine	0	0	0	0	0	0	1.2 ± 0.02	16.6 ± 0.4
Leucine	0.5 ± 0.2	0	61.4 ± 0.01	38.2 ± 4.14	0	0	3.01 ± 0.4	13.3 ± 0.5
Methionine	0.03 ± 0.04	0	0	2.5 ± 1.1	0	0	1.9 ± 0.02	17.3 ± 0.01

Bioreactor cultivation

KmJA.090 in ammonium sulphate with 1L/min aeration

Time	Residual glucose (g/L)	Acetate (mg/g)	Ethanol (mg/g)	Glycerol (mg/g)	2-Phenylethanol (mg/g)	Tyrosol (mg/g)	Cinnamic acid (mg/g)	Coumaric acid (mg/g)
0	15.1 ± 0.7	0	0	0	0	0	0	0
14	13.8 ± 1	1.1 ± 0.4	222.8 ± 25	0	0	0	0	0.4 ± 0.3
18	11.9 ± 1.9	1.4 ± 2.4	250.6 ± 27	4.1 ± 4.7	0	0	0	4.6 ± 4.2
22	7.9 ± 2.5	3.9 ± 6.7	225.3 ± 0.04	23.9 ±10.7	0	0	0	13.9 ± 6.3
38	1 ± 0.8	10 ± 0.6	267.3 ± 12.3	57.1 ± 1.3	0	0	0.5 ± 0.1	21.8 ± 5.4
42	0.9 ± 0.9	10.8 ± 0.6	333.1 ± 13.1	56.7 ± 1.8	0	0	0.7 ± 0.1	21.82 ± 5.4

KmJA.090 in ammonium sulphate with 2L/min aeration

Time	Residual glucose (g/L)	Acetate (mg/g)	Ethanol (mg/g)	Glycerol (mg/g)	2-Phenylethanol (mg/g)	Tyrosol (mg/g)	Cinnamic acid (mg/g)	Coumaric acid (mg/g)
0	15.1 ± 0.1	0	0	0	0	0	0	0
14	12 ± 0.3	0.9 ± 0.43	88.7 ± 19.1	5.9 ± 1.4	0	0	0	5.5 ± 0.6
18	7.2 ± 0.2	3.1 ±0.3	11.9 ± 10.1	42.2 ± 3.6	0	0	0.002 ± 0.003	22.6 ± 0.1
22	4.9 ± 0.3	4.1 ± 0.52	20.9 ± 11.6	65.3 ± 1.8	0	0	0.005 ± 0.005	28.9 ±1.3
38	3.5 ± 0.32	2.6 ± 1.4	62.32 ± 24.2	84.4 ± 10.9	0	0	0.005 ± 0.005	32.2 ± 2.4
42	3.5 ± 0.3	3.8 ± 1.3	72.8 ± 12.7	85 ± 11.3	0	0	0.006 ± 0.006	31.9 ± 2.2

KmJA.090 in Glutamate with 2L/min aeration

Time	Residual glucose (g/L)	Acetate (mg/g)	Ethanol (mg/g)	Glycerol (mg/g)	2-Phenylethanol (mg/g)	Tyrosol (mg/g)	Cinnamic acid (mg/g)	Coumaric acid (mg/g)
0	14.23 ± 0.9	0	0	0	0	0	0	0
14	8.8 ± 0.6	33.6	23.5 ± 0.6	13.8 ± 1.3	0	0	1.2 ± 1	12.4 ± 1.7
18	0.1 ±0.1	8.3	11.5 ± 2.9	20.5 ± 5.9	0	0	0	38.3 ± 0.7
22	0.02 ±0.02	0.5 ± 0.9	1 ± 1.7	17.3 ± 3.1	0	0	0	41 ± 1
38	0.01 ± 0.01	0	0	15.4 ± 0.9	0	0	0	42.1 ± 1
42	0.01 ± 0.01	0	0	12.9 ± 0.5	0	0	0	42.38 ± 1.2

KmJA.188 in ammonium sulphate with 1L/min aeration

Time	Residual glucose (g/L)	Acetate (mg/g)	Ethanol (mg/g)	Glycerol (mg/g)	2-Phenylethanol (mg/g)	Tyrosol (mg/g)	Cinnamic acid (mg/g)	Coumaric acid (mg/g)
0	13.8 ± 0.3	0	0	0	0	0	0	0
14	9.9 ± 0.6	0	126.2 ± 44.9	18.5 ± 3.3	0	0	0	14.9 ± 3.2
18	8.4 ± 0.54	0	88.5 ± 29.3	27.7 ± 4.7	0	0	0	23.3 ± 3.4
22	7.2 ± 0.9	0	6.1 ± 1.6	40.5 ± 5.9	0	0	0	30.2 ± 3.2
38	4.3 ± 0.9	0	6.2 ± 4.1	73.4 ± 2.1	0	0	1.2 ± 0.2	41.1 ± 1.3
42	4.2 ± 0.3	0	60.2 ± 24	75.3 ± 0.8	0	0	1.6 ± 0.3	41.5 ± 1.26

KmJA.188 in ammonium sulphate with 2L/min aeration

Time	Residual glucose (g/L)	Acetate (mg/g)	Ethanol (mg/g)	Glycerol (mg/g)	2-Phenylethanol (mg/g)	Tyrosol (mg/g)	Cinnamic acid (mg/g)	Coumaric acid (mg/g)
0	14.43 ± 0.5	0	0	0	0	0	0	0
14	11.8 ± 0.5	0	0	8.8 ± 2.9	0	0	0	7.2 ± 1.8
18	10.5 ± 0.6	0	0	20 ± 3.2	0	0	0	13.1 ± 0.6
22	9.5 ± 0.6	0	0.3 ± 0.5	33.9 ± 3.8	0	0	0	20.8 ± 3.4
38	6.8 ± 0.8	0	3.2 ± 5.6	67.5 ± 8	0	0	1.7 ± 0.5	36.3 ± 3.2
42	6.4 ± 1	0	13.6 ± 0.7	60.9 ± 1.7	0	0	2.3 ± 0.5	37.1 ± 3.1

Chapter 4. Enzymatic and functional characterisations reveal the contributions of *Kluyveromyces marxianus* 2-oxo acid decarboxylase genes to glycolytic and Ehrlich pathways in the yeast

Joel A. Akinola, Andrea Rizzo, Jean-Marc Daran, John P. Morrissey.

This chapter has been prepared for submission to the journal Yeast.

JAA constructed all deletion strains, interpreted the data and prepared the manuscript

AR assisted in the characterisation of deletion strains as part of a masters thesis

JMD provided supervision during the completion of enzyme assays at Delft University of Technology, The Netherlands

JPM provided supervision and edited the manuscript

Abstract

The decarboxylation of 2-oxo acids by 2-oxo acid decarboxylases is an important step in sugar metabolism via glycolysis and in amino acids metabolism via the Ehrlich pathway. Although the genes encoding 2-oxo acid decarboxylases in Kluyveromyces yeasts are known, they have not been comprehensively characterised. Through bioinformatic analysis and *in vitro* enzyme assays, we identified two 2-oxo acid decarboxylases in *K. marxianus* as *PDC1* and *ARO10.* Pdc1 showed activities for linear 2-oxo acid substrates (pyruvate and 2-oxo butanoate) and the disruption of *PDC1* eradicated endogenous pyruvate decarboxylase activity and reduced glucose consumption and biomass formation. Aro10 was specific for 2-oxo acid substrates that are derived from aromatic amino acid phenylalanine (phenylpyruvate) and from the branched chain amino acid valine (ketoisovalarate). Unlike the *S. cerevisiae* Aro10, *K. marxianus* Aro10 showed some pyruvate decarboxylase activity and Pdc1 displayed some phenylpyruvate decarboxylase activity. Although *ARO10* contributed majorly to endogenous phenylpyruvate decarboxylase activity during *K. marxianus* growth in phenylalanine, its disruption was unable to eradicate the activity until *PDC1* was additionally disrupted. A third gene which was previously regarded as *PDC5* is homologous to *S. cerevisiae THI3.* The product of this gene showed no 2-oxo acid decarboxylase activity, but its disruption reduced the transcription of thiamine biosynthetic genes (*THI4* and *THI11*), confirming that this gene is the *THI3* of *K. marxianus.* This study identified the genes encoding the mediators of crucial steps in the glycolytic and Ehrlich pathways and revealed their functional contributions to the pathways, furthering our understanding of how these genes contribute to sugar and amino acid metabolism. Specifically, it is anticipated that knowing the contributions of *PDC1* and *ARO10* to pyruvate and phenylpyruvate decarboxylase activities in *K. marxianus* will provide information regarding the disruption of these genes in order to prevent precursor degradation and improve supply in engineering this yeast for the production of aromatic molecules from aromatic amino acid precursors.

Keywords

Pyruvate decarboxylase, 2-oxo acids decarboxylase, Thiamine biosynthesis, *Kluyveromyces marxianus,* Ehrlich pathway, Glycolysis, Amino acids degradation, Aromatic molecules.

Take Away

- *Kluyveromyces marxianus* has only *PDC1* and *ARO10* as 2-oxo acid decarboxylase genes, with Pdc1 and Aro10 showing varying levels of activity for all linear, aromatic, branched chain, and sulphur-containing 2-oxo acid substrates tested.

- Phenylpyruvate decarboxylase Aro10 displays some activities for the linear 2-oxo acids, pyruvate and 2-oxo butanoate but does not contribute to endogenous PDC activity during growth on ammonium.
- PDC-null *K. marxianus* shows growth defect on glucose.
- The disruption of a gene previously designated as *PDC5* in *K. marxianus* reduced the transcription of thiamine biosynthetic genes, indicating that it is actually a *THI3.*

Introduction

Two oxo-acid decarboxylases (2ODCs) are thiamine pyrophosphate (TPP)-dependent decarboxylases that catalyse the decarboxylation of 2-oxo acids (2OAs) into aldehydes (figure 1, reactions 2 and 3). The 2OA substrates of these enzymes include pyruvate which is formed in the metabolism of sugars as fermentable carbon sources via the glycolytic pathway and the degradation products arising from the catabolic deamination of amino acids in the Ehrlich pathway. The decarboxylation of 2OAs is necessary for driving the assimilation of sugars and amino acids in these pathways (Hazelwood *et al.*, 2008). Thus, 2ODCs are important in yeast metabolism as their absence can limit proper sugar and amino acid metabolism. The aldehyde products of 2ODCs are further oxidised or reduced into carboxylic acids or alcohols by aldehyde and alcohol dehydrogenases (reactions 4 and 5), depending on the intracellular redox status (Hazelwood *et al.*, 2008). Pyruvate decarboxylases which are involved in central carbon metabolism via the glycolytic pathway are examples of 2ODCs that are so called because they take pyruvate, a 2OA, as substrate in order to produce acetaldehyde (reaction 3). Amino acids-derived 2OAs are decarboxylated into fusel aldehydes (reaction 2) that are further converted into fusel acids and alcohols some of which are applicable as flavour & fragrances and as biofuels that can replace bioethanol due to their superior physicochemical properties (Atsumi, Hanai and Liao, 2008). Thus, kinetically efficient 2ODCs are industrially attractive and because these compounds contribute majorly to the flavour properties of fermented foods and beverages, the characterisation of *K. marxianus* 2ODCs as has been extensively done for *S. cerevisiae* (Romagnoli *et al.*, 2012) can foster its application for the fermentation of foods and beverages.

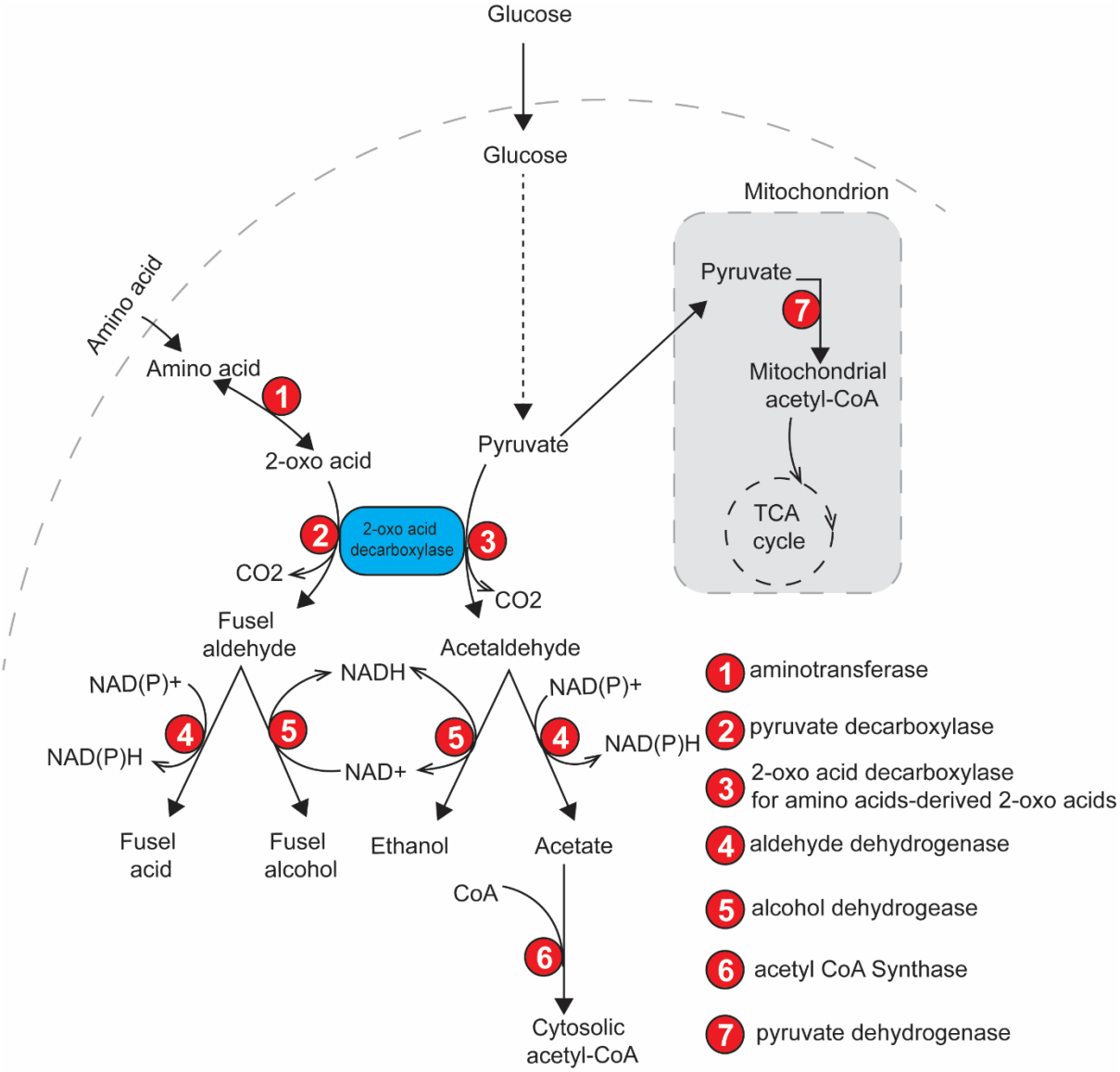


Figure 1. Role of 2-oxo acid decarboxylases in glycolytic and Ehrlich pathways. In the glycolytic pathway, glucose is converted into pyruvate which is a 2-oxo acid, while amino acids are deaminated by aminotransferases into their 2-oxo acid substrates. The pyruvate and amino acid-derived substrates are further converted into acetaldehyde or fusel aldehyde, respectively, through their decarboxylation by 2-oxo acid decarboxylases. Acetaldehyde is further oxidised into acetate by aldehyde dehydrogenase or reduced into ethanol by alcohol dehydrogenase. These enzymes can also take fusel aldehydes as substrates in order to produce fusel acids and alcohols. The oxidoreduction reactions are also important for achieving redox balance which can become compromised by the absence of 2-oxo acid decarboxylases. Acetate in the cytosol is used for the biosynthesis of cytosolic acetyl-CoA which is important for proper cell function. Pyruvate can alternatively enter the TCA cycle where it is used to generate mitochondrial acetyl-CoA and TCA cycle intermediates.

S. cerevisiae has four 2ODC genes, the products of which have been enzymatically characterised (Romagnoli *et al.*, 2012). These enzymes are interesting in that they decarboxylate a broad and overlapping array of 2OA substrates. Thus, when a gene, the

product of which mainly catalyses the decarboxylation of a substrate is disrupted, other gene(s) are transcriptionally upregulated in order to provide enzymes that will, although with poorer kinetics, take on the catalytic role of that gene product (Seeboth, Bohnsack and Hollenberg, 1990; Vuralhan *et al.*, 2005). Therefore, the disruption of a 2ODC gene may not guarantee the eradication of an enzyme activity that is associated with it, meaning that the characterisation of these genes requires the complete elimination of 2ODC activities so as to express and characterise each gene individually.

Previous studies identified three putative 2ODCs in *K. marxianus* and named the genes encoding them as *KmPDC1, KmPDC5* and *KmARO10* (Choo *et al.*, 2018; Hassing *et al.*, 2021). Of these genes, only the product of *KmPDC1* was demonstrated to have pyruvate decarboxylase (PDC) activity (Choo *et al.*, 2018). Evidence showed that disrupting *PDC1* eradicates PDC activity in *K. marxianus* and *K. lactis,* when grown in glucose medium. This shows that Pdc1 is the main pyruvate decarboxylase enzyme in both yeasts. PDC activity drives the assimilation of fermentable carbon sources by metabolising pyruvate into acetaldehyde from which ethanol and acetate are produced (reactions 3, 4, 5). These reactions are required for co-factor reoxidation and cytosolic acetyl-CoA supply in *S. cerevisiae* (Pronk, Steensma and Van Dijken, 1996). The importance of PDC activity is illustrated by the fact that PDC-null *S. cerevisiae* is unable to grow on glucose and this phenotype is linked to the lack of cytosolic acetyl-CoA (Flikweert *et al.*, 1999). PDC-null *K. marxianus* and *K. lactis*, on the other hand, maintain their ability to grow on fermentable carbon sources and the question of how these yeasts circumvent the need for PDC activity remains largely unanswered. Well known alternative routes for cytosolic acetyl-CoA supply include its production in other compartments such as the mitochondria, peroxisomes, and nucleus. However, the inability of activated acyl groups to freely move across intracellular membranes means that special strategies are required to transport acetyl-CoA from these cellular compartments to the cytosol (Van Roermund *et al.*, 1995). This can be mediated by transport proteins or by deactivation into acyl group for free membrane crossing and then reactivation in the cytosol (Chen, Siewers and Nielsen, 2012). Given that these yeasts are able to efficiently metabolise sugars through respiration (Kiers *et al.*, 1998; Sakihama *et al.*, 2019), unlike *S. cerevisiae*, the source of cytosolic acetyl-CoA is suspected to be the mitochondrion but the specific mediators involved are yet to be identified. A prerequisite for identifying these mediators is a comprehensive characterisation of the yeast's 2ODC genes in order to improve our understanding of their roles.

Contrary to *ScPDC5, KmPDC5* does not contribute to endogenous PDC activity (Choo *et al.*, 2018) nor has any 2ODC activity been reported for its product. This protein is reminiscent of Thi3 which is a regulatory protein that is involved in the transcriptional regulation of thiamine

biosynthesis. Despite its high sequence homology with 2ODCs, Thi3 has also not been demonstrated to show any enzyme activity but is instead involved in the regulation of thiamine biosynthetic genes in thiamine depleted medium towards thiamine biosynthesis (Nosaka, 2006).

In the current study, the functions of the three putative *K. marxianus* 2ODC genes were elucidated. Enzyme kinetics (V_{max} and K_m values) for pyruvate and 2OAs generated from linear, aromatic, branched chain and sulphur-containing amino acids were investigated. Finally, the functional roles of these genes during *K. marxianus* growth in the assimilation of glucose and aromatic amino acids were tested.

Materials and Methods

1. Strains and cultivation

All yeast strains used in this study are listed on Table 1. Yeasts were routinely cultured at 30°C in YPD/E broth (2% yeast extract, 1% peptone and 2% glucose or ethanol; Formedium, Hunstanton, UK). For growth analysis, yeasts were cultivated in minimal medium (3g/L KH2PO4, 0.5g/L MgSO4.7H2O, appropriate nitrogen source, with vitamins and trace elements, pH5.5) with 2% ethanol as carbon source for preculturing and 15-20g/L glucose for the actual analysis. All growth analysis were done in 50mL cultures in 250mL Erlenmeyer flasks. Solid media for yeast growth were supplemented with 2% agar (Formedium). *E. coli* DH5α grown in LB broth (1% NaCl, 1% peptone, 0.5% yeast extract; Formedium) or LB supplemented with 1.5% agar and the appropriate antibiotic (100mg/L ampicillin, 50mg/L chloramphenicol, or 50mg/L kanamycin; Fisher Scientific (Dublin, Ireland)) were used for bacterial transformations.

2. *In silico* identification of *K. marxianus* 2-oxo acid decarboxylase genes

Homologous sequences from budding yeasts were recovered and sequence alignment was performed using the MUSCLE algorithm (Edgar, 2004). For this, hits in the genomes of *K. marxianus, K. lactis, Lachancea kluyveri, Lachancea waltii, Lachancea thermotolerans, Eremothecium gossypii*, *Zygosaccharomyces rouxii, Torulaspora delbrueckii*, *S. cerevisiae, Naumovozyma castellii Naumovozyma dairensis, Kazachstania naganishii, Kazachstania africana, Tetrapisispora blattae* and *Tetrapisispora phaffii* were considered. Phylogenetic analysis was performed by neighbour-joining method with bootstrap values inferred from 600 replicates using MEGAX (Kumar *et al.*, 2018). Synteny and evolutionary conservation were assessed on the Yeast Gene Browser (YGOB, http://ygob.ucd.ie/) (Byrne and Wolfe, 2005).

Table 1. Strains used in this study.

Strain	Genotype/ Description	Source
NBRC1777	Wildtype *K. marxianus*	NITE Biological Resource Centre
CBS2359	Wildtype *K. lactis*	Westerdijk Institute
KmASR.005	NBRC1777; *dln4*	Rajkumar et al, 2019
KmJA.001	KmASR.005; *aro10*	This work
KmJA.002	KmASR.005; *pdc1*	This work
KmJA.051	KmASR.005; *thi3*	This work
KmJA.005	KmASR.005; *aro10pdc1*	This work
KmJA.164	KmASR.005; *thi3aro10pdc1*	This work
KmJA.147	KmASR.005; *thi3aro10pdc1/* small colonies after KmJA.008 was grown on glucose	This work
KmJA.148	*thi3aro10pdc1/* large colonies after KmJA.008 was grown on glucose	This work
KmJA.113	KmJA.148; *ura3*	This work
KmJA.114	KmJA.113; p*KmPDC1pr-KmPDC1-ScPGK1ter [URA]*	This work
KmJA.115	KmJA.113; p*KmPDC1pr- KmTHI3-ScPGK1ter [URA]*	This work
KmJA.116	KmJA.113; p*KmPDC1pr- KmARO10-ScPGK1ter [URA]*	This work
KmJA.117	KmJA.113; p*KmPDC1pr-Empty-ScPGK1ter [URA]*	This work
KmJA.178	KmJA.051; *thi4*	This work
KmJA.244	KmJA.178; I4::*THI3* locus	This work

KmJA.151	NBRC1777; *pdc1*	This work
KlJA.210	CBS2359; *pdc1*	This work
KmJA.179	KmJA.051; *ura3*	This work
KmJA.198	KmJA.179 with plasmid pP2-*KmTHI3*-T5	This work
KmJA.199	KmJA.179 with plasmid pP2-*ScTHI3*-T5	This work
KmJA.200	KmJA.179 with plasmid pP2-*Empty*-T5	This work

3. Construction of plasmids

3.1. Construction of CRISPR plasmids

All plasmids and oligonucleotides used in this study are listed in Table 2 and Table S1, respectively. pUCC001 is a CRISPR plasmid with a pangenomic replication origin that allows its use for CRISPR application in several yeast species (Juergens *et al.*, 2018; Varela *et al.*, 2019). For high copy number of CRISPR plasmid in *K. marxianus,* we constructed pJA_U $gRNA_{empty}$ that has *K. marxianus* high copy number replication origin (*ARS2*). To construct this plasmid, pUCC001 was digested with EcoRI restriction enzyme and the single guide RNA transcriptional unit from the linearised plasmid was amplified with primers JA603 and JA612 using Q5 High-Fidelity Polymerase (New England Biolabs, Ipswich, UK). The amplification product with 5' and 3' BsmBI/T7 ligase Golden Gate overhangs was purified and put together in a Golden Gate reaction with plasmids pP7Cas9T1 and pMTU-DO-HPH-C4 to generate the empty CRISPR plasmid pJA_U $gRNA_{empty}$ from which codon optimised *Streptococcus pyrogenes Cas9* from the yeast toolkit (Lee *et al.*, 2015) can be expressed under the control of the *KmTEF1* promoter and *KmINU1* terminator in *K. marxianus*. The plasmid pMTU-DO-HPH-C4 provided a backbone containing a bacterial kanamycin resistant gene, a yeast hygromycin resistant gene and *ASR2* in pJA_U $gRNA_{empty}$. It was then possible to introduce a protospacer DNA insert for a specific target into pJA_U $gRNA_{empty}$ in the same manner as pUCC001 (Rajkumar and Morrissey, 2022). Briefly, complementary oligonucleotides of the gRNA protospacer sequence are designed and synthesised with 5′ and 3′ overhangs (5′-CGTC-3′ and 5′-AAAC-3′) on the sense and antisense strands, respectively. The sequence of CRISPR targets used in this work are listed in Table S2. The oligonucleotides were denatured, annealed, and phosphorylated with 1 µL of T4 polynucleotide kinase (10U/µL, NEB) in a total reaction volume of 10 µL. Fifty femto mole of the annealed protospacer insert was then put together with 100ng of pJA_U $gRNA_{empty}$ in a BsaI/T7 ligase Golden Gate reaction. The reaction product was transformed and selected on kanamycin supplemented LB plates and bacterial transformants harbouring plasmids with a correctly inserted gRNA protospacer DNA were confirmed by PCR using primer JA603 and the antisense oligonucleotide of a gRNA protospacer DNA.

3.2. Construction of plasmids for genes expressions

Plasmids for expressing 2ODC genes were constructed by inserting *KmPDC1pr-gene ORF-ScPGK1ter* into vectors pTU-DO-URA. This was done by means of BsaI/T7 ligase Golden Gate reactions consisting of pTU-DO-URA, pKmK.P2 (*KmPDC1* promoter) and pYTK054

(*ScPGK1* terminator) alongside pL1_*KmPDC1,* pL1_*KmARO10*, pL1_*KmTHI3* or pJA_L1 *Mock gene25*. The complementation plasmid for *KmTHI3* locus was constructed by inserting the *THI3* open reading frame sequence and its upstream and downstream intergenic sequences into integrative vector pI4-MTU-DO-HPH in a BsaI/T7 ligase Golden Gate reaction consisting of this vector and pL1_*KmTHI3* locus to generate pI4-*KmTHI3* locus. All expression plasmids were confirmed by colony PCR and subsequently by NotI restriction.

Table 2. Plasmids used in this study.

Name	Description	Source
pMTU-DO-HPH-C4	*GFP drop-out- KanR-KmARS2/C4- hphMX*	Addgene ID. 160343
pUCC001	*AaTEF1pr-SpCas9-ScPHO5t ScTDH3pr-HH-gRNAempty-HDV-ScCYC1t hphMX*	Addgene ID. 124451
pP7Cas9T1	*KmTEF1pr-SpCas9-KmINU1t- KmARS7/CenD-AmpR- ScURA3*	This work
pJA_U_$gRNA_{empty}$	*KmTEF1pr-SpCas9-KmINU1t- KmARS7/CenD-ScTDH3pr-HH-gRNAempty-HDV-ScCYC1t-KanR- KmARS2/C4- hphMX*	This work
pJA_U_$gRNA_{KmPDC1}$	pJA_U_*gRNAempty* with gRNA protospacer sequence *targeting KmPDC1* inserted	This work
pJA_U_$gRNA_{KmTHI3}$	pJA_U_*gRNAempty* with gRNA protospacer sequence targeting *KmTHI3* inserted	This work
pJA_U_$gRNA_{KmARO10}$	pJA_U_*gRNAempty* with gRNA protospacer sequence targeting *KmARO10* inserted	This work
pJA_U_$gRNA_{KmURA3}$	pJA_U_*gRNAempty* with gRNA protospacer sequence targeting *KmURA3* inserted	This work
pJA_U_$gRNA_{KmTHI4}$	pJA_U_*gRNAempty* with gRNA protospacer sequence targeting *KmTHI4* inserted	This work
pTU-DO-URA	*GFP drop-out- AmpR-KmARS7/CenD -ScURA3*	This work
pP2-*KmPDC1*	pTU-DO-URA with insert *KmPDC1pr- KmPDC1 -ScPGK1t*	This work
pP2-*KmTHI3*	pTU-DO-URA with insert *KmPDC1pr- KmTHI3 -ScPGK1t*	This work
pP2-*KmARO10*	pTU-DO-URA with insert *KmPDC1pr- KmARO10 -ScPGK1t*	This work
pP2-*Empty*	pTU-DO-URA with insert *KmPDC1pr- Empty -ScPGK1t*	This work
pKmK.P2	*KmPDC1* promoter in pYTK001	Addgene ID. 125035
pYTK054	*ScPGK1* promoter in pYTK001	Addgene ID. 65161

pL1_*KmPDC1*	*K. marxianus PDC1*in pYTK001	This work
pL1_*KmTHI3*	*K. marxianus THI3* in pYTK001	This work
pL1_*KmARO10*	*K. marxianus ARO10* in pYTK001	This work
pL1_*ScTHI3*	*S. cerevisiae THI3* in pYTK001	This work
pJA_L1 *Mock gene25*	Level I plasmid of 25bp DNA fragment for connecting YTK standard promoters to terminators	This work
pL1_*KmTHI3* locus	*K. marxianus THI3* open reading frame sequence and its upstream and downstream intergenic sequences (*KmTHI3* locus sequence) in pYTK001	This work
pI4-MTU-DO-HPH	Integrative plasmid with GFP dropout targeting locus downstream of *KmHSP104*; *hphMX* marker	Addgene ID. 160216
pI4-*KmTHI3* locus	pI4-MTU-DO-HPH with *KmTHI3* locus	This work
pYTK036	*SpCas9* in pYTK001	Addgene ID. 65143
pKmK.P7	*KmTEF1* promoter in pYTK001	Addgene ID. 125040
pKmK.T1	*KmINU1* terminator in pYTK001	Addgene ID. 125053
pKmK.T5	*KmPGK1* terminator in pYTK001	
pMTU-DO-URA-C4	*GFP drop-out- KanR-KmARS2-ScURA3*	Addgene ID. 160339
pP2-*KmTHI3*-T5	pMTU-DO-URA-C4 with insert *KmPDC1pr-KmTHI3-KmPGK1t*	This work
pP2-*ScTHI3*-T5	pMTU-DO-URA-C4 with insert *KmPDC1pr-ScTHI3-KmPGK1t*	This work
pP2-*Empty*-T5	pMTU-DO-URA-C4 with insert *KmPDC1pr-Empty-KmPGK1t*	This work

4. Construction of yeast strains

KmASR.005 was used as parental strain from which all the mutants used in this study were derived unless indicated otherwise. Yeast transformations were done by a variation of the lithium acetate/PEG method (Rajkumar and Morrissey, 2022). Gene disruptions in *K. marxianus* and *K. lactis* were undertaken as previously described (Rajkumar and Morrissey, 2022). Briefly, cultures were transformed with CRISPR plasmids alongside repair DNA with 95bp homology flanks for each gene. Transformants were selected on solid YPD or YPE medium supplemented with 400mg/L hygromycin and those with successful gene deletions were identified by colony PCR. The primers used were complementary to sequences within genes but outside the deleted region. Before further work, isolated deletion strains were cured of CRISPR plasmids by passaging on liquid and solid YPD or YPE media. Strains with disruptions in putative 2ODC genes were routinely cultured in YPE for all subsequent work until growth studies in glucose were undertaken. For individual expression of *KmPDC1, KmARO10* and *KmTHI3* in the triple deletion strain KmJA.148, *KmURA3 was* disrupted in the strain to generate KmJA.113 which was subsequently transformed with the expression plasmids pP2-*KmPDC1*, pP2-*KmTHI3*, pP2-*KmARO10* and pP2-Empty to generate strains KmJA.114, KmJA.115, KmJA.116 and KmJA.117, respectively. For *KmTHI3* complementation, integrative cassette of *KmTHI3* locus sequence was first obtained by digestion of pI4-*KmTHI3* locus with SgsI (FD1894, Fisher Scientific). The digestion product was subsequently used to transform into KmJA.051 and selected on YPD agar supplemented with 400mg/L hygromycin for integration into its genome. Isolates with the cassette correctly integrated were identified by colony PCR.

5. Determination of enzyme K_m and V_{max}

Cell extracts were obtained from batch (flask) culture of strains growing in minimal glucose medium as previously described (Romagnoli *et al.*, 2012) . Briefly, cell cultures at mid-exponential growth phase were washed twice with freezing solution (10mM potassium phosphate buffer (pH7.5) and 2mM EDTA) and stored at -20^{o}C in this solution until the assay was carried out. The cells were thawed at room temperature, and then washed with and resuspended in the sonication solution (100mM potassium phosphate buffer (pH7.5) containing 2mM magnesium chloride and 2mM dithiothreitol) for sonication. Three grams of 0.7mm glass beads were added to the resuspended culture and sonicated by 6 bursts lasting 30 seconds each at an amplitude of 8µm using an MSE sonicator at 0^{o}C. Sonicated cultures were centrifuged at 4^{o}C at 36,000 Xg for 20 minutes to remove cell debris and the supernatant (purified cell extract) was used for enzyme assays.

Decarboxylase activities with linear 2OAs (pyruvate and 2-oxo butanoate) in freshly prepared cell extracts was determined in reaction mixtures of 1mL total volume consisting of the cell extracts, 40mM imidazole-HCl (pH6.5), 5mM $MgCl_2$, 0.2mM thiamine pyrophosphate, 0.15mM NADH, 88U/mL alcohol dehydrogenase (Sigma-Aldrich). Reactions were started after determining the mixtures' basal absorbance at 340nm by adding substrates to the mixture, causing the absorbance to decrease due to NADH oxidation into NAD+. K_m and V_{max} values were calculated from reaction rates with pyruvate concentrations between 0 to 50mM for PDC activity and 50mM of 2-oxo butanoate. For phenylpyruvate, ketoisovalarate and α-keto-γ-(methylthio)butyrate (KMBA), activities were also measured in 1mL reaction mixtures but consisting of the cell extract, 100mM KH_2PO_4-K_2HPO_4 buffer (pH7.0), 0.2mM thiamine diphosphate, 5mM $MgCl_2$, 15mM pyrazole, 2mM NAD+ and 1.75U of yeast aldehyde dehydrogenase dissolved in 1mM dithiothreitol (Sigma-Aldrich). The reactions were started after the determination of the mixtures' basal absorbance at 340nm by adding substrates to the mixtures after which the absorbance increased due to NAD+ reduction into NADH. K_m and V_{max} values were calculated from the rates of reactions activated by substrate concentrations from 0.025mM to 10mM for phenylpyruvate and 0.25mM to 20mM for ketoisovalarate, and 20mM for KMBA. Reaction rates, given as the difference between the slope of the line representing the change in absorbance before and after substrates addition, were linearly proportional to the amount of cell extract added. Total protein concentrations in extracts were determined by previously described method (Lowry *et al.*, 1951) using bovine serum albumin (Sigma-Aldrich) as standard, and were used to normalise the reaction rates. K_m and V_{max} values were calculated by fitting kinetic data with GraphPad Prism 9 (Graph-Pad Software) using nonlinear regression of the Michaelis-Menten equation.

6. RNA preparation and quantification by RT-qPCR

One millilitre of yeast cell cultures at early exponential growth phase from three independent cultures of each strain were harvested by centrifugation at 17000 Xg for 1 minute at 4°C after which 900µL of the supernatant was replaced with RNAlater (AM7021, Fisher Scientific), the cells were resuspended, incubated at 4°C for 24 hours and stored at -80°C until total RNA isolation. Total RNA was prepared from cells using RNeasy Kit (Cat. No./ID: 74904, Qiagen) and treated with RNase-Free DNase Set (Cat. No./ID: 79254, Qiagen) according to the manufacturer's instructions. cDNA was generated from 1µg of total RNA using ProtoScript II First Strand cDNA Synthesis Kit (E6560, NEB) and oligo d(T) primers in 20µL reactions according to manufacturer's instruction.

For RT-qPCR, 20µl of reaction mixtures consisting of 10µl of Light Cycler 480 SYBR Green I Master (Roche), 0.5µM of each primer (Life Technologies-Invitrogen, Milan, Italy) and

diethylpyrocarbonate-treated water (DEPC-water) was used in a LightCycler 480 Real-Time PCR System (Roche). Before usage, primers for housekeeping genes (*KmACT1* and *KmALG9*) and genes of interest were validated with serial diluted genomic DNA and confirmed to be specific based on single peaks of melting curves and to have amplification efficiencies between 90 and 105%. To obtained relative transcript abundance, sample Ct values were determined in duplicates and normalised by Ct values of the housekeeping genes.

7. Quantification of extracellular metabolites

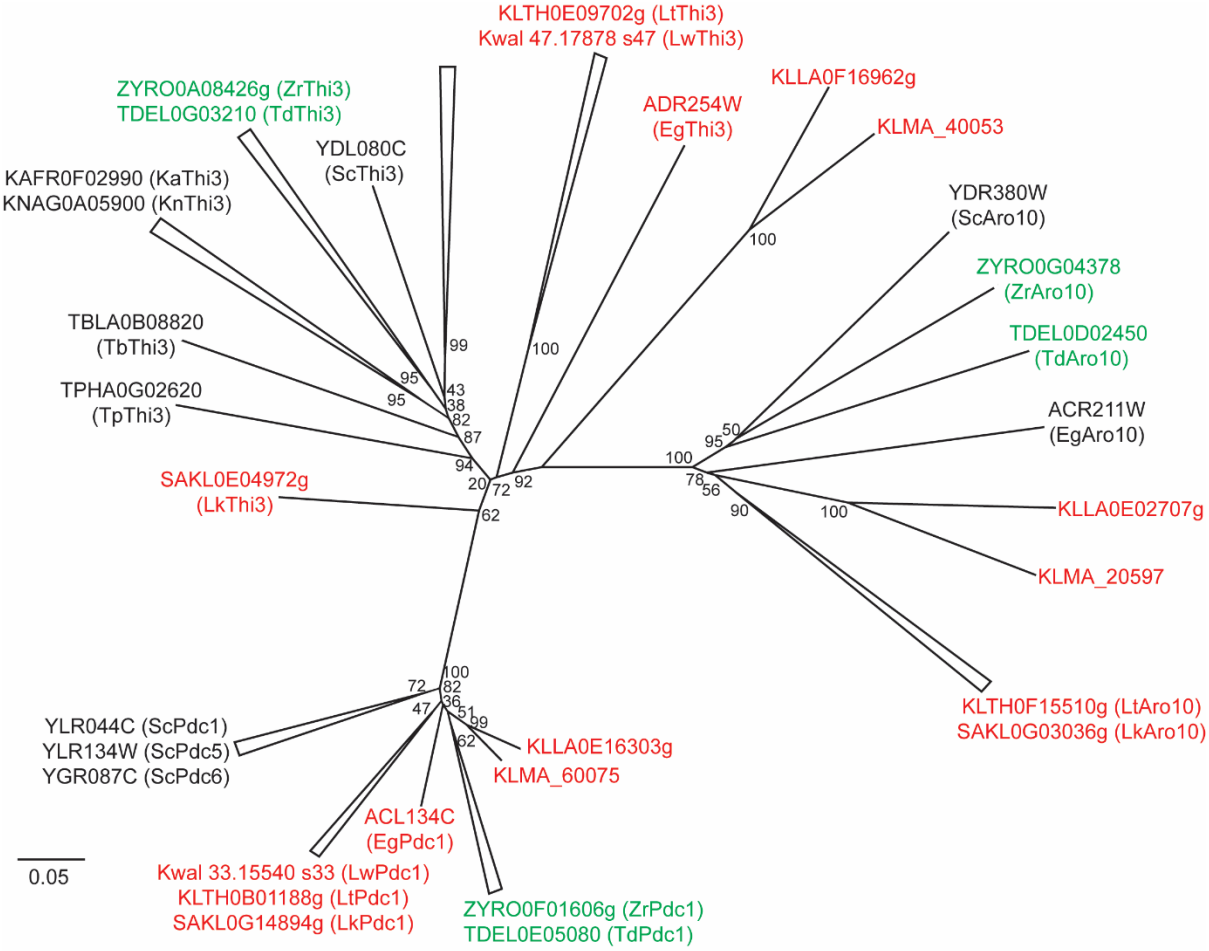
For profiling extracellular metabolite production, culture supernatants were obtained by centrifuging 1mL of cultures at 17000 ×g for 5 minutes and stored at -20°C until metabolite concentrations were determined by high performance liquid chromatography (HPLC). For aromatic compounds, 700µL of cultures were mixed with 96% ethanol at 1:1 ratio before centrifugation. Metabolite concentrations in supernatants were determined by HPLC at appropriate dilutions. Phenylpyruvate, 2-phenylethanol and phenylacetic acid were analysed as previously described (Koopman *et al.*, 2012). Briefly, Agilent 1260 Infinity HPLC system with a Zorbax Eclipse Plus C18 column (4.6 x 150mm, 3.5µm; Agilent, Little Island, Ireland) operating at 40°C was used. As mobile phase, 20mM of KH_2PO_4 and 1% acetonitrile (pH2)/acetonitrile at flow rate of 0.8mL/min with an elution gradient in which acetonitrile was increased from 0-10% for the first 6 minutes, then increased to 40% between 6 to 23 minutes, and then switched to 100% KH_2PO_4 for 23 to 27 minutes. Detection was by VWD100 multi-wavelength detector measuring at 200nm. The concentrations of glucose and extracellular acetate, pyruvate, ethanol, and glycerol were determined using Agilent 1200 HPLC system (Agilent Technologies, CA, USA) with a REZEX 8µL 8% H+ organic acid column (300 ×7.8 mm) (Phenomenex, CA, USA) operating at 65°C using 0.01 N H2SO4 as mobile phase at a flow rate of 0.6mL/min. Detection was by a refractive index detector. HPLC data were analysed using OpenLab CDS (Agilent) and standards were obtained from Sigma-Aldrich.

Results

1. Prediction of *K. marxianus* 2-oxo acid decarboxylases

To predict the function of the 3 putative 2ODC protein hits from *K. marxianus* and *K. lactis,* phylogenetic analysis was conducted. Based on the evidence-supported evolutionary paradigm that the whole genome duplication observed in yeasts of the *S. cerevisiae* lineage is as a result of interspecies hybridization of the genome of budding yeasts belonging to the Kluyveromyces, Lachancea, Eremothecium (KLE) species and the Zygosaccharomyces, Torulaspora (ZT) species (Marcet-Houben and Gabaldón, 2015), protein hits in yeasts from these species and those from post-whole genome duplication (post-WGD) species were

included in the phylogenetic analysis. With 2ODCs from *S. cerevisiae* previously characterised (Romagnoli *et al.*, 2012), they were used to predict the functions of orthologues from other species (figure 2A). The proteins distributed into clusters containing *S. cerevisiae* Aro10, Thi3 and Pdc and their functions can be inferred according to the *S. cerevisiae* proteins present in each cluster. Aro10 and Thi3 from ZT species clustered with orthologues from the post-WGD species, suggesting that *ARO10* and *THI3* genes in post-WGD species come from ZT species. On the other hand, *S. cerevisiae* Pdc1, Pdc5 and Pdc6 clustered together and separated from proteins in both the KLE and ZT species in the Pdc cluster. Thus, the origin of *S. cerevisiae* Pdc proteins cannot be deduced from the tree. Three putative 2ODC genes are common in Kluyveromyces species and are located on chromosomes II (*KLMA_20597*), IV (*KLMA_40053*) and VI (*KLMA_60075*) in *K. marxianus*. It can be deduced from the tree that *KLMA_20597* is *KmARO10*, *KLMA_60075* is *KmPDC1, KmPDC5,* or *KmPDC6*. Although the tree also showed that *KLMA*_40053 and its *K. lactis* orthologue are Thi3, it is interesting that the Kluyveromyces Thi3 proteins are divergent from the Thi3 proteins from other yeasts. Synteny analysis confirmed the identities of these genes due to the presence of orthologues of neighbouring genes of the *S. cerevisiae* 2ODC genes in the vicinities of their predicted orthologues in both Kluyveromyces species (figure 2B). Although the orthologues of the genes surrounding *PDC5* in *S. cerevisiae* are also found in the Kluyveromyces yeasts, *PDC5* itself is absent at the locus.


KLTH0E09702g (LtThi3)
Kwal 47.17878 s47 (LwThi3)
KLLA0F16962g
ADR254W
(EgThi3)
KLMA_40053
ZYRO0A08426g (ZrThi3)
TDEL0G03210 (TdThi3)
YDL080C
(ScThi3)
KAFR0F02990 (KaThi3)
KNAG0A05900 (KnThi3)
YDR380W
(ScAro10)
ZYRO0G04378
(ZrAro10)
TBLA0B08820
(TbThi3)
TDEL0D02450
(TdAro10)
TPHA0G02620
(TpThi3)
ACR211W
(EgAro10)
SAKL0E04972g
(LkThi3)
KLLA0E02707g
KLMA_20597
KLTH0F15510g (LtAro10)
SAKL0G03036g (LkAro10)
YLR044C (ScPdc1)
YLR134W (ScPdc5)
YGR087C (ScPdc6)
KLLA0E16303g
KLMA_60075
ACL134C
(EgPdc1)
0.05
Kwal 33.15540 s33 (LwPdc1)
KLTH0B01188g (LtPdc1)
SAKL0G14894g (LkPdc1)
ZYRO0F01606g (ZrPdc1)
TDEL0E05080 (TdPdc1)

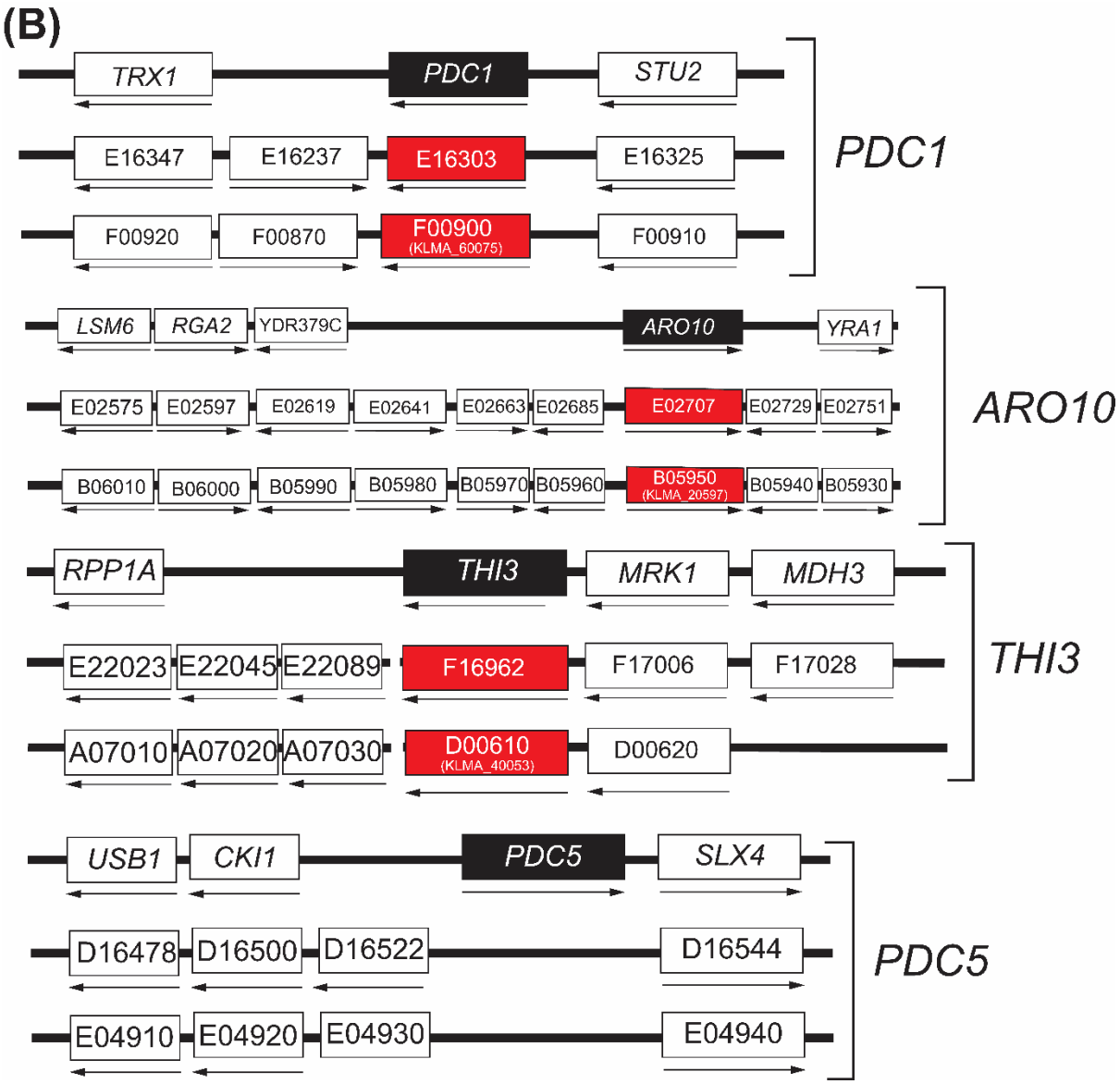


Figure 2. Prediction of the function of putative *K. marxianus* 2ODC genes by bioinformatic analysis. (A) Tree showing the closeness based on sequence identity between 2ODCs of yeasts belonging to the Kluyveromyces, Lachancea, and Eremothecium (KLE) species (red texts), Torulaspora and Zygosaccharomyces (ZT) species (green texts) and post-whole genome duplication species (black texts). The tree was constructed using neighbour-joining method. (B) Illustration of synteny for demonstrating that the genes found in the vicinities of the *ARO10*, *PDC1* and *THI3* in *S. cerevisiae* (top) are also found around the *K. lactis* (middle) and *K. marxianus* (bottom) 2ODC genes as displayed on YGOB (http://ygob.ucd.ie/).

2. *K. marxianus* strain with triple disruption of *PDC1, ARO10* and *THI3* showed delayed growth on glucose.

To construct strains that individually express *PDC1, ARO10* and *THI3,* we needed to first disrupt the three genes in order to generate a strain that is devoid of 2ODC activities. The genes were disrupted in wildtype *K. marxianus* (KmASR.005) by CRISPR-mediated clean

deletions as illustrated in figure S1A on YPE to generate strain KmJA.164. Growth analysis on glucose from ethanol precultures showed that KmJA.164 grew with a lag phase time of about 50 hours compared to 4 hours for KmASR.005 (Figure 3A). To determine whether the delayed growth observed in the first round of cultivation persisted, KmJA.164 cultures were re-inoculated for a second round of cultivation in liquid medium at the end of 250 hours of cultivation in the first round with an intermediate ethanol preculturing step. The ethanol preculture was also spread on solid medium to obtain single colonies. In liquid minimal glucose medium, the culture showed considerably shorter lag phase time in this second round than in the first round, and large and small colonies emerged on solid medium (Figure 3A). The cultivation of one of the small colonies (KmJA.147) and one of the large colonies (KmJA.148) in liquid medium showed that both strains grew within 24 hours during which the growth of KmJA.164 was undetectable (Figure S2A). Large and small colonies emerged from KmJA.147 but only large colonies emerged from KmJA.148 on solid medium (Figure S2B). Put together, it is clear from these observations that most of cells of the triple deletion strain were unable to grow in glucose. The media, however, selects for cells that are able to grow, probably due to mutation. These cells populated the medium in the first cultivation round and the delayed lag phase time can be explained by the fact that these cells are few in the preculture and at the beginning of the cultivation. The lag phase time decreased in the second round because the cultivation was started with a preculture that is abundant in cells that have the ability to grow on glucose. The quantification of extracellular metabolites during growth on glucose showed that KmJA.148 accumulated pyruvate and acetate but produced no ethanol unlike the wildtype strain which did not accumulate pyruvate and produced acetate and ethanol (Figure 3B)

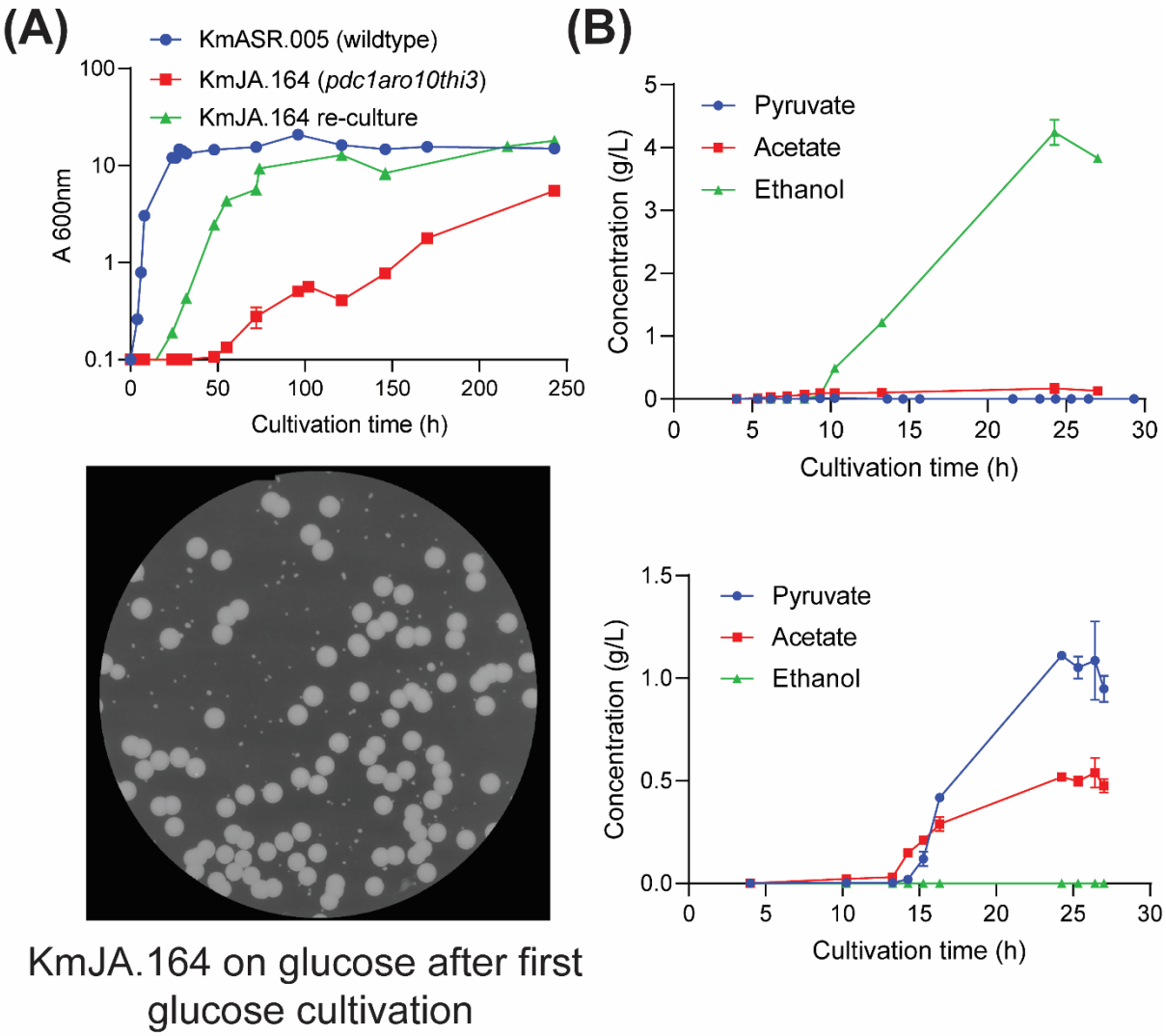


Figure 3. Disruption of the putative 2ODC genes of *K. marxianus*. (A) *PDC1, ARO10* and *THI3* were disrupted in wildtype *K. marxianus* to generate KmJA.164 and both strains were cultivated in minimal glucose media (top panel). At the end of the cultivation, KmJA.164 culture was re-cultivated in fresh liquid (green line in the top panel) and solid (bottom panel) media. (B) The concentrations of extracellular pyruvate, acetate and ethanol were determined for wildtype (top panel) and for one of the large colonies of KmJA.164 on solid media (bottom panel). Data shown are mean ± standard deviation of three replicates.

3. 2-oxo acid decarboxylase activities of KmPdc1, KmAro10 and KmThi3.

To determine 2ODC kinetics of the three genes, KmJA.148 was transformed with plasmids individually harbouring each gene for constitutive expression under the control of *KmPDC1* promoter and *ScPGK1* terminator on glucose. Cell extracts were obtained from strains KmJA.114, KmJA.115 and KmJA.116, expressing *KmPDC1, KmTHI3* and *KmARO10*, respectively, and from KmJA.117 which is devoid of the 3 genes but contains the plasmid

backbone and KmASR.005. The extracts were uses for enzyme assays with various 2OAs as substrate (Table 1). Serving as positive control, activities were detected in assays with KmASR.005 extracts for all substrates while KmJA.117 extracts displayed no activity for any of the substrates, confirming that the disruption of the three genes eradicated 2ODC activities.

3.1. KmPdc1 and KmAro10 decarboxylate pyruvate

As a linear 2OA substrate derived from and crucial in the metabolism of fermentable carbon sources like glucose during glycolysis, higher V_{max} values of 3.54 U/mg of protein was detected for pyruvate in KmJA.114 compared to 0.19 U/mg of protein in KmJA.116, indicating that Pdc1 shows higher PDC activity than Aro10. Furthermore, lower K_m value was measured in KmJA.114 (5.38mM) than KmJA.116 whose kinetics could not be fitted on a Michaelis-Menten plot (Figure S3), meaning that Pdc1 is more specific for pyruvate than Aro10. PDC activity was not detected for KmJA.115 expressing Thi3. PDC activity were also examined *in vivo* for these proteins based on the detection of ethanol production during the cultivation of the strains in glucose medium. With ethanol production detected at a maximum titre of 4.2g/L in KmASR.005 and completely lost in KmJA.117, the individual expression of *KmPDC1* in KmJA.114 restored ethanol production to titres close to that in KmASR.005, *KmARO10* in KmJA.116 produced some ethanol but at low titre (Figure 4A). These results were accompanied by the ability of KmASR.005 and KmJA.114 to completely metabolise glucose while KmJA.117 and KmJA.116 were unable to do so (Figure 4B). The results here agree with the conclusions of our bioinformatics analysis that *KLMA_60075* is the pyruvate decarboxylase *PDC1* gene in *K. marxianus*. It is interesting that PDC activity was detected for KLMA_20597, which was predicted to be Aro10 because *S. cerevisiae* Aro10 has not been reported to show this activity.

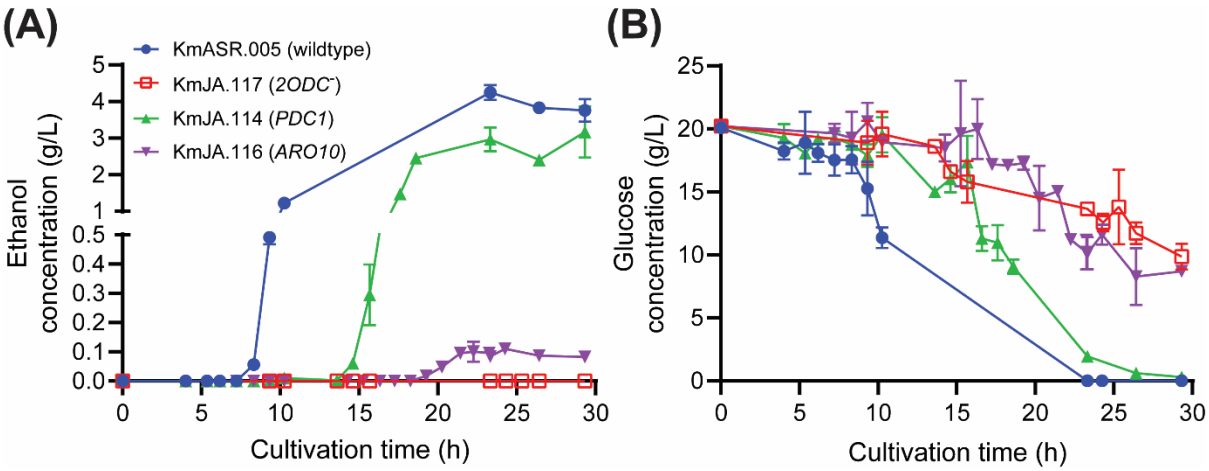


Figure 4. PDC activity of *K. marxianus* 2ODCs *in vivo*. Strains KmASR.005 (wildtype); KmJA.117 (*pdc1thi3aro10pEMPTY*); KmJA.116 (*pdc1thi3aro10pARO10*); KmJA.114 (*pdc1thi3aro10pPDC1*) were grown on minimal glucose medium to determine their (A) extracellular concentration of ethanol

produced, and (B) glucose consumption profile. Data shown are mean ± standard deviation of duplicates.

3.2. KmPdc1 and KmAro10 decarboxylate amino acids-derived 2-oxo acids.

For 2OAs generated by the deamination of amino acids, KmJA.116 displayed higher V_{max} than KmJA.114 in the cases of phenylpyruvate, ketoisovalarate and KMBA but not for the linear substrate, 2-oxo butanoate, for which, similar to pyruvate, a higher V_{max} was obtained in KmJA.114 than KmJA.116. Very low K_m values of 0.18mM and 0.6mM were measured for KmJA.116 with phenylpyruvate and ketoisovalarate, respectively, compared to KmJA.114 whose kinetics for both substrates could not be fitted on a Michaelis-Menten plot (Figure S3), showing that Aro10 is more specific than Pdc1 for phenylpyruvate as an example of aromatic 2OAs derived from aromatic amino acids, and for ketoisovalarate, an example of branched chain 2OAs derived from branched chain amino acids. Activity was not detected in KmJA.115 for all four substrates. Phenylpyruvate decarboxylase activity was examined *in vivo* by means of biomass-normalised titres of 2-phenylethanol and phenylacetic acid (2PE+PAA) production during the cultivation of the expression strains in glucose medium with phenylalanine as sole nitrogen source. The PAA+2PE produced by KmASR.005 was completely lost in KmJA.117, restored to in KmJA.116, and KmJA.114 produced some but at low titres (Figure 5A). Thus, the catalysis of Aro10 is in line with our prediction by bioinformatic analysis that this gene is indeed *ARO10*, and Pdc1 also displayed decarboxylase activities for amino acid-derived 2OA substrates. The *in vivo* results were accompanied by the detection of the high titres of phenylpyruvate in KmJA.117 and KmJA.114 which was lost in KmJA.116, while production in KmASR.005 was detected at lower titres than in KmJA.117 and KmJA.114 but higher than KmJA.116 (Figure 5B).

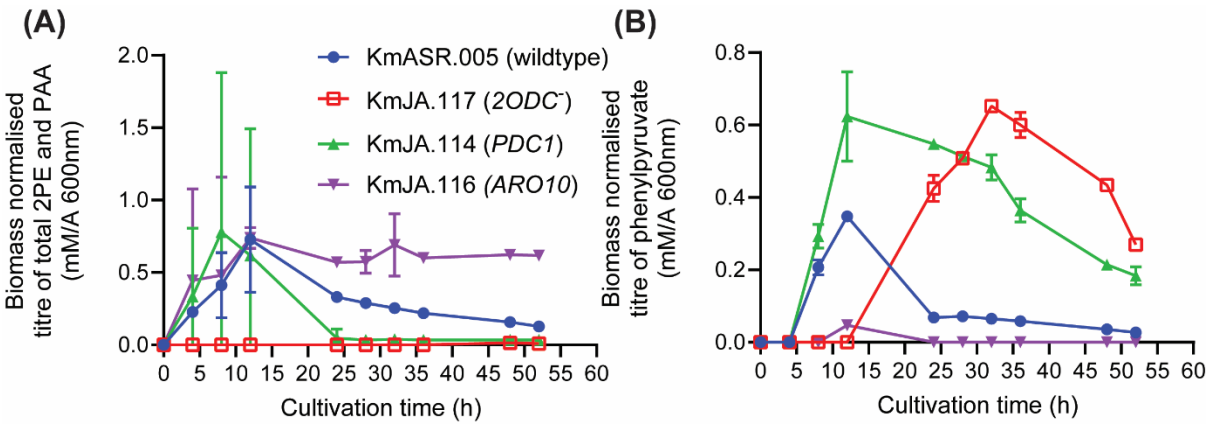


Figure 5. Phenylpyruvate decarboxylase activity of putative *K. marxianus* 2ODCs *in vivo*. Strains KmASR.005 (wildtype); KmJA.117 (*pdc1thi3aro10pEMPTY*); KmJA.116 (*pdc1thi3aro10pARO10*); KmJA.114 (*pdc1thi3aro10pPDC1*) were grown on minimal glucose medium with phenylalanine as the sole nitrogen source to determine their production of Ehrlich metabolites. (A) The titres of 2-phenylethanol (2PE) and phenylacetic acid (PAA) produced were normalised by A 600nm values of the

culture at the time of sampling, and (B) The titres of accumulated extracellular phenylpyruvate were normalised by A 600nm values of the culture at the time of sampling. Data shown are mean ± standard deviation of duplicates.

Table 1. *in vitro* determination of 2ODC kinetics of *KmPdc1, KmThi3* and *KmAro10. K. marxianus* strain which has the three putative 2ODC genes disrupted and is able to grow on glucose (KmJA.148) was transformed with plasmids bearing each gene for individual expression under the control of a constitutive promoter. Decarboxylase activities for 2-oxo acid substrates were determined in cell extracts obtained from cultures of the generated strains at mid-exponential growth phases on minimal glucose medium. Data shown are mean ± standard deviation of duplicates. NA, not applicable; BD, below the detection limit with a Vmax value of < 0.02 U/mg of protein; ND, not determined.

2ODC class	Substrate (Amino acid)	Strain (Gene regulation)	Mean K_m (mM) ± SD	Mean V_{max} (U/mg of protein) ± SD
Linear	Pyruvate	KmASR.005 (wildtype)	NA	1.77 ± 0.28
		KmJA.117 (*2ODC* $^{-}$)	NA	BD
		KmJA.116 *(KmARO10)*	15.79 ± 1.57	0.19 ± 0.01
		KmJA.115 (*KmTHI3*)	NA	BD
		KmJA.114 (*KmPDC1*)	5.38 ± 0.76	3.54 ± 0.06
	2-oxo butanoate (Threonine)	KmASR.005 (wildtype)	ND	0.48 ± 0.03
		KmJA.117 (*2ODC* $^{-}$)	ND	BD
		KmJA.116 *(KmARO10)*	ND	0.63 ± 0.05
		KmJA.115 (*KmTHI3*)	ND	BD
		KmJA.114 (*KmPDC1*)	ND	1.74 ± 0.6
Aromatic	Phenylpyruvate (Phenylalanine)	KmASR.005 (wildtype)	NA	0.02 ± 0.001
		KmJA.117 (*2ODC* $^{-}$)	NA	BD
		KmJA.116 *(KmARO10)*	0.18 ± 0.08	0.81 ± 0.08
		KmJA.115 (*KmTHI3*)	NA	BD
		KmJA.114 (*KmPDC1*)	12.29 ± 5	0.1 ± 0.003

Branched chain	Ketoisovalarate (Valine)	KmASR.005 (wildtype)	NA	0.17 ± 0.02
		KmJA.117 (*2ODC* $^{-}$)	NA	BD
		KmJA.116 *(KmARO10)*	0.6 ± 0.04	0.6 ± 0.005
		KmJA.115 (*KmTHI3*)	NA	BD
		KmJA.114 (*KmPDC1*)	19.78 ± 6.4	0.4 ± 0.04
Sulphur- containing	KMBA (Methionine)	KmASR.005 (wildtype)	ND	0.06 ± 0.001
		KmJA.117 (*2ODC* $^{-}$)	ND	BD
		KmJA.116 *(KmARO10)*	ND	0.42 ± 0.02
		KmJA.115 (*KmTHI3*)	ND	BD
		KmJA.114 (*KmPDC1*)	ND	0.27 ± 0.02

4. KmThi3 regulates thiamine biosynthesis.

4.1. *K. marxianus* is unable to grow on thiamine-free medium when *THI3* is disrupted.

In line with the bioinformatic evidence that the product of *KmTHI3* is a Thi3 protein, the expression of KmThi3 in KmJA.115 was so far, like ScThi3, unable to decarboxylate any of the 2OA substrates tested. With KmASR.005 able to grow in thiamine-free medium due to its ability to biosynthesise thiamine, the disruption of thiamine biosynthetic thiazole synthase *THI4* gene in KmASR.005 (KmJA.178) rendered it unable to grow in the medium (Figure 6A). To determine whether *KmTHI3* is indeed a *THI3*, the gene was disrupted in KmASR.005 to generate KmJA.051. Like KmJA.178, KmJA.051 was also unable to grow in thiamine-free medium and the genomic integration of the entire *THI3* locus (the sequence of the open reading frame and its intergenic regions) into KmJA.051 to generate strain KmJA.244 reversed the growth phenotype (Figure 6A), strongly indicating that the gene product is a Thi3 protein.

4.2. KmThi3 regulates the transcription of thiamine biosynthetic genes.

The cultivation of KmASR.005 in minimal glucose medium supplemented with 250µg/L of thiamine resulted in 70-fold and 25-fold higher relative transcript abundance than in 1000µg/L for the thiamine biosynthetic genes *THI4* and *THI11*, respectively (Figure 6B). Therefore, to determine whether the transcription of these genes is repressed when *KmTHI3* is disrupted, the relative transcript abundance of both genes was compared in KmASR.005 and KmJA.051 during growth on 250µg/L thiamine. The relative transcript abundance for these genes was lower in KmJA.051 than in KmASR.005 in this condition, confirming that KmThi3 regulates the transcription of *THI4* and *THI11.*

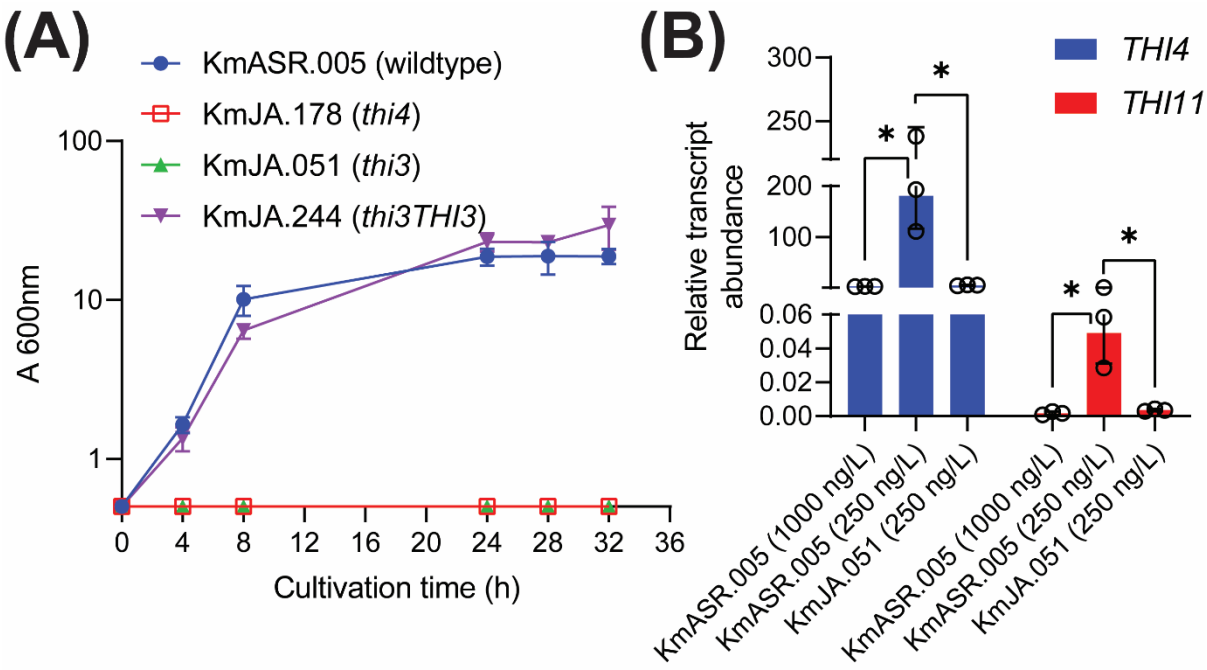


Figure 6. KmThi3 regulates thiamine biosynthesis. (A) The strains shown were grown on minimal glucose medium without thiamine (thiamine-free medium). The inability of KmJA.178 (*thi4*) to grow indicated that the medium was truly thiamine-free. While KmASR.005 which has the wildtype *THI3* gene grew in thiamine-free medium, the disruption of this gene in KmJA.051 (*thi3*) resulted in loss of growth. The re-introduction of functional *KmTHI3* by integrating a cassette of the *KmTHI3* open reading frame sequence and its upstream and downstream intergenic sequences into the genome of KmJA.051 (KmJA.244) revived the strain. (B) RT-qPCR of thiamine biosynthetic genes *THI4* and *THI11* in KmASR.005 and KmJA.051 harvested at early exponential growth phase on minimal glucose medium with 1000 or 250 µg/L of thiamine. Data shown are mean ± standard deviation of three independent experiments. Statistically significant differences (examined by independent ttest) are denoted by an asterisk (*, $p<0.05$).

5. Contributions of *PDC1* and *ARO10* to 2-oxo acid decarboxylase activity in wildtype *K. marxianus.*

5.1. Wildtype PDC activity is supplied solely by Pdc1.

To determine the contributions of *PDC1* and *ARO10* to endogenous PDC activity in *K. marxianus,* the genes were individually disrupted in KmASR.005. This generated strains KmJA.002 (*pdc1*) and KmJA.001 (*aro10*). The specific PDC activities in cell-free extracts from minimal glucose medium-grown cultures of the three strains were determined and the results showed that while PDC activity was completely eradicated in KmJA.002, KmJA.001 showed no difference in activity compared to KmASR.005 (figure 7A). Additionally, metabolite profiling showed that ethanol was not produced by KmJA.002 but persisted at levels similar to KmASR.005 in KmJA.001 (figure 7B). These results show that Pdc1 contributes all the PDC activity in wildtype *K. marxianus* during growth on glucose. The loss of PDC activity in

KmJA.002 resulted in poorer glucose consumption compared to KmASR.005 and KmJA.001, and similar to KmASR.005, low titres of acetate was produced and later consumed in KmJA.001 and KmJA.002 with the latter reaching higher titre than the former before consumption (figure 7B). Furthermore, while the profile of glycerol was similar to that of acetate in KmASR.005 and KmJA.001 in that it was produced and later consumed in these strains, glycerol accumulated in KmJA.002, similar to reports in PDC-null *S. cerevisiae*.

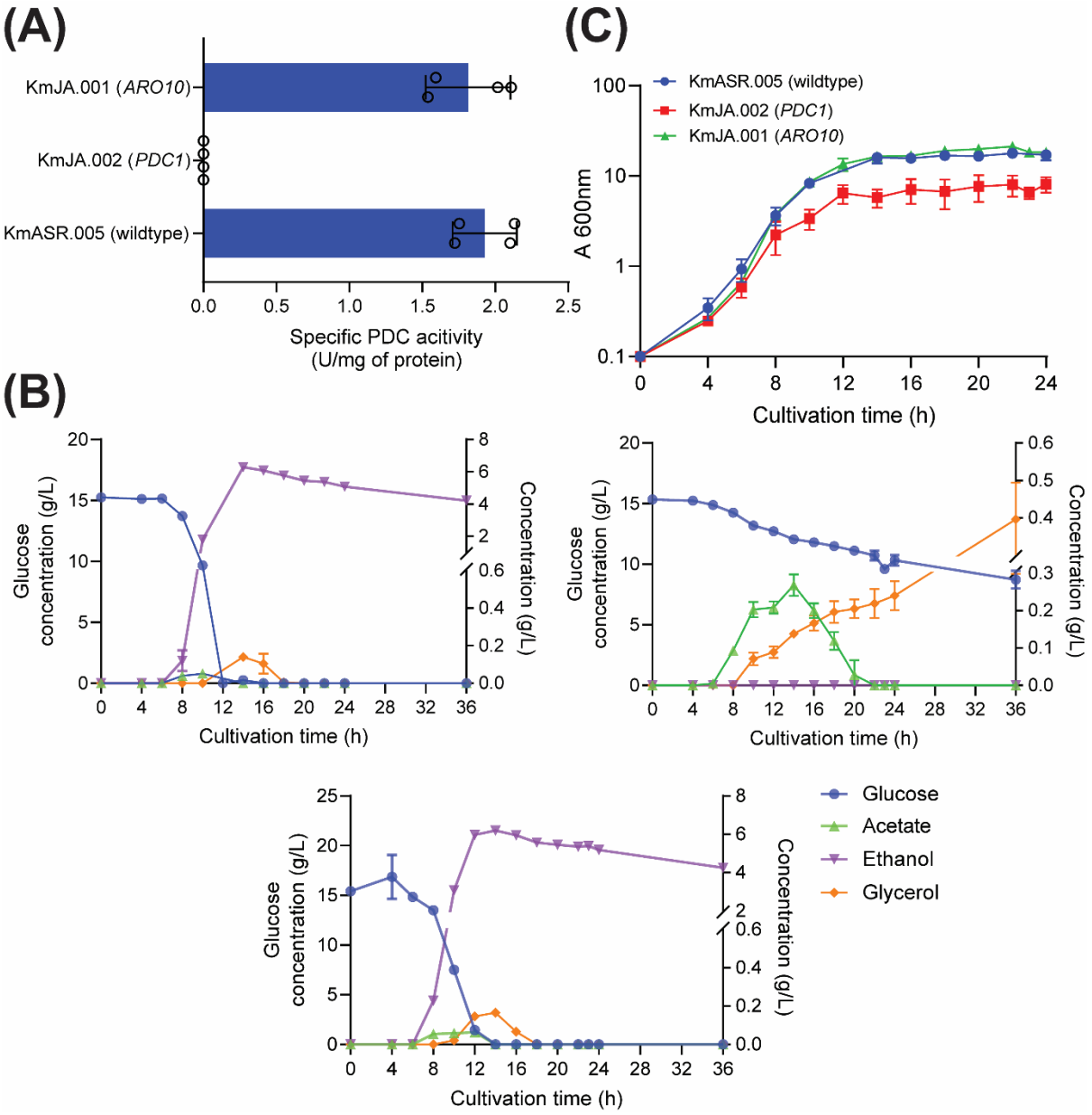


Figure 7. Contributions of *PDC1* and *ARO10* to endogenous pyruvate decarboxylase activity in *K. marxianus*. (A) Specific pyruvate decarboxylase (PDC) activity was determined in cell extracts obtained from cultures at mid-exponential growth phase on minimal glucose medium. Data shown are mean ± standard deviation of two independent experiments. (B) The concentrations of glucose and extracellular metabolites were determined in KmASR.005 (wildtype) at the top left panel, KmJA.002 (*pdc1*) at the

top right panel and KmJA.001 (*aro10*) at the bottom panel during growth in minimal glucose medium. (C) The growth profiles of the 3 stains were determined. Data shown are mean ± standard deviation of three independent experiments for A 600nm, and three replicates for metabolite titres.

5.2. PDC-null *K. marxianus* displays growth defect in glucose medium.

Growth study during the cultivation of KmASR.005, KmJA.001 and KmJA.002 in glucose showed that while KmJA.001 displayed no observable growth defect, KmJA.002 displayed similar growth profile as KmASR.005 until 8 hours into the cultivation after which the growth of KmJA.002 slowed down and at the end of the cultivation, it produced lower biomass than KmASR.005 (figure 7C). This observation contradicts previous reports by others that PDC-null *K. marxianus* (Choo et al., 2018) and *K. lactis* (Bianchi et al., 1996) show no growth defect. KmASR.005, which is the background strain in which *PDC1* was disrupted in this experiment, has *DLN4* disrupted. Thus KmJA.002 is in fact a *dln4pdc1* double deletion strain. To eliminate the possibility that the additional *DLN4* disruption caused the growth defect observed in KmJA.002, *PDC1* was disrupted in truly wildtype *K. marxianus* NBRC1777 and in *K. lactis* CBS2359 strains and both strains were cultivated in glucose medium. However, similar to the results in the *DLN4*-disrupted *K. marxianus*, *PDC1* disruption caused growth defect in wildtype strains of both yeasts in minimal glucose medium (figure S4A). Based on the fact that others previously used complex media in their study, it is possible that their results reported are due to media complexity. However, cultivation in YPD showed that *pdc1* strains of both yeasts produced less biomass than wildtypes (figure S4B)

5.3. Wildtype phenylpyruvate decarboxylase activity is supplied by Aro10 with contribution from Pdc1.

To elucidate the contributions of *PDC1* and *ARO10* to endogenous phenylpyruvate decarboxylase activity of *K. marxianus,* the extracellular metabolite profiles of KmASR.005 and KmJA.001 during cultivation in minimal glucose media with phenylalanine as the sole nitrogen source were determined. The profiles showed that total production of 2-phenylethanol and phenylacetic acid titres was considerably lower in KmJA.001 than in KmASR.005 (figure 8A). It is interesting that these metabolites were produced at all in the absence of *ARO10* in KmJA.001. Since results from the enzyme assays showed that Pdc1, though with poorer specificity than Aro10, has phenylpyruvate decarboxylase activity, it could contribute to the activity under the control of its native promoter. Thus, *PDC1* was additionally disrupted in KmJA.001 to generate KmJA.005. Indeed, 2-phenylethanol and phenylacetic acid production was completely eradicated in KmJA.005, showing that *PDC1* is responsible for the production of these metabolites in KmJA.001. To determine the expression of *ARO10* and *PDC1* at transcriptional level during growth on phenylalanine, RT-qPCR analysis was performed on total RNA samples obtained from KmASR.005 that was cultivated in ammonium or

phenylalanine as sole nitrogen sources. The relative transcript abundance of *ARO10* was higher in phenylalanine than in ammonium with levels in ammonium near to zero, while no significant difference was observed between both nitrogen sources for *PDC1.* To gain insight into the possibility that compensatory regulation through transcriptional upregulation may be responsible for phenylpyruvate decarboxylase activity by *PDC1* in the absence of *ARO10,* the relative transcript abundance of *PDC1* was determined in KmJA.001 but was not significantly different to the relative transcript abundance in KmASR.005 (Figure 8B). This shows that the contribution of phenylpyruvate decarboxylase activity from *PDC1* in the absence of *ARO10* is not through compensatory regulation and that the basal transcription level of *PDC1* is sufficient to mediate the activity *in vivo.*

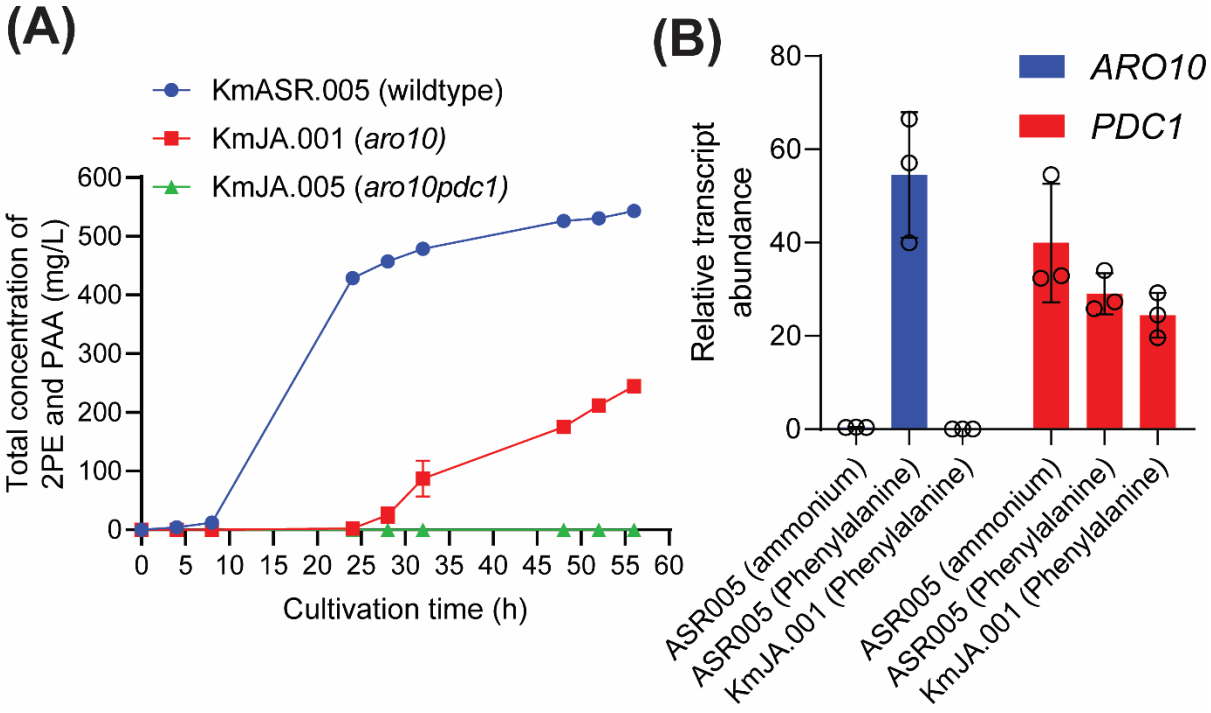


Figure 8. Contributions of *PDC1* and *ARO10* to endogenous phenylpyruvate decarboxylase activity in *K. marxianus*. (A) The total titre of 2-phenylethanol (2PE) and phenylacetic acid (PAA) was determined for the three strains during their growth on minimal glucose medium with phenylalanine as the sole nitrogen source. (B) RT-qPCR of *PDC1* and *ARO10* was determined in KmASR.005 (wildtype) KmJA.001 (*aro10*) and KmJA.005 (aro10pdc1) at early exponential growth phase during growth on minimal glucose media with ammonium or phenylalanine as sole nitrogen sources. Data shown are mean ± standard deviation of three replicates for metabolites and of duplicates for relative transcript abundance.

Discussion

Bioinformatic analysis in this work predicted that *PDC1* and *ARO10* are the two 2ODC genes in *K. Marxianus* and that a gene that was previously thought to be a *PDC5* is actually a *THI3*.

Synteny analysis showed that *K. marxianus* does not have *PDC5.* The enzyme kinetics of the proteins of the three genes were systematically analysed by their individual expression in a *K. marxianus* strain that has the three genes disrupted. Based on *in vitro* enzyme assays with cell-free extracts of the expression strains, 2ODC activities were detected for Pdc1 and Aro10 but not for Thi3, establishing that *K. marxianus* contains only *PDC1* and *ARO10* as its 2ODC genes. Furthermore, Pdc1 takes pyruvate and 2-oxo butanoate as examples of linear 2OA substrates and Aro10 takes aromatic, branched chain and sulphur-containing substrates. Although Pdc1 also showed activity for Aro10 substrates, and unlike *S. cerevisiae* according to reports by Vuralhan and colleagues (Vuralhan *et al.*, 2005)*, K. marxianus* Aro10 showed activity for Pdc1 substrates, the kinetics of both enzymes was unable to fit on Michaelis-Menten plot for the respective substrates (figure S2). Although they clustered with Thi3 proteins from other species and show synteny with ScThi3, Thi3 proteins from Kluyveromyces, Lachancea and Eremothecium species are divergent (figure 2). Thi3 from Zygosaccharomyces and Torulaspora species clustered closely with ScThi3 and other members of the Saccharomyces lineage, suggesting that Thi3 in the lineage comes from ZT species. In line with this, although experimental data showed in this current study that *KmTHI3* is a bona fide *THI3* gene (figure 6), *ScTHI3* was unable to rescue *thi3* deletion strain of *K. marxianus* during growth on thiamine-free medium (figure S5), suggesting that ScThi3 may not be able to regulate thiamine biosynthetic genes in *K. marxianus*.

With *PDC1* disruption causing complete loss of PDC activity and ethanol production, no significant difference from wildtype was found in regard to these two phenotypes for *ARO10* disruption (figure 7). This shows that *PDC1* is the sole contributor of PDC activity during the growth of wildtype *K. marxianus* in glucose and ammonium. In agreement with previous reports (Bianchi, Tizzani and Frontal, 1996; Choo *et al.*, 2018), *PDC1* deletion strains of *K. marxianus* and *K. lactis* displayed similar growth rates as wildtypes in this study. However, they showed growth defect in that they formed lower biomass at the end of the cultivation than wildtypes (figure 7C and figure S4). This observation is corroborated by the fact that glucose consumption was poorer in *pdc1* than in wildtype (figure 7B). *K. marxianus* and *K. lactis* are respiratory yeasts and can grow by directing pyruvate to the TCA cycle in the absence of PDC activity. Since respiratory capacity is a factor of oxygen availability, the difference in our observations here and by others in regard to the growth profile of *pdc1* strains of Kluyveromyces yeasts is explainable by difference in aeration. Indeed, the growth of *pdc1 K. marxianus* is poorer in oxygen-limited condition as has been shown by others. This can also explain report by Choo et al. (2018) that *PDC1* disruption in *K. marxianus* KCTC17555 reduced extracellular glycerol production which is contrary to the glycerol accumulation observed here for NBRC1777 (figure 7B). Pyruvate decarboxylation is a major route for

maintaining cytosolic redox balance as it supplies acetaldehyde which is acted upon by alcohol dehydrogenases to produce ethanol, re-oxidising NADH to NAD+ in the process. In the absence of PDC activity, redox balance is maintained by respiration when oxygen is available and by glycerol-3-phosphate reaction during oxygen limitation or anaerobiosis (Weusthuis *et al.*, 1994). The latter produces glycerol. Thus, the possibly higher aeration by others than in our work here as opposed to difference in strain types is a likely explanation for why we observed glycerol accumulation and they did not. This is particularly relevant in the potential application of *K. marxianus* as yeast cells factories for producing heterologous products. It may be feasible to reduce ethanol production, which can hamper product yield, by deleting *PDC1* and drive respiration through ample aeration in order to maintain redox balance without directing flux towards glycerol production.

Being a transcriptional regulator that is involved in glucose sensing, in-frame internal deletion in *MTH1* (*MTH1-ΔT*) was reported to reduced glucose uptake and restored the ability of PDC-null *S. cerevisiae* to grow on glucose (Oud *et al.*, 2012). The deleted segment is required for Mth1 degradation and reduction in Mth1 degradation is known to decrease the transcription of hexose transporter genes (Ozcan and Johnston, 1995). Due to this, coupled with the fact that PDC-null *S. cerevisiae* that was evolved to grow on glucose displayed reduced transcription of several of those transporter genes (Van Maris *et al.*, 2004), it is thought that *MTH1-ΔT* restricts glycolytic flux and prevents pyruvate accumulation to toxic levels. The two major functions of alcoholic fermentation are the rapid reoxidation of NADH formed in glycolysis back into NAD+ and the supply of cytosolic acetyl-CoA. The combination of lack of PDC activity, unrestricted glycolytic flux, lack of cytosolic acetyl-CoA and limited capacity of the mitochondrial respiratory chain for reoxidation of cytosolic NADH can reduce cytosolic NAD+ and can cause growth defect or complete growth loss as in the case of *S. cerevisiae.* Subsequent disruption of *ACH1* encoding acetyl-CoA transferase which releases the acetyl group of mitochondrial acetyl-CoA and transfers CoA onto acceptor molecules such as succinate (Fleck and Brock, 2009) caused *MTH1-ΔT* PDC-null *S. cerevisiae* to lose its ability to grow on glucose (Chen *et al.*, 2015). This indicates that the growth phenotype conferred by *MTH1-ΔT* is related to improved cytosolic acetyl-CoA supply. It is possible that *MTH1-ΔT*-induced reduction in glucose uptake relieved glucose catabolite repression of pyruvate dehydrogenase complex (PDH) which is known to be in effect for at least the gene encoding dihydrolipoamide dehydrogenase component of the complex, *LPD1* (Bowman *et al.*, 1992), and maybe even the TCA cycle. This will increase pyruvate metabolism via for oxidative decarboxylation into acetyl-CoA by PDH, thereby increasing the availability of mitochondrial acetyl-CoA. Ach1 can free up the acetyl group of mitochondrial acetyl-CoA for export to the cytosol where it can be converted into cytosolic acetyl-CoA by acetyl-CoA synthetase. We

show here that PDC-null *K. marxianus* displays slower glucose consumption (Figure 7B) with no pyruvate accumulation detected. This indicates that the strain is unlikely to be in a condition of glycolytic flux overdrive. In fact evidence by others showed that glycolysis plays a minor role in the growth of *K. marxianus* and that growth is conferred by respiration based on poor or loss of growth when the strain's respiration is compromised on complex medium (Choo ***et al.***, 2018). Therefore, it is likely that cytosolic acetyl-CoA is sourced from the mitochondrial pool in PDC-null *K. marxianus*, mediated possibly through the action of *ACH1* but alternative mediators such as threonine aldolase Gly1 and the carnitine-shuttle exist. An interesting study would be the disruption of these mediators to determine which of them increases the growth defect of PDC-null *K. marxianus* and *K. lactis*.

In vivo phenylpyruvate decarboxylase activity was not completely eradicated when *ARO10* was disrupted but required the additional disruption of *PDC1* to reach this phenotype (figure 8), showing that although Pdc1 is not as efficient a phenylpyruvate decarboxylase enzyme as Aro10, its native expression is sufficient to provide the activity in *K. marxianus*. Transcriptional analysis of both genes showed that similar to reports in *S. cerevisiae, ARO10* transcription is induced by phenylalanine as an example of aromatic amino acids and that the transcription of *PDC1* was not different in phenylalanine compared to ammonium, or even when *ARO10* is disrupted. This is relevant in engineering *K. marxianus* for the production aromatic compounds from aromatic amino acid precursors which requires the attenuation of phenylpyruvate decarboxylase activity as production could benefit not only from disrupting *ARO10* but also from additional disruption of *PDC1.*

Supplementary materials

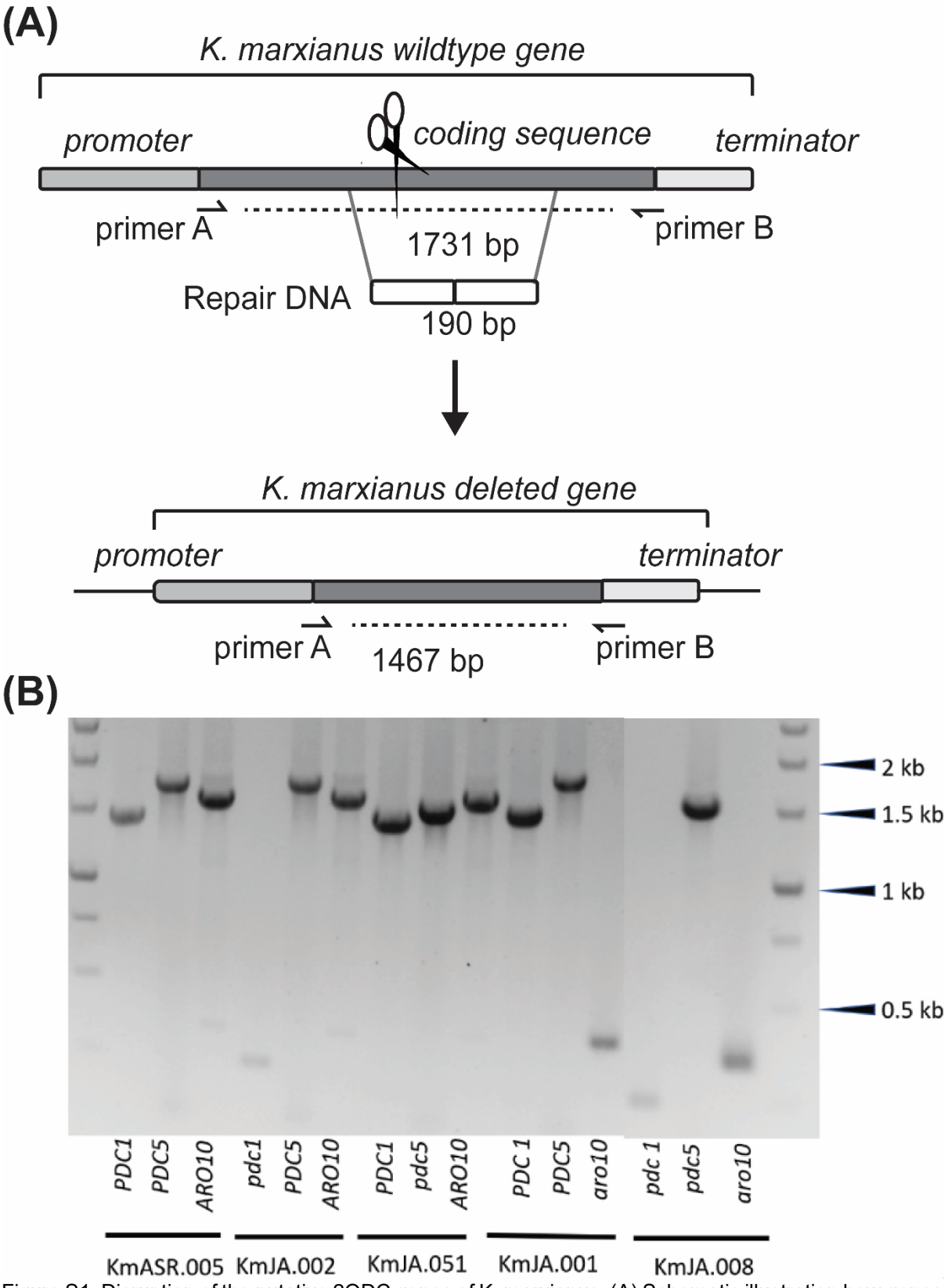
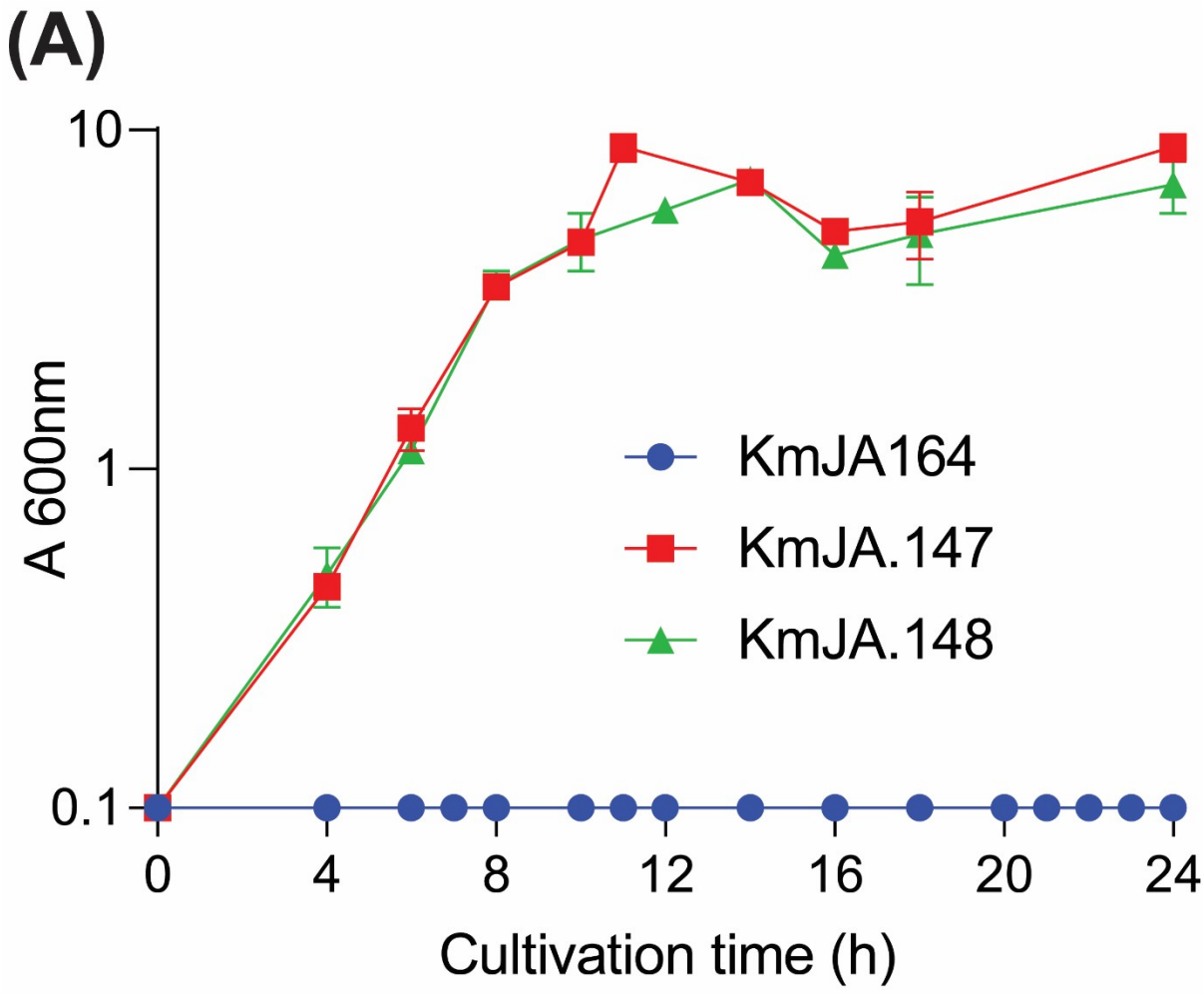


Figure S1. Disruption of the putative 2ODC genes of K. marxianus. (A) Schematic illustration how genes were disrupted in this work. Double strand breakage created by CRIPSR at a gene locus was repaired by homology directed repair with a short 190bp repair DNA, leading to the deletion of a segment and

loss of function of the gene. Gene deletions were confirmed by PCR with primers complementary to sequences within a gene, but outside the deleted segment.



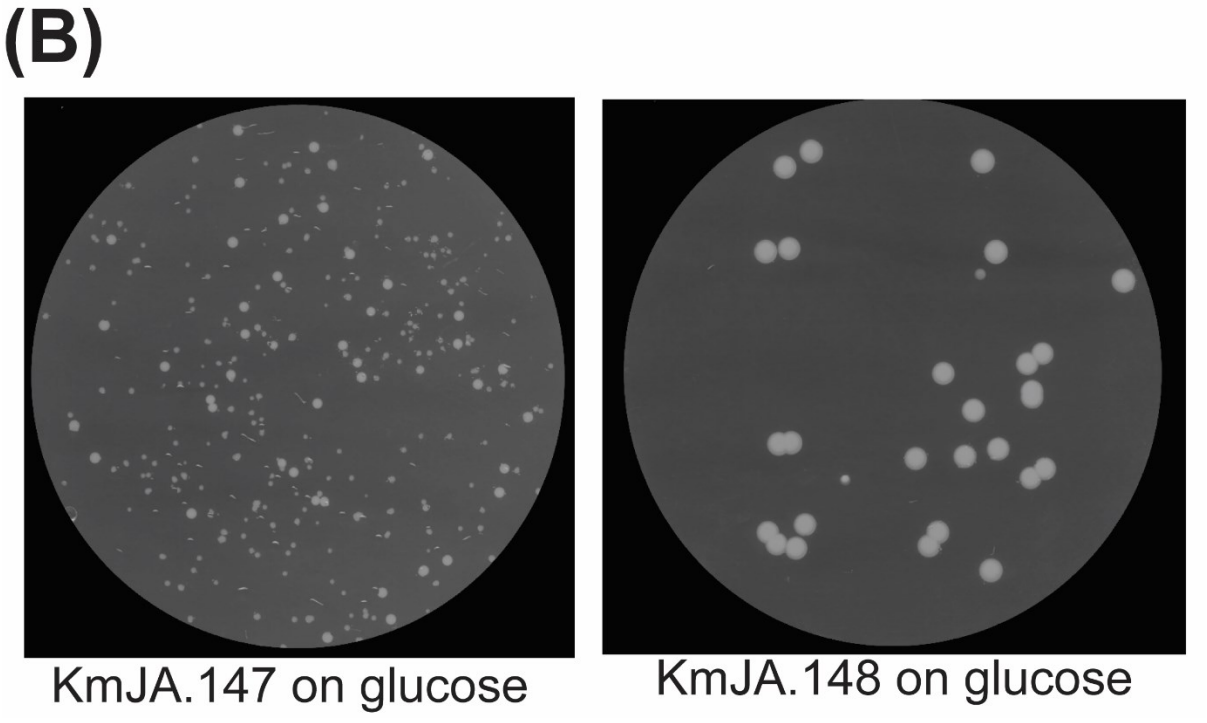


Figure S2. KmJA.164, which harbours disruptions of the three putative *K. marxianus* 2-oxo acid decarboxylase genes, was used to inoculate liquid minimal glucose medium (MGM). After 250 hours of cultivation, the culture was spread on solid MGM from which small and large colonies emerged. One small colony (KmJA.147) and a large colony (KmJA.148) were used to inoculate liquid minimal ethanol medium after which the cultures were used to inoculate liquid MGM (A) and spread again on solid MGM (B). Data shown for A 600nm are mean ± standard deviation from three independent experiments.

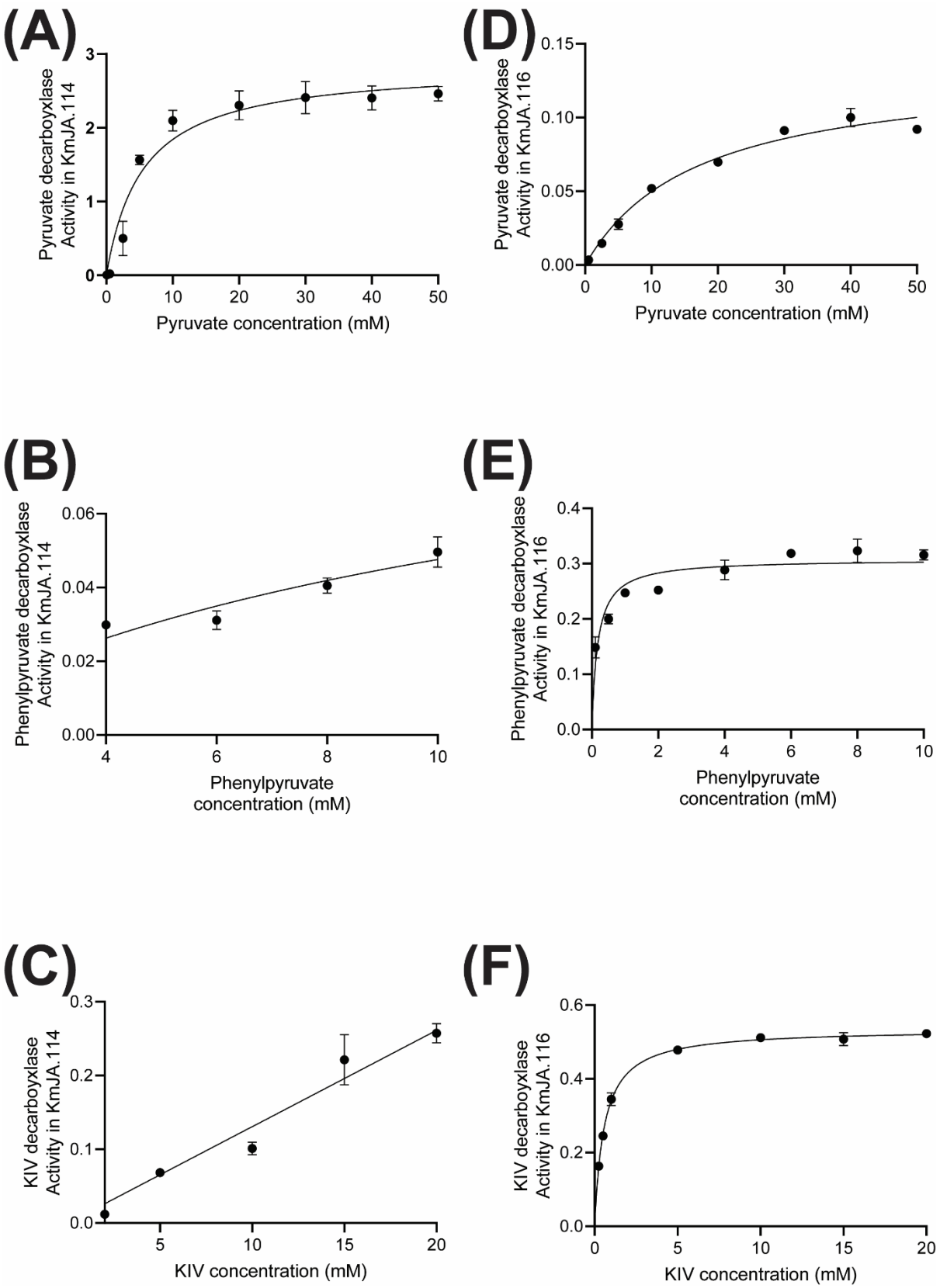


Figure S3. Michaelis-Menten fitting of kinetics of with cell extracts from KmJA.114 (*PDC1*) and KmJA.116 (*ARO10*). Data points are duplicates of specific activities at each concentration.

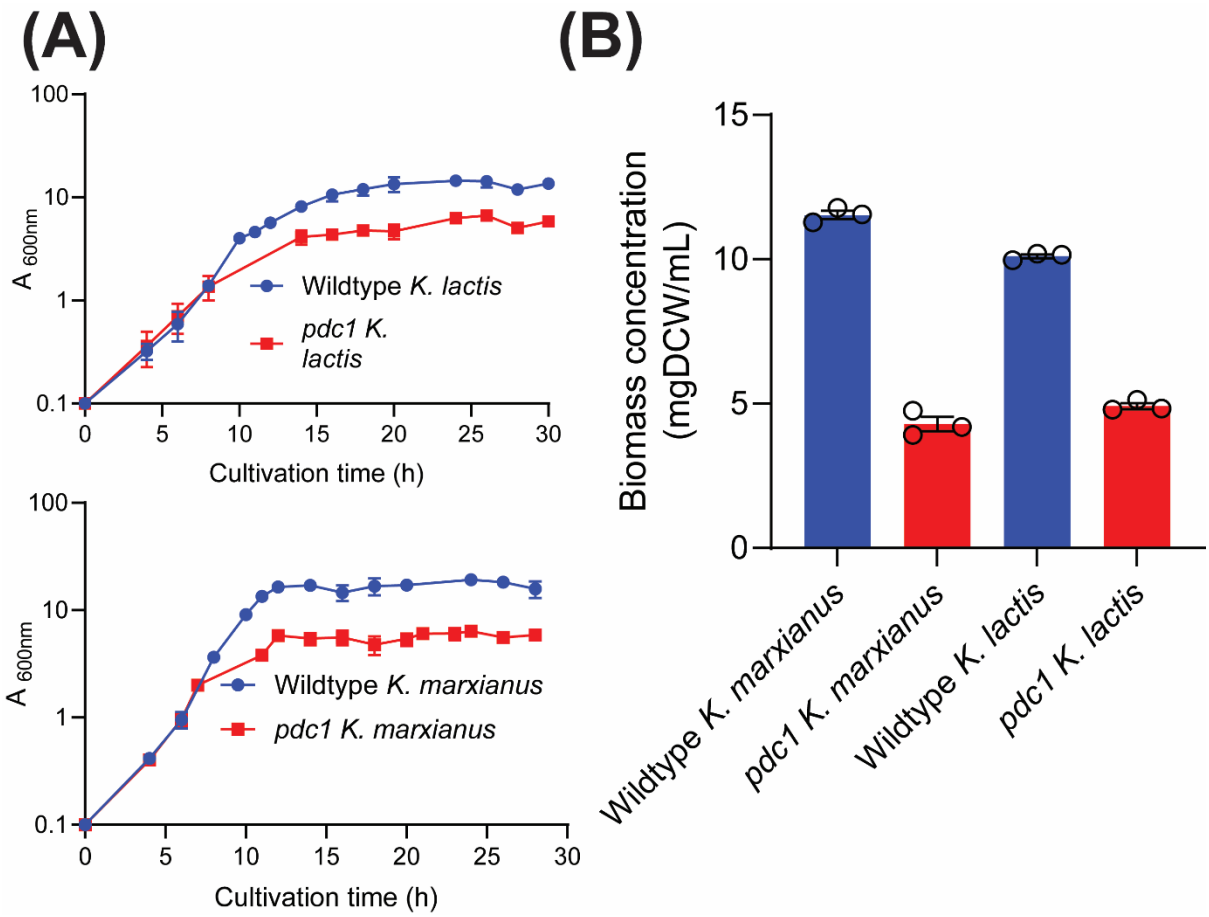


Figure S4. Wildtype *K. marxianus* and *K. lactis* display growth defect when *PDC1* is disrupted (A) *PDC1* was disrupted in wildtype strains of *K. marxianus* and *K. lactis* and the generated strains were grown on minimal glucose medium. (B) *PDC1* deletion and wildtype strains of *K. lactis* and *K. marxianus* were grown on YPD complex medium and biomass concentrations (dry cell weights/mL) were determined. Data shown are mean ± standard deviation from three independent experiments.

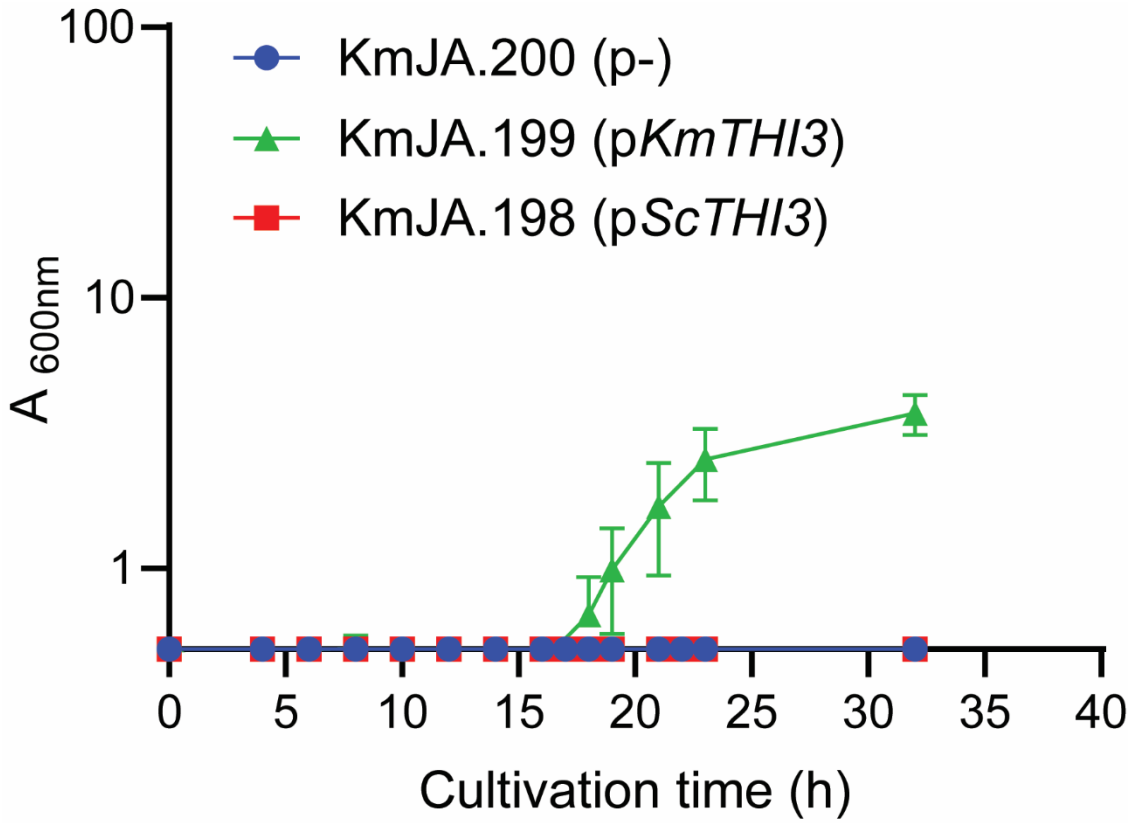


Figure S5. *ScTHI3* was unable to revive *thi3* strain of *K. marxianus* during growth in thiamine-free medium. Plasmids harbouring *ScTHI3* or *KmTHI3* for expression under the control of *KmPDC1* promoter or only the plasmid backbone were transformed into a *thi3* strain of *K. marxianus*. The three strains were then cultivated in thiamine-free medium.

Table S1. Primers used in this study.

Primer Name	Sequence	Application
JA603	CGTCTCACCAAAAGGGAGTTAGAATCATTTTGAATAAAAAA	Forward primer for amplifying pUCC001 gRNA transcriptional unit and for confirming correct insertion of protospacer sequences in pJA_U $gRNA_{empty}$
JA612	CGTCTCATGCTCCAAAACCTTCTCAAGCAAGG	Reverse primer for amplifying pUCC001 gRNA transcriptional unit
JA192	CGTCCCAGCCTTCGTTACCCCAAT	Sense oligonucleotide for *PDC1* protospacer sequence
JA193	AAACATTGGGGTAACGAAGGCTGG	Anti-sense oligonucleotide for *PDC1* protospacer sequence
JA92	CGTCCATGTTGCAAAACTTCCTTG	Sense oligonucleotide for *THI3* protospacer sequence
JA93	AAACCAAGGAAGTTTTGCAACATG	Anti-sense oligonucleotide for *THI3* protospacer sequence
JA202	CGTCAACCAAGTGGTGAACATTGT	Sense oligonucleotide for *ARO10* protospacer sequence
JA203	AAACACAATGTTCACCACTTGGTT	Anti-sense oligonucleotide for *ARO10* protospacer sequence
JA582	CGTCAGGAAGTTACGAAAGAACCT	Sense oligonucleotide for *KmURA3*
JA583	AAACAGGTTCTTTCGTAACTTCCT	Anti-sense oligonucleotide for *KmURA3* protospacer sequence
JA190	TTGTTGGTGTTCCATCCATCTCTTCCCAAGCTAAGCAATTGT TGTTGCACCACACCTTGGGTAACGGTGACTTCACTGT TTTCCACAGAATGTCTTCTAAGATCAGATTG	Forward primer for construction of *PDC1* DNA repair fragment
JA191	TTGGCGTTGATGGAAGCAGTCAATTGAGCTTGCTT GACCAAGTTGGATGGAGCATCCATGACTGGCAA CATGACTTCGATCAATCTGATCTTAGAAGACATT CTGTGGAA	Reverse primer for construction of *PDC1* DNA repair fragment
JA100	GATGTTCCTATTCATGAGATTCTTCCTCAGCTTATAGCAAAAATTGA CGCATCAAAGCTCTCTAATCTGTACGCCGTCACAGTTCCCGATATG GTGGGGAAGCTGGACAA	Forward primer for construction of *THI3* DNA repair fragment
JA101	AACCTTGAGTGAGAATGGTACTCAATTCTT GGAAAGTGAATTGTAGCGAACCATCACCT ACGATCAAAATGATCCGATGCTTGTCCAG CTTCCCCACCATATCGGGAACT	Reverse primer for construction of *THI3* DNA repair fragment
JA198	CTTGAATAGTCATTTGAGCAGCACCAT CACCTTCACACAAGATCAACTTTGGAA	Forward primer for construction of *ARO10* DNA repair fragment

	CATAGTCTGCTGGGACAGCAGATTGGT TGATAATGTGAGATTTCTAGCAAAGGC AA	
JA 199	TCACTCTTGGTCGTTATGTGTTCGA GAGATTGCTCAACTGTGGTTCCAA GACCATTTTCGGTGTTCCAGGTGA CTTCAACTTGCCTTTGCTAGAAAT CTCACATTATCAA	Reverse primer for construction of *KLMA_20597* DNA repair fragment
JA63	AAACAGCAGAGTTGTTAAGATTAGTTGAGGTTTT GGGTCCATATATCTGTCTATTGAAGACACATGTA GATATCTTGGAGGATTTCAGCTTTGAGTTTGGGA CAACAATA	Forward primer for construction of *KmURA3* DNA repair fragment
JA64	GATCTCTTCCCTTTGCGAAAAGACCTCTACCAAC AATAATGATGTCTGATCCACCGGCAACAACTTC ATCCACAGTTCTGTATTGTTGTCCCAAACTCAAA GCTGAAATC	Reverse primer for construction of *KmURA3* DNA repair fragment
JA206	TTGTCTGCCTTGAACGGTATT	Forward primer for confirming *PDC1* deletion
JA207	TTGATGGAAGCAGTCAATTGA	Reverse primer for confirming *PDC1*deletion
JA102	GTATGCTGATAGGTACAACTTGG	Forward primer for confirming *THI3* deletion
JA103	CTAAAGTCTTACTCTTTTGCTTTGAGG	Reverse primer for confirming *THI3* deletion
JA204	AGTGTAACCATTGTTGTTGAA	Forward primer for confirming *ARO10* deletion
JA205	CCAGTAGTTCTAGACGATAAATC	Reverse primer for confirming *ARO10* deletion
JA80	GGAAAGAGCAGCTGCTCATAGAAGT	Forward primer for confirming *KmURA3* deletion
JA81	CGGCAACAACTTCATCCACAGTTC	Reverse primer for confirming *KmURA3* deletion
JA623	GGCTGAACGTGGTTACTCCT	Forward primers for RT-qPCR of *ACT1* housekeeping gene
JA624		Reverse primers for RT-qPCR of *ACT1* housekeeping gene
JA625	CCATCTCAGGATCCCTCTTC	Forward primers for RT-qPCR of *ALG9* housekeeping gene
JA626	CGATTCCAGCGAATAGTTGA	Reverse primers for RT-qPCR of *ALG9* housekeeping gene
JA627	ACCACATTTGGTGTCGGTGA	Forward primers for RT-qPCR of *ARO10*
JA628	ACCGACAATGTGCAAGACCT	Reverse primers for RT-qPCR of *ARO10*
JA504	AAGTCGTGTGCCTGGGTAAC	Forward primers for RT-qPCR of *THI4*
JA505	TGAGGGCACTTATGTGGTTG	Reverse primers for RT-qPCR of *THI4*

JA508	TGAAGCCGTCAGATGTGGTA	Forward primers for RT-qPCR of *THI11*
JA509	CAAACAAGCGAGCTCATCAA	Reverse primers for RT-qPCR of *THI11*
JA194	CGTCCACTTCCCAAATGACACCTA	Forward primers for RT-qPCR of *PDC1*
JA197	AAACTCTGATCATGGTGGAGATTT	Reverse primers for RT-qPCR of *PDC1*

Table S2. CRISPR targets used in this study.

Gene	**Target sequence**	**Source**
KmTHI3	GAATCCGGTACATCCGCGAT	Work in chapter 3
KmARO10	GTCCAAAGAAGACTTAGCCA	Work in chapter 3
PDC1	CCAGCCTTCGTTACCCCAAT	This work
KmURA3	AGGAAGTTACGAAAGAACCT	(Rajkumar et al. 2019)
KmTHI4	GTTAGTAACCACACCTGCAG	This work

Chapter 5. Derivatisation of coumaric acid into naringenin in *Kluyveromyces marxianus*

Joel A. Akinola, John P. Morrissey.

The data in this chapter are being prepared for inclusion in a manuscript

JAA performed all experiments, interpreted the data and prepared the manuscript

JPM provided supervision

Abstract

Having reported the successful engineering of *K. marxianus* for the production of coumaric acid in chapter 3, we tried to derivatise coumaric acid into specific phenylpropanoids in this study. Flavonoids are highly valuable phenylpropanoids that are produced from an intermediate flavonoid called naringenin. In this work, the expression of singles copies of the genes encoding the six enzymes required for naringenin biosynthesis from phenylalanine, including phenylalanine ammonia lyase (Pal), cinnamic acid 4-hydroxylases and cytochrome P450 reductase (Cpr), 4-coumaric acid:coenzyme A ligase (4Cl), chalcone synthase (Chs) and chalcone isomerase (Chi) in strains that are pre-engineered for increased aromatic amino acids biosynthesis did not result in detectable naringenin production until the copy number of the Chs gene was increased. However, coumaric acid production persisted, showing that this precursor was only partially depleted and that there may be bottlenecks below coumaric acid in the naringenin pathway. To further deplete coumaric acid into naringenin, a gradual approach was applied by first engineering coumaric acid production in the yeast and then additionally expressing only the gene encoding 4Cl, the enzyme responsible for converting coumaric acid into the metabolite immediately below it, coumaroyl-CoA. This resulted in the loss of coumaric acid production, showing that it was consumed by 4Cl. However, when Chs and Chi were additionally expressed, coumaric acid was not consumed. This hinted that the reactions catalysed by 4Cl and Chs are bottlenecks in the pathway. To address these bottlenecks, increased biosynthesis of coenzyme A which is required for both reactions and malonyl-CoA of which 3 molecules are required for the Chs reaction were engineered in addition to multi-copy integration of 4Cl and Chs. Interestingly, the overexpression of acetyl-CoA synthase and acetyl-CoA carboxylase by themselves to increase malonyl-CoA biosynthesis did not improve naringenin production but their co-expression with coenzyme A biosynthetic genes did. Production was then further improved by multi-copy 4Cl and by multi-loci integration of Chs at delta site sequences distributed around the yeast's genome, resulting in a strain producing the highest yield (1231µg/g glucose) of naringenin in this work.

Introduction

Coumaric acid is a branch point metabolite that serves as an intermediate from which many aromatic molecules belonging to the phenylpropanoid family of compounds are biosynthesised in plants. Thus, phenylpropanoids are characterised by the phenyl (6 carbon) ring and the three carbon aliphatic chain of their coumaric acid precursor. They are structurally

classified into flavonoids, monolignols, phenolic acids, stilbenes and coumarins. In particular, Flavonoids are highly useful molecules in the food, agrochemical, pharmaceutical and nutraceutical industries and are commercially produced through extraction from plants materials and by chemical synthesis. Limited plant materials for which large amounts are required for extraction, and difficult downstream processing are the major challenges in the commercial production of flavonoids and plant chemicals in general. On the other hand, the complexity of many valuable flavonoids complicates their chemical synthesis, causing production to require multiple steps and toxic chemicals in addition to the production of toxic byproducts that are harmful to the environment (Milke *et al.*, 2018). Given that the pathways for the biosynthesis of many flavonoids, the precise enzymes that catalyse the reactions and regulation of the pathways in plants have been mostly elucidated, the combination of plant metabolic engineering and tissue culture can potentially increase the production of flavonoids (Chandran *et al.*, 2020). However, this will still require difficult downstream processing in order to obtain pure individual flavonoids. In plants, flavonoids are produced from a derivative of coumaric acid called naringenin. In the naringenin biosynthetic pathway (figure 1), coumaric acid is activated by 4-coumaric acid:coenzyme A ligase (4Cl) to form coumaroyl-CoA which can be condensed with 3 molecules of malonyl-CoA through the action of the type III polyketide synthase, chalcone synthase (Chs), to form an open ring isoform of naringenin (naringenin chalcone). The isomerisation of naringenin chalcone by chalcone isomerase (Chi) closes the ring to form naringenin flavanone from which other flavonoids are produced.

In recent years, microorganisms have emerged as suitable cell factories for the production of flavonoids and other plant molecules because they do not naturally produce many plant molecules, allowing them to be engineered through the heterologous expression of the enzymes that are involved in the biosynthetic pathway for specific molecules. Thus, production in microbial cell factories can make downstream processing easier due to the more focused nature of the production than in plants. Although bacteria have been engineered to produce some flavonoids, yeasts are more suitable cell factories for producing them because the biosynthesis of many flavonoids require cytochrome P450 enzymes which function poorly or not at all in bacteria due to their lack of membrane bound organelles such as the endoplasmic reticulum (Cao *et al.*, 2020; Liu *et al.*, 2020). The availability of genetic engineering tools, detailed knowledge of its physiology and its widespread usage for commercial biotechnological applications have made *S. cerevisiae* the go-to yeast for metabolic engineering, but some non-saccharomyces yeasts like *Yarrowia lipolytica* and *Kluyveromyces marxianus* are now being engineered in order to avail of the physiological properties that are present in them but not in *S. cerevisiae*. For example, the superior ability of *Yarrowia lipolytica* to biosynthesise fatty acids which are derived from malonyl-CoA, a limiting metabolite in the

biosynthesis of some valuable plant flavonoids has resulted in recent improvements in flavonoid production in yeasts (Madzak, 2021). The ability of *K. marxianus* to metabolise more diverse substrates, to grow faster on glucose than *S. cerevisiae* and thrive at high temperatures makes this yeast attractive for commercial production (Lane *et al.*, 2011; Lertwattanasakul *et al.*, 2015; Morrissey *et al.*, 2015) .

We previously reported the engineering of *K. marxianus* for *de novo* biosynthesis of coumaric acid at a yield of over 40mg/g glucose in batch bioreactors with minimal glucose medium in chapter 3 based on heterologous overexpression of phenylalanine ammonia lyase from *Photorhabdus luminescens* (*PlPAL*) and cinnamic acid 4-hydroxylase and cytochrome P450 reductase from *Arabidopsis thaliana* (*AtC4H* and *AtCPR1*) (the coumaric acid module) in addition to an improved aromatic amino acid (AAA) precursor supply. It should be possible to further convert the coumaric acid produced in this yeast to naringenin by expressing genes encoding 4Cl, Chs and Chi in addition to the coumaric acid module as has been demonstrated in *S. cerevisiae* (Koopman *et al.*, 2012) and *Y. lipolytica* (Lv *et al.*, 2019). However, the condensation of 4-coumaroyl-CoA and malonyl-CoA is a bottleneck reaction in the pathway in plants (Knogge, Schmelzer and Weissenböck, 1986), is also well-known in yeasts and could be due to limited malonyl-CoA for which 3 molecules are required for every naringenin molecule formed (figure 1). The high requirement of malonyl-CoA for proper cell function makes it a highly expensive metabolite and Chs must compete against reactions in native pathways that may have better catalysis and efficiently utilise malonyl-CoA such as in lipid synthesis. The provision of multiple copies of Chs genes has been demonstrated as a strategy for successfully alleviating this bottleneck in yeasts (Koopman *et al.*, 2012; Lv *et al.*, 2019; Cao *et al.*, 2020). While this strategy theoretically increases the abundance of Chs enzyme molecules that is intracellularly available for simultaneously catalysing the bioconversion at any given time and for improving the reaction's competition for available malonyl-CoA, this strategy can reduce the intracellular malonyl-CoA molecules that are available to native reactions. Malonyl-CoA biosynthetic genes can be expressed in addition to the naringenin pathway in order to increase its biosynthesis, resulting in higher abundance of intracellular malonyl-CoA for the Chs and native reactions. Malonyl-CoA is produced from acetyl-CoA by the action of yeast native acetyl-CoA carboxylase (Acc1) and the overexpression of its wildtype or mutant versions by themselves or in addition to acetyl-CoA synthase (Acs) which converts acetate to acetyl-CoA was demonstrated to increase naringenin production in *Y. lipolytica* (Lv *et al.*, 2019). The Coenzyme A part of acetyl and malonyl-CoA biosynthesis has not received a lot of attention but could be a limiting metabolite for both molecules and also naringenin. Indeed, a study showed that the engineering of Coenzyme A biosynthesis by overexpressing the first enzyme in the coenzyme A biosynthetic pathway, pantothenate kinase

PanK, increased naringenin production, pointing at coenzyme availability as a bottleneck of naringenin biosynthesis in *S. cerevisiae* (Liu, Zhang and Jiang, 2017)

In this current study, we engineered *K. marxianus* for increased biosynthesis of AAAs by increasing flux through the shikimate pathway and by overexpressing the genes that positively contribute to the biosynthesis of the precursors of this pathway. We also attempted for the first time to derivatise coumaric acid to naringenin in *K. marxianus* by further expressing *At4-CL3*, *AtCHS3* and *AtCHI* in coumaric-acid producing strains. To increase malonyl-CoA availability, the native *ACC1* and mutant *Salmonella enterica acs* genes were overexpressed from strong constitutive promoters. The entire coenzyme A biosynthetic pathway genes were also overexpressed in order to increase the abundance of coenzyme A.

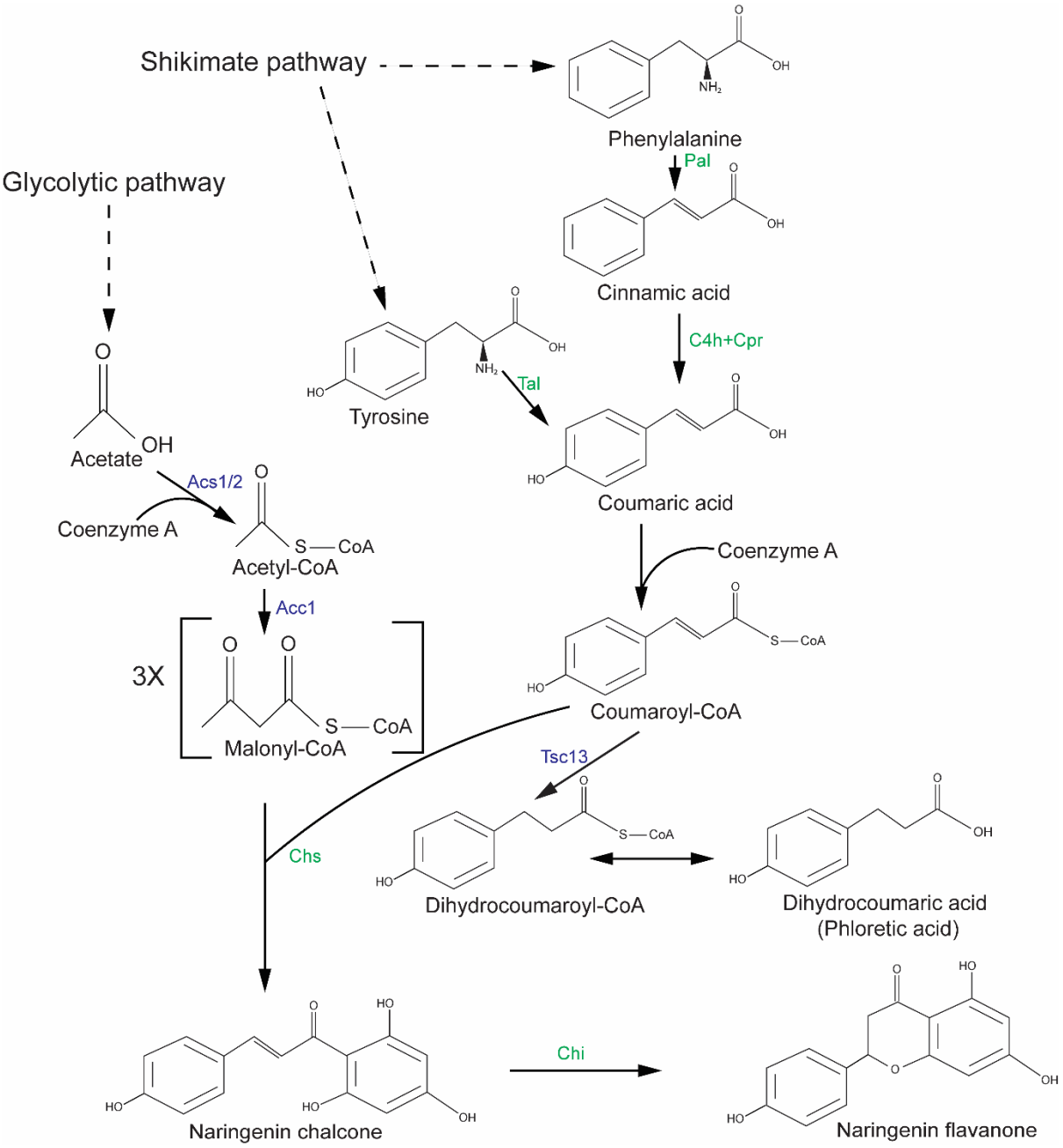


Figure 1. Biosynthetic pathway of naringenin. Phenylalanine and tyrosine are the precursors of naringenin in plants with phenylalanine ammonia lyase (Pal), cinnamic acid 4-hydroxylase and cytochrome P450 reductase acting on phenylalanine or tyrosine ammonia lyase (Tal) upon tyrosine to produce coumaric acid. Coumaric acid is then converted into coumaroyl-CoA by 4-coumaric acid:coenzyme A ligase (4Cl). Chalcone synthase (Chs) catalyses the decarboxylative condensation of coumaroyl-CoA and 3 molecules of malonyl-CoA into naringenin chalcone. Chalcone isomerase then closes the open ring structure of the chalcone to form the final naringenin flavanone. Coumaroyl-CoA can also be converted by the yeast native enoyl-CoA reductase (Tsc13) into dihydrocoumaroyl-CoA from which phloretic acid is produced. Thus, the production of phloretic acid competes against naringenin production for coumaroyl-CoA. The central carbon metabolite, acetate, and coenzyme A are important in the pathway because they are the precursors that are required for malonyl-CoA biosynthesis, and coenzymes A is additionally required for the formation of coumaroyl-CoA. Acetate is converted into acetyl-CoA by acetyl-CoA synthase (Acs1/2). Acetyl-CoA is then further converted into malonyl-CoA by acetyl-CoA carboxylase (Acc1). The enzymes in green texts from plants, are not found in yeasts and require heterologous expression for naringenin production, while those in blue are native to yeasts and can be overexpressed for naringenin production.

Materials and method

1. Strains and growth conditions

Kluyveromyces marxianus strain NBRC1777 from the Biological Resource Centre, NITE (NBRC; Tokyo Japan) was used as background in this study. All yeast strains used in this study are listed in Table 1. Yeasts were routinely cultured at 30^{o}C in YPD broth (2% yeast extract, 1% peptone and 2% glucose) or in appropriately supplemented minimal glucose medium (5g/L $(NH_4)_2SO_4$, 3g/L KH_2PO_4, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 15g/L glucose, and filter-sterilized vitamins and trace elements) brought to pH 5.5-6 with KOH(Verduyn *et al.*, 1992). For yeast growth on solid medium, 2% agar supplementation was applied. Nutritional reagents and agar were purchased from Formedium, Hunstanton, UK. Bacterial transformation was with *E. coli* DH5α grown in LB broth (1% NaCl, 1% peptone, 0.5% yeast extract) or LB supplemented with 1.5% agar and appropriate antibiotic (50mg/L chloramphenicol, 100mg/L ampicillin or 50mg/L kanamycin; Fisher Scientific (Dublin, Ireland)).

Table 1. Strains used in this study.

Strain	Genotype	Source
KmASR.047	*dnl4 aro4 aro7 lac4Δ::KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmTDH3pr-KmRPE1-mPDC1t KmTSA1pr-KmRKI1-ScPGK1t aro3Δ::KmPDC1pr-KmARO4^{fbr}-ScADH1t KmENO1pr-ARO1-KMXK_A03020t KmPGK1pr-ARO2-KmLAC4t KmTEF1pr-KmARO7^{fbr}-KmPDC1t KmENO1pr-KmPHA2-KmINU1t KmHSP150pr-KmARO9-ScPGK1t*	(Rajkumar and Morrissey, 2020)
KmJA.112	*Ku80*	Work in chapter 2
KmJA.019	KmJA.112, *his3*	This study
KmJA.285	KmJA.019, I3::*KmPDC1pr-KmARO4^{fbr}-ScADH1t*	This study
KmJA.293	KmJA.285, *lac4Δ::KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t*	This study
KmJA.294	KmJA.285, *lac4Δ::KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t*	This study
KmJA.297	KmJA.285, *lac4Δ::KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t KmPDC1-KmTRP2(S64R)-ScPGK1t*	This study
KmJA.298	KmJA.285, *lac4Δ::KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t KmPDC1-KmTRP2(S74L)-ScPGK1t*	This study
KmJA.299	KmJA.285, *lac4Δ::KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t KmPDC1-KmTRP2(S64R, S74L)-ScPGK1t*	This study
KmJA.300	KmJA.285, *lac4Δ::KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t KmPDC1-KmGLN1-ScPGK1t KmPDC1-KmTRP2(S64R)-ScPGK1t*	This study
KmJA.026	KmASR.047, *leu2*	This work
KmJA.033	KmJA.026, I7/*ARO10*::*KmPDC1-PlPAL-ScPGK1t*	
KmJA.313	KmJA.033, I3::*KmPDC1-ScFEN2-ScPGK1t* 5X (*KmTDH3pr-At4-CL3-PDC1t*)	
KmJA.159	KmJA.026, I4::*KmTDH3pr-At4-CL3-PDC1t KmPDC1pr-PlPAL-ScPGK1t KmPDC1pr-AtCHI1-INU1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work

KmJA.168	KmJA.159, I3::*KmTEF1pr-AtCHS3- KMXK_A03020t*	This work
KmJA.169	KmJA.159, I3::2X (*KmTEF1pr-AtCHS3-KMXK_A03020t*)	This work
KmJA.170	KmJA.159, I3::3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*)	This work
KmJA.233	KmJA.159, I3::3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*) *KmPDC1pr-KmCAB1-KmPGK1t*	This work
KmJA.173	KmASR.047, I4::*KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work
KmJA.174	KmASR.047, I4::*KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t KmTDH3pr-At4-CL3-PDC1t*	This work
KmJA.025	KmJA.112, *ura3his3leu2*	This work
KmJA.247	KmJA.025, I6::*KmPDC1pr-KmARO4^{fbr}-ScADH1t KmENO1pr-KmARO1-KMXK_A03020t KmPGK1pr-ARO2-KmLAC4t KmTEF1pr-KmARO7^{fbr}-KmPDC1t KmENO1pr-KmPHA2-KmINU1t KmHSP150pr-KmARO9-ScPGK1t*	This work
KmJA.252	KmJA.247, I3::3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*)	This work
KmJA.253	KmJA.252, I4::*KmTDH3pr-At4-CL3-PDC1t KmPDC1pr-PlPAL-ScPGK1t KmPDC1pr-AtCHI1-INU1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work
KmJA.160	KmJA.025, I4::*KmTDH3pr-At4-CL3-PDC1t KmPDC1pr-PlPAL-ScPGK1t KmPDC1pr-AtCHI1-INU1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work
KmJA.162	KmJA.160, *aro10*	This work
KmJA.212	KmJA.162, I6::*KmPDC1pr-KmARO4^{fbr}-ScADH1t KmENO1pr-KmARO1-KMXK_A03020t KmPGK1pr-ARO2-KmLAC4t KmTEF1pr-KmARO7^{fbr}-KmPDC1t KmENO1pr-KmPHA2-KmINU1t KmHSP150pr-KmARO9-ScPGK1t*	This work
KmJA.225	KmJA.212, I3::3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*)	This work
KmJA.220	KmJA.212, I3::3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*) *KmPDC1pr-KmACC1-ScPGK1t*	This work
KmJA.234	KmJA.220, *lac4*::*KmPDC1pr-KmCAB1-ScPGK1t*	This work
KmJA.221	KmJA.212, I3::3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*) *KmPDC1pr-Seacs*-ScPGK1t*	This work
KmJA.235	KmJA.221, *lac4*::*KmPDC1pr-KmCAB1-ScPGK1t*	This work

KmJA.222	KmJA.212, I3::3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*) *KmPDC1pr-KmACC1-ScPGK1t KmPDC1pr-Seacs*-ScPGK1t*	This work
KmJA.236	KmJA.222, *lac4*::*KmPDC1pr-KmCAB1-ScPGK1t*	This work
KmJA.284	KmJA.222, *trp1*	This work
KmJA.287	KmJA.284, I3::5X (*KmTDH3pr-At4-CL3-PDC1t*)	This work
KmJA.245	KmJA.222, *lac4*::6X (*KmPDC1pr-KmCAB1-ScPGK1t*)	This work
KmJA.246	KmJA.222, *lac4::KmPDC1pr-ScCAB1-ScPGK1t*	This work
KmJA.254	KmJA.222, *lac4::KmPDC1pr-ScCAB1*-ScPGK1t*	This work
KmJA.273	KmJA.222, *lac4::KmPDC1pr-ScCAB1*-ScPGK1t KmPDC1pr-ScCAB2-ScPGK1t KmPDC1pr-ScCAB3-ScPGK1t KmPDC1pr-HAL3-ScPGK1t KmPDC1pr-ScCAB4-ScPGK1t KmPDC1pr-ScCAB5-ScPGK1t*	This work
KmJA.275	KmJA.273, I3::*KmPDC1pr-ScFEN2-ScPGK1t*	This work
KmJA.274	KmJA.222, *lac4::KmPDC1pr-ScCAB1*-ScPGK1t KmPDC1pr-ScCAB4-ScPGK1t KmPDC1pr-ScCAB5-ScPGK1t*	This work
KmJA.277	KmJA.273 *trp1*	This work
KmJA.288	KmJA.277, I3::5X (*KmTDH3pr-At4-CL3-PDC1t*)	This work
KmJA.282	KmJA.277, I3::*KmPDC1-ScFEN2-ScPGK1t* 5X (*KmTDH3pr-At4-CL3-PDC1t*)	This work
KmJA.333	KmJA.282 with *KmTEF1pr-AtCHS3- KMXK_A03020ter KmPDC1pr- Venus-KmINU1*ter *KmTEF1pr-AtCHS3- KMXK_A03020ter* integrated at delta site loci of its genome	This work

2. Plasmid construction

2.1. Construction of plasmids for gene expression

All expression plasmids used in this study were integrative, constructed by Golden Gate assembly based on the hierarchical Yeast Toolkit standard (Lee *et al.*, 2015; Rajkumar *et al.*, 2019) and are listed in

Table 2. . Briefly, genes were sourced by Q5 High-Fidelity Polymerase (M4092L, New England Biolabs (NEB), Ipswich, UK) PCR amplification using primers bearing BsaI and BsmBI sites at their 5' and 3' ends or as gene blocks (Twist Biosciences, San Francisco, U.S.A.). All non-yeast native genes expressed in this study were codon optimised for *S. cerevisiae* and are listed on Table S1. Genes were then cloned by a BsmBI (R0739L, NEB)/T7 DNA ligase (M0318L, NEB) Golden Gate reaction into pYTK001 to construct level-I gene plasmids. When a gene was to be individually expressed, the level-I plasmid of that gene was put together with level-I promoter and terminator plasmids and an integrative dropout plasmid in a BsaI (R0535L, NEB)/T7 DNA ligase Golden Gate reaction to construct a level-II plasmid of the gene. The integrative dropout plasmids used were from previous work by Rajkumar and colleagues (Rajkumar *et al.*, 2019), and are listed in Table S2. To make the construction of level II plasmids easy, the ten level-II red florescence protein (RFP) dropout plasmids required for constructing level-III plasmids with up to six concatenated TUs that were constructed in our previous work were used in this study. For constructing the level-III plasmids with multiple concatenated transcriptional units (TU), the genes to be expressed from the plasmid were first individually cloned alongside the desired yeast promoters and terminators into level-II plasmids using BsaI/ T7 DNA ligase reactions to construct a set of level-II plasmids that individually harboured transcriptional units for each gene of interest. The level-II dropout plasmid in which a TU is placed is specified by the predetermined position for that TU in the level-III plasmid to be constructed. Once a set of level-II TU plasmids were constructed and confirmed, they were assembled and inserted into appropriate dropout plasmids by BsmBI/ T7 DNA ligase reaction to generate the level-III biosynthetic plasmid. Golden Gate reaction products were transformed into *E. coli* DH5α and selected on LB supplemented with chloramphenicol for level-I plasmids, ampicillin for level-II plasmids and kanamycin for level-III plasmids. The plasmids were confirmed by PCR using OneTaq Quick-Load Polymerase (M0486L, NEB) and sequencing for level I gene plasmids or NotI (FD0596, Fisher Scientific) digestion for level II and III plasmids. All supporting plasmids and primers used in this study are listed in Table S2 and Table S3, respectively.

Table 2. Expression plasmids used in this study.

Name	Description	Source
pI6-AAA/Phe	pI6-MTU-DO-URA with insert *KmPDC1pr-KmARO4^{fbr}-ScADH1t KmENO1pr-KmARO1-KMXK_A03020t KmPGK1pr-ARO2-KmLAC4t KmTEF1pr-KmARO7^{fbr}-KmPDC1t KmENO1pr-KmPHA2-KmINU1t KmHSP150pr-KmARO9-ScPGK1t*	(Rajkumar and Morrissey, 2020)
pJA_VM L2. 35	pI3 MTU-DO-G418 with insert *KmPDC1pr-KmARO4^{fbr}-ScADH1t*	This work
pJA_VM L3.16	pI1 MTU-DO-HIS with insert *KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t*	This work
pJA_VM L3.17	pI1 MTU-DO-HIS with insert *KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t*	This work
pJA_VM L3.28	pI1 MTU-DO-HIS with insert *KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t KmPDC1-KmTRP2(S64R)-ScPGK1t*	This work
pJA_VM L3.29	pI1 MTU-DO-HIS with insert *KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t KmPDC1-KmTRP2(S74L)-ScPGK1t*	This work
pJA_VM L3.30	pI1 MTU-DO-HIS with insert *KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t KmPDC1-KmTRP2(S64R, S74L)-ScPGK1t*	This work
pJA_VM L3.31	pI1 MTU-DO-HIS with insert *KmTEF1pr-KmTKL1-KmINU1t KmPGK1pr-KmTAL1- KMXK_A03020t KmSSA2pr-KmENO1-ScADH1t KmPDC1-KmGLN1-ScPGK1t KmPDC1-KmTRP2(S64R)-ScPGK1t*	This work
pJA_NM L3.52	pI4 MTU-DO-G418 with insert *KmTDH3pr-At4-CL3-PDC1t KmPDC1pr-PlPAL-ScPGK1t KmPDC1pr-AtCHI1-INU1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work
pJA_NM L2.37	pI3 MTU-DO-LEU with insert *KmTEF1pr-AtCHS3- KMXK_A03020t*	This work
pJA_NM L3.18	pI3 MTU-DO-LEU with insert 2X (*KmTEF1pr-AtCHS3- KMXK_A03020t*)	This work
pJA_NM L3.43	pI3 MTU-DO-LEU with insert 3X (*KmTEF1pr-AtCHS3- KMXK_A03020t*)	This work
pJA_NM L3.74	pI3 MTU-DO-LEU with insert 3X (*KmTEF1pr-AtCHS3- KMXK_A03020t*) *KmPDC1pr-KmCAB1-KmPGK1t*	
pJA_NM L3.55	pI4 MTU-DO-G418 with insert *KmPDC1pr-PlPAL-ScPGK1t KmENO1pr-AtC4H-KmPGK1t KmENO1pr-AtCPR1-KmPGK1t*	This work
pJA_NM L3.57	pI3 MTU-DO-LEU with insert 3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*) *KmPDC1pr-KmACC1-ScPGK1t*	This work
pJA_NM L3.58	pI3 MTU-DO-LEU with insert 3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*) *KmPDC1pr-Seacs*-ScPGK1t*	This work
pJA_NM L3.59	pI3 MTU-DO-LEU with insert 3X (*KmTEF1pr-AtCHS3-KMXK_A03020t*) *KmPDC1pr-KmACC1-ScPGK1t KmPDC1pr-Seacs*-ScPGK1t*	This work
pJA_NM L2.114	pI1 MTU-DO-HIS with insert *KmPDC1pr-ScCAB1*-ScPGK1t*	This work
pJA_NM L2. 24	pI7/*ARO10* MTU-DO-LEU with insert *KmPDC1-PlPAL-ScPGK1t*	Work in chapter 3
pJA_NM L3.80	pI1 MTU DO- HIS3 with insert *KmPDC1pr-ScCAB1*-ScPGK1t KmPDC1pr-ScCAB2-ScPGK1t KmPDC1pr-ScCAB3-ScPGK1t KmPDC1pr-HAL3-ScPGK1t KmPDC1pr-ScCAB4-ScPGK1t KmPDC1pr-ScCAB5-ScPGK1t*	This work
pJA_NM L3.75	pI1 MTU DO- HIS3 with insert *KmPDC1pr-KmCAB1-ScPGK1t KmPDC1pr- KmCAB1-ScPGK1t KmPDC1pr- KmCAB1-ScPGK1t KmPDC1pr- KmCAB1-ScPGK1t KmPDC1pr- KmCAB1-ScPGK1t*	This work
pJA_NM L3.82	pI1 MTU DO- HIS3 with insert *KmPDC1pr-ScCAB1*-ScPGK1t KmPDC1pr-ScCAB4-ScPGK1t KmPDC1pr-ScCAB5-ScPGK1t*	This work
pJA_NM L2.116	pI3 MTU-DO-Hph with insert *KmPDC1pr-ScFEN2-ScPGK1t*	This work
pJA_NM L3.88	pI3 MTU-DO-TRP1 with insert 5X (*KmTDH3pr-At4-CL3-PDC1t*)	This work
pJA_NM L3.84	pI3 MTU-DO-TRP1 with insert *KmPDC1-ScFEN2-ScPGK1t* 5X (*KmTDH3pr-At4-CL3-PDC1t*)	This work
pJA_NM L3.54	pIDelta MTU-DO with insert *KmTEF1pr-AtCHS3- KMXK_A03020ter KmPDC1pr-Venus-KmINU1*ter *KmTEF1pr-AtCHS3- KMXK_A03020ter*	Work in chapter 2

2.2. Bacterial transformation and confirmation of plasmids

E. coli DH5α was transformed with Golden Gate reaction products and selected on LB medium supplemented with appropriate antibiotics. The presence of GFP dropout cassette in pYTK001 before part insertion and in integrative dropout plasmids, and RFP dropout cassette in the level II plasmids allowed green/white and red/white colony screening. The plasmids were confirmed by PCR using OneTaq Quick-Load Polymerase (M0486L, NEB) and sequencing for level I gene plasmids or NotI (FD0596, Fisher Scientific) digestion for level II and III plasmids.

3. Yeast strain construction

3.1. Transformations and confirmation of constructed strains

A variation of the lithium acetate/PEG transformation method was used for introduction of plasmids into *K. marxianus* (Rajkumar and Morrissey, 2022). All gene expressions were established by genomic integrations. For integrative plasmids, two micrograms of plasmid DNA were digested with SgsI (FD1894, Fisher Scientific), used to transform yeasts, and selected on appropriate selective medium. Successful integrations were confirmed by PCR screening of transformants using two sets of primers: one amplifying the product between a region upstream of an integration site and the promoter of the integrated fragment (or the first promoter of integrated fragments harbouring multiple concatenated TUs) while the second set of primers amplified the product between a region downstream the integration site and the selective marker gene sequence.

3.2. Gene disruption

Gene disruptions were achieved by clean deletion of a large segment of the gene according to previously described method (Rajkumar and Morrissey, 2022). Briefly, DSB was generated at the gene locus using CRISPR plasmid described elsewhere and were repaired by homology directed repair with repair DNAs that have flanks that are homologous to the 5' and 3' ends of the gene. For single deletions of *ARO10, HIS3 LEU2* or *TRP1,* pJA_U *KmARO10*, pJA_U *KmHIS3* pJA_U *KmLEU2* and pJA_U *KmTRP1* was used, respectively. KmJA.025 which has *URA3, HIS3* and *LEU2* deleted was created with plasmid pJA_M3 *UHL* for simultaneous targeting and disruption of the three genes. These plasmids are described in Table S2, and CRISPR targets are listed in Table S4. Repair DNAs were generated by synthesizing forward and reverse overlapping primers. The primers were annealed and extended in 50 µL PCR reactions using Q5 High-Fidelity Polymerase and purified and concentrated by ethanol precipitation. Yeasts were then co-transformed with 4.5µg of repair DNAs and 150ng of

plasmids for single gene targeting or 500ng for triple gene targeting and selected on YPD agar supplemented with 400mg/L hygromycin. Gene deletions were confirmed by colony PCR using diagnostic primers that are complementary to the internal regions of genes but outside the deleted regions. Plasmids were removed from confirmed deletion strains by passaging on YPD before establishing further modifications.

3.3. Construction of strains with increased production of aromatic amino acids

To engineer increased aromatic amino acids biosynthesis in *K. marxianus* the genealogy of strains constructed is illustrated in figure 2B. First, non-homologous end joining was inactivated by deleting *YKU80* in a wildtype strain to generate KmJA.112. We have previously showed elsewhere that the disruption of this gene improves genome engineering by homologous recombination in *K. marxianus*. A histidine auxotrophic version of KmJA.112 was constructed by deleting *HIS3* to generate KmJA.019 which served as the parent strain that was engineered for increased production of these aromatic amino acids. Flux through the shikimate pathway was increased by overexpressing feedback resistant *ARO4* (*ARO4*^{FBR}) in strain KmJA.285. Precursor supply was then engineered integrating *TAL1, TKL1* and *ENO1* in KmJA.285 to generate strains KmJA.293 and KmJA.294. This strain was also used to test the overexpression of mutant *TRP2* and *GLN1* genes for their ability to increase the direction of chorismate towards tryptophan biosynthesis.

3.4. Construction of strains for naringenin production

The genealogy of strains constructed for naringenin production is illustrated in figure 3. To test the potential of a *K. marxianus* strain that we engineered elsewhere for increased phenylalanine production as platform for producing naringenin, the entire naringenin biosynthetic pathway was expressed in strain KmJA.047. This strain overexpresses *ARO4*^{FBR}, *ARO1, ARO2* and *ARO7*^{FBR} of the shikimate pathway and*TKL1, TAL1, RPE1* and *RKI1* of the pentose phosphate pathway to increase the supply of erythrose 4-phosphate precursor to the shikimate pathway. While KmASR.047 also overexpresses the aromatic aminotransferase encoding gene *ARO9* and prephenate dehydratase encoding gene *PHA2,* leading to high production phenylalanine-derived Ehrlich metabolites that can limit the phenylalanine available as precursors for producing phenylalanine-derived non-Ehrlich metabolite molecules, the strain can serve as starting point for establishing the naringenin pathway. Thus, the genes of the entire pathway were integrated into the genome of KmASR.047. To do this, an integrative cassette bearing *PlPAL, AtC4H, AtCPR1, At4-CL3, AtCHI* (naringenin cassette) was first used to transform a *lue2* version of KmASR.047 (KmJA.026) to generate KmJA.159 which served as the parent strain for testing the effect of expressing up to 3 copies of *AtCHS3*

and *KmCAB1.* To test whether coumaric acid is converted to downstream metabolites by *At-4CL3,* the coumaric acid module (*PlPAL, AtC4H* and *AtCPR1*) was integrated into KmASR.047 genome with and without *At-4CL3* to generate strains KmJA.173 and KmJA.174, respectively. KmJA.033 which has *PlPAL* integrated at the *ARO10* locus was also constructed from KmJA.026. KmJA.313 which bears 5 copies of *At4-CL3* and an *ScFEN2* copy was then generated from KmJA.033.

For gradual engineering of the pathway, the uracil, histidine and leucin auxotrophic strain, KmJA.025 served as parent strain. KmJA.253 which overexpresses the shikimate pathway and the entire naringenin pathway and 3 copies of *AtCHS3* was generated from KmJA.025. To test the effect of attenuating phenylpyruvate activity, several attempts to delete *ARO10* in KmJA.253 were unsuccessful. This was therefore achieved by first integrating the naringenin cassette into KmJA.025 to generate KmJA.160 and subsequent deletion of *ARO10,* generating KmJA.162. The cassette for overexpressing shikimate pathway genes was then integrated into KmJA.162 genome to generate KmJA.212. For testing the effect of the overexpression of malonyl-CoA biosynthetic genes, cassettes harbouring 3 copies of *AtCHS3* and single copies of *ACC1* and/or *Seacs** were used to transform KmJA.212 to generate KmJA.220, KmJA.221, KmJA.222 from which KmJA.234, KmJA.235 and KmJA.236 overexpressing *KmCAB1* were derived, respectively. KmJA.222 was further used as parent strain for testing the effect on naringenin production of increased *At4-CL3* copies, PanK engineering, overexpression of CoA biosynthetic genes, and multi-loci integration of *AtCHS3*. For multi-copy integration of *AtCHS3* at delta sites of KmJA.282 genome, pJA_M4 delta and SgsI digestion product of pJA_NM L3.54 were used to transform the strain and selected on hygromycin supplemented YPD plate. Six transformants were obtained of which one displayed fluorescence. For fluorescence detection in transformants, yellow fluorescent protein (YFP) signals were determined by cultivation in YPD overnight, followed by reinoculation in fresh YPD by diluting down to 100 fold and further cultivating for 3 hours. Cultures were then harvested at mid-exponential growth phase by a single wash with water, resuspended in water and transferred at volumes of 200µL into 96-well plates. Using a Spark Multimode Microplate reader (Tecan), we determined culture A 600nm and fluorescence with excitation and emission of 485 and 535nm, respectively. After correcting for the autofluorescence of wildtype *K. marxianus*, the fluorescence signal was normalised to cell number by dividing by the A 600nm.

4. Cultivation and sampling of production strains

For determining the titres of aromatic molecules produced by strains, they were preculture overnight in 10mL of minimal glucose medium. Cultures were started at A 600nm 0.1 in 50 mL of fresh medium in 250 mL Erlenmeyer flasks and incubated at 30^{o}C in a shaker at 200 rpm

for 5 days. Extracellular metabolites concentrations were determined in culture supernatants. For aromatics, supernatant was collected by mixing 700µL of cultures with 96% ethanol at 1:1 ratio, vortexed and pelleted by centrifugation using a benchtop centrifuge at 4°C for 5 minutes at maximum speed. Glucose, ethanol, acetate, and glycerol concentrations were determined in undiluted culture. Supernatants were stored at -20°C until analysis.

5. Analytical methods

The concentrations of extracellular aromatics were determined by an Agilent 1260 Infinity HPLC Infinity system with Zorbax Eclipse Plus C18 column (4.6 x 150mm, 3.5µm; Agilent, Little Island, Ireland) operated at 40°C with 20mM KH_2PO_4 brought to pH2 with phosphoric acid and supplemented with 1% acetonitrile/acetonitrile as the mobile phase at a flow rate of 0.8 mL/min as previously described(Koopman *et al.*, 2012). For elution, acetonitrile was increased from 0-10% for the first 6 minutes, then to 40% between 6 to 23 minutes, and then switched to 100% KH_2PO_4 for 23 to 27 minutes. Aromatic metabolites were detected by a VWD100 multi-wavelength detector with the absorbance of Ehrlich pathway metabolites, naringenin and phloretic acid measured at 200nm and at 280nm for cinnamic and coumaric acids. Data were analysed using OpenLab CDS (Agilent). All HPLC standards for metabolites were purchased from Sigma-Aldrich or Fisher Scientific. The concentrations of glucose were determined with an Agilent 1200 HPLC system (Agilent Technologies, CA, USA) with a REZEX 8µL 8% H+ organic acid column (300 ×7.8 mm) (Phenomenex, CA, USA) at 65°C using 0.01 N H_2SO_4 as the mobile phase at a flow rate of 0.6 mL/min and were detected by a refractive index detector.

Results and Discussion

1. Metabolic engineering of *K. marxianus* to increase aromatic amino acid biosynthesis

The biosynthesis of AAAs is illustrated in figure 2A and wildtype *K. marxianus* was engineered for increased production of these amino acids by expressing genes in the pathway according to the strain genealogy provided in figure 2B. Firstly, the flux through the shikimate pathway was increased by overexpressing a mutant *ARO4* which encodes feedback resistant DAHP synthase (Dahps) in wildtype *K. marxianus* to generate strain KmJA.285. This resulted in the production of tyrosol, 2-phenylethanol and tryptophol compared to the parent strain (KmJA.019) for which the three molecules were not detected (figure 2C). This result shows that the biosynthesis of tyrosine, phenylalanine and tryptophol was increased in this stain. To test whether production can be further increased by overexpressing pentose phosphate

pathway and glycolytic genes in order to increase the supply of erythrose 4-phosphate and phosphoenolpyruvate which are the precursors of the shikimate pathway, the native genes involved in both pathways were overexpressed in KmJA.285 to generate KmJA.293 overexpressing *TKL1* and *TAL1* and KmJA.294 overexpressing *TAL1, TKL1* and *ENO1.* Both strains produced lower cumulative titre of the three aromatic alcohols, showing that precursor engineering did not increase AAA production in this strain. Due to low production of tryptophan in these strains and the fact that the production of tyrosol and 2-phenylethanol but not tryptophan in *K. marxianus* has been demonstrated elsewhere (Rajkumar and Morrissey, 2020), we attempted to improve tryptophan production by two strategies. The first strategy which involved the expression of mutant *TRP2* to encode for feedback resistant isoforms of anthranilate synthase, S64R and S74L mutants in strain KmJA.297 and KmJA.298, respectively, increased the production of tryptophol by over 7-folds, but KmJA.299 expressing an isoform with both mutations did not show increased tryptophol production. For the second strategy we looked at the possibility that increased production of glutamine could increase tryptophan production because it acts as amino donor for the conversion of chorismate into anthranilate (figure 2A) by overexpressing the native *GLN1,* but this strategy surprisingly reduced tryptophol production (figure 2C).

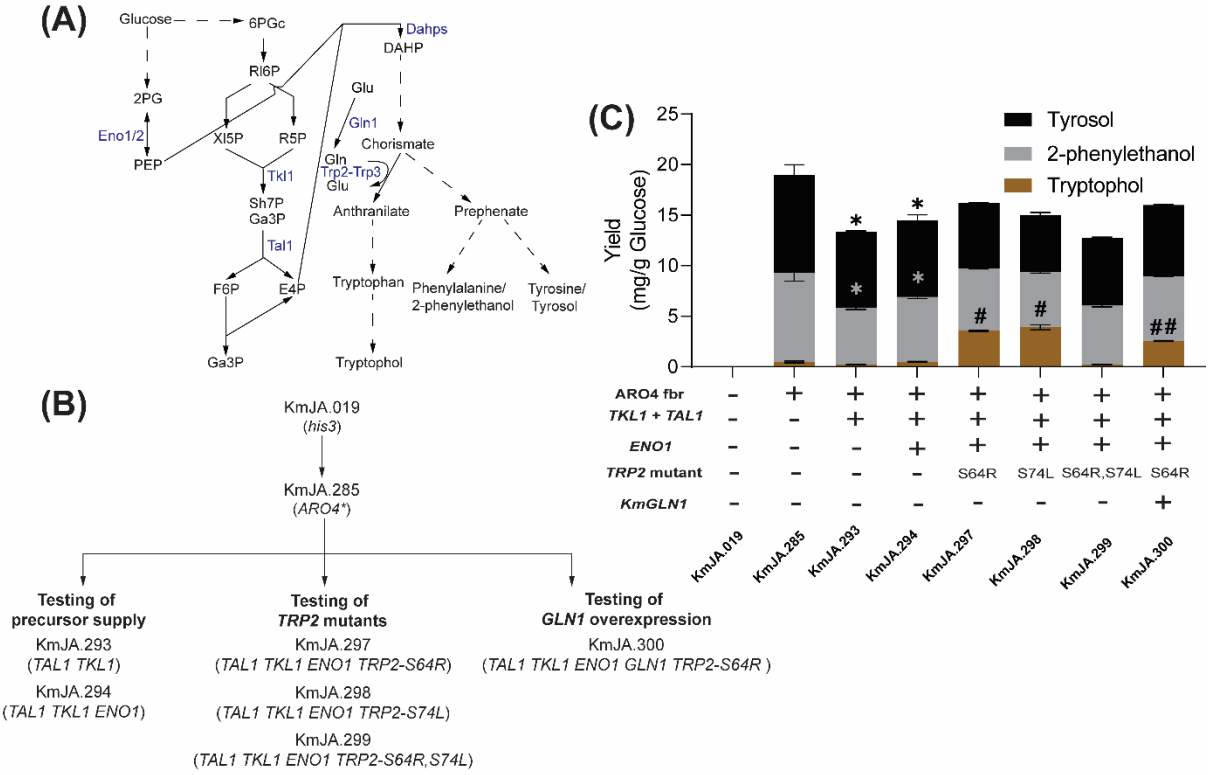


Figure 2. Metabolic engineering of *K. marxianus* for increased production of aromatic amino acids. (A) Phosphoenolpyruvate (PEP) from the glycolytic pathway and erythrose 4-phosphate (E4P) from the pentose phosphate pathway are precursors for the shikimate pathway. Genes encoding the native enzymes shown in blue texts were overexpressed in *K. marxianus* to engineer it to produce aromatic

amino acids at higher levels than in the wildtype yeast. Strains were constructed according to the genealogy illustrated in (B). The yields of tyrosol, 2-phenylethanol and tryptophan (aromatic alcohol derivatives of tyrosine, phenylalanine, and tryptophan, respectively) during the cultivation of the strains in minimal glucose medium are reported (C). Data shown are the mean ± standard deviation of three replicates. Statistically significant differences (examined by independent ttest) in extracellular metabolite yields by strains relative to control strains, KmJA.285, KmJA.294 and KmJA.297 are denoted by an asterisk (*, $p<0.05$), a hash (#, $p<0.05$) and two hash symbols (##, $p<0.05$), respectively. The symbols are directly placed above the bar representing the metabolites that they refer to.

2. Expression of the entire naringenin biosynthetic pathway in *K. marxianus* produced coumaric acid and low yield of naringenin.

We previously constructed *K. marxianus* strain KmASR.047 elsewhere. This strain produces 2-phenylethanol at a yield of 24.2mg/g glucose. To engineer *K. marxianus* for naringenin production from phenylalanine, the entire naringenin pathway consisting of *PlPAL* and genes from *A. thaliana* were integrated in single copies into the genome of KmASR.047 for expression under the control of strong constitutive promoters. The derivative strain KmJA.168 produced coumaric at 6.3mg/g (figure 4A). Naringenin was not detected in this strain but phloretic acid was detected at a peak area of just above 3000 (figure 4B). An increase in the number of *AtCHS3* integrated to two and three in strains KmJA.169 and KmJA.170, respectively, resulted in the production of detectable naringenin in both strains. This was accompanied by decrease in the production of coumaric acid and phloretic acid in these strains compared to KmJA.168. The three strains did not show difference in 2-phenylethanol production compared to the parent strain, KmASR.047 (figure 4A).

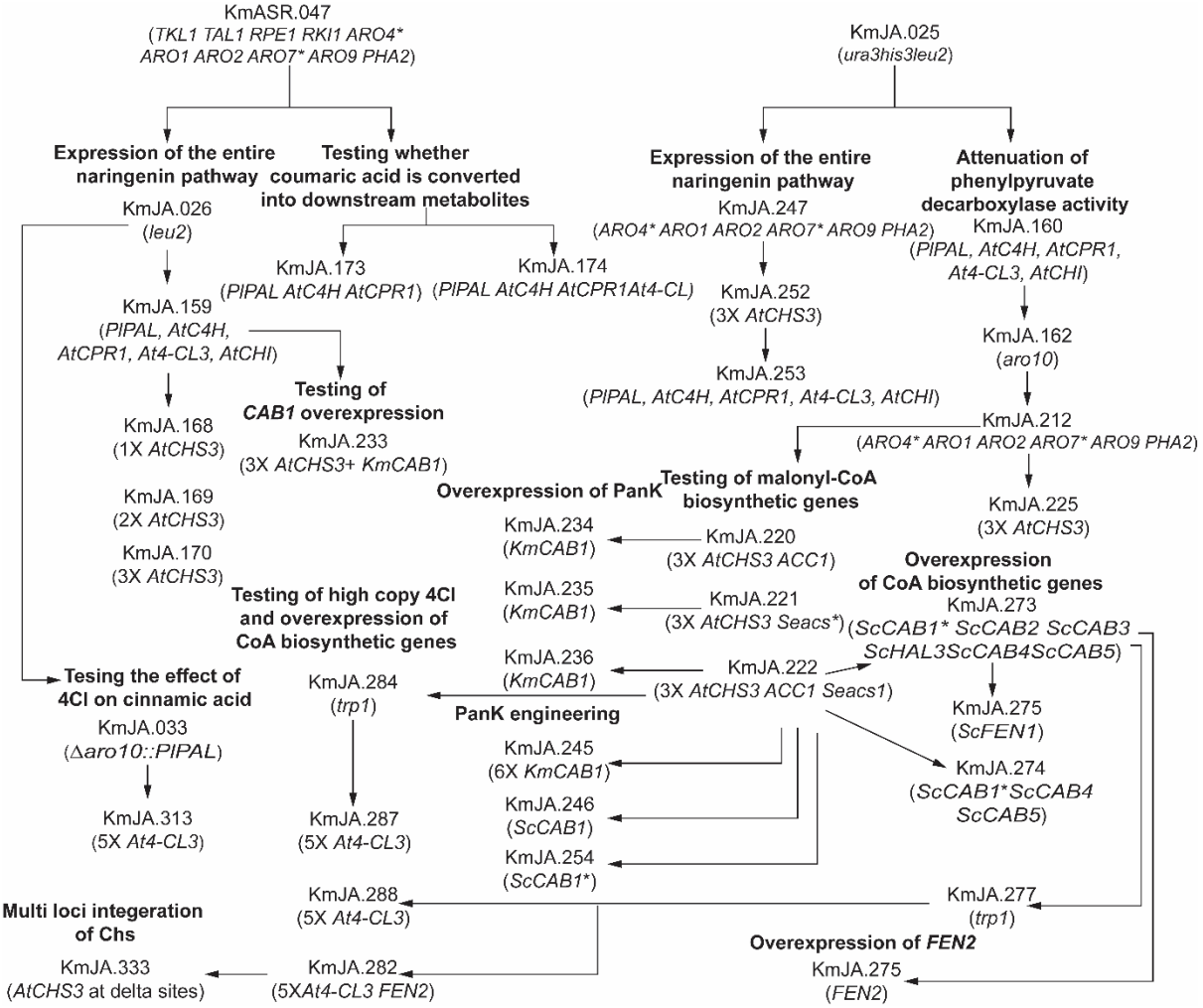


Figure 3. Genealogy of *K. marxianus* strains constructed for naringenin production.

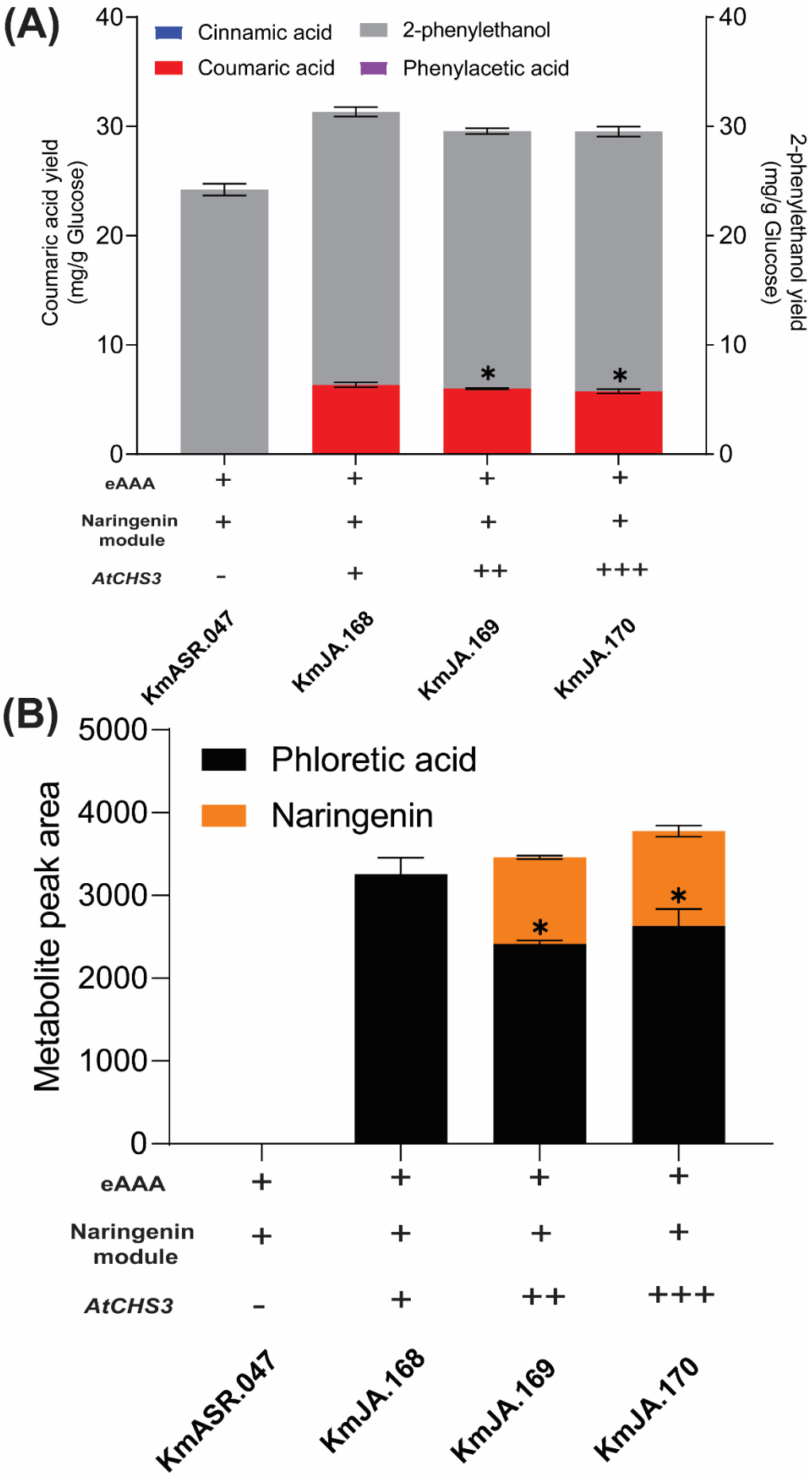

(A)
Cinnamic acid
2-phenylethanol
Coumaric acid
Phenylacetic acid
Coumaric acid yield (mg/g Glucose)
2-phenylethanol yield (mg/g Glucose)
40
30
20
10
0
*
*
eAAA
Naringenin module
AtCHS3
+
+
+
+
+
+
+
+
-
+
++
+++
KmASR.047
KmJA.168
KmJA.169
KmJA.170
(B)
Phloretic acid
Naringenin
Metabolite peak area
5000
4000
3000
2000
1000
0
*
*
eAAA
Naringenin module
AtCHS3
+
+
+
+
+
+
+
+
-
+
++
+++
KmASR.047
KmJA.168
KmJA.169
KmJA.170

Figure 4. Expression of naringenin biosynthetic genes. All genes of the naringenin pathway were integrated into the genome of KmASR.047 to generate the other three strains. KmASR.047 is a *K. marxianus* strain that already has its aromatic amino acid biosynthetic pathway enhanced towards phenylalanine production (eAAA). KmJA.168, KmJA.169 and KmJA.170 have one, two or three copies of *AtCHS3* integrated in their genomes. Metabolite yields were determined for the four strains during growth in minimal glucose medium (A). Peak areas are reported for phloretic acid and naringenin because they were produced at concentrations below the limits of accurate measurements (B). Data shown are the mean ± standard deviation of three replicates. Statistically significant difference (examined by independent ttest) in coumaric acid yield by strains relative to control strain, KmJA.168, are denoted by an asterisk (*, $p<0.05$). The symbol is directly placed above the bar representing the metabolites it refers to.

3. Gradual engineering of *K. marxianus* for naringenin production

To enable the establishment and and exploration of other engineering targets for naringenin production, a moderately engineering *K. marxianus* strain, KmJA.247, which overexpresses only the *ARO4, ARO1, ARO2* and *ARO7* genes of the shikimate pathway, in addition to *ARO9* and *PHA2* from a single locus in its genome was used as the parent strain for gradual engineering of naringenin production. Importantly, this strain is different to the previous parent strain used above (KmASR.047) in that the genes of the non-oxidative phase of the PPP which are integrated at a separate locus are not present in KmJA.247. Thus, KmJA.247 produces lower yield of 2-phenylethanol (14.2mg/g versus 24.2mg/g) and frees up the integration site for additional gene expression that can contribute towards naringenin production. Similar to KmASR.047, the integration of the entire naringenin pathway in KmJA.247 to generate KmJA.253 caused phenylalanine to enter the pathway, resulting in low production of cinnamic acid but not coumaric acid, naringenin nor phloretic acid, and 2-phenylethanol production was similar in both strains. To increase the availability of phenylalanine to the pathway, KmJA.225 was constructed. This strain has the modifications present in KmJA.247 in addition to the disruption of *ARO10* which contributes most of the phenylpyruvate decarboxylase activity in *K. marxianus*. It produced cinnamic and coumaric acid at a cumulative yield of 6.7mg/g and 2-phenylethanol production persisted at 14.2mg/g (figure 2A). Naringenin was not detected but phloretic acid was produced at peak area of 1100 in this strain.

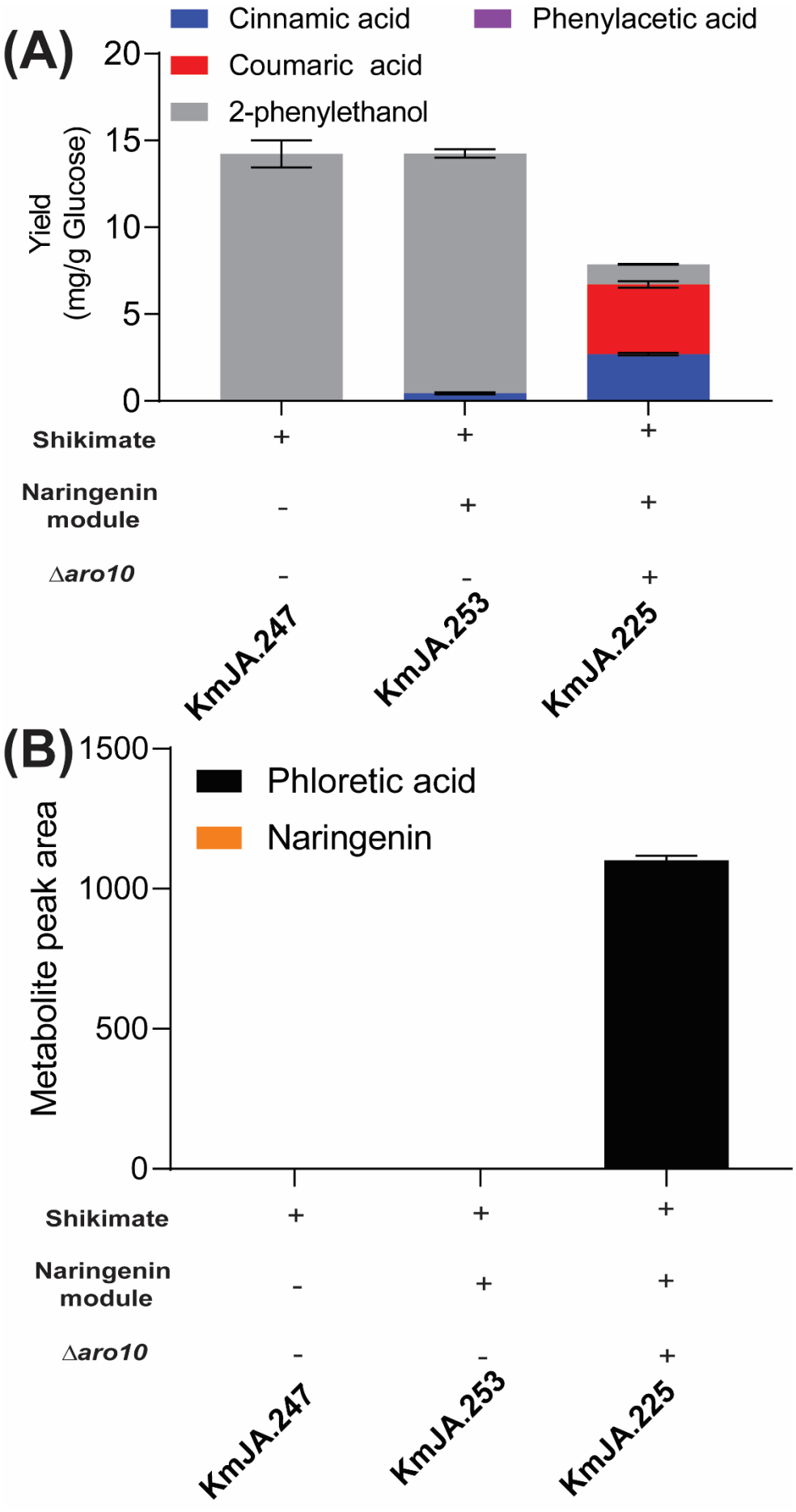

(A)
Cinnamic acid
Phenylacetic acid
Coumaric acid
2-phenylethanol
Yield (mg/g Glucose)
20
15
10
5
0
Shikimate
Naringenin module
Δaro10
KmJA.247
KmJA.253
KmJA.225
(B)
Phloretic acid
Naringenin
Metabolite peak area
1500
1000
500
0
Shikimate
Naringenin module
Δaro10
KmJA.247
KmJA.253
KmJA.225

Figure 5. Disruption of *ARO10* increased the cumulative yield of cinnamic and coumaric acid. The entire naringenin pathway was integrated into the genome of KmJA.247 which overexpresses the shikimate pathway to generate KmJA.253. Strain KmJA.225 is similar to KmJA.253 in its modifications but has its *ARO10* disrupted. The yields of 2-phenylethanol, cinnamic acid and coumaric acid during the cultivation of the three strains in minimal glucose medium are reported in (A). Peak areas are reported for phloretic acid and naringenin to reflect their production at concentrations below the limits of accurate measurements (B). Data shown are the mean ± standard deviation of three replicates.

3.1. Overexpression of malonyl-CoA biosynthetic genes

To increase the biosynthesis of malonyl-CoA for Chs catalysed conversion of coumaroyl-CoA into naringenin chalcone, L641P mutant of *S. enterica acs* and the native *ACC1* were overexpressed in strains of KmJA.220 (*ACC1*), KmJA.221 (mutant *Seacs*), KmJA.222 (both genes). These strains also contain the modifications in KmJA.225 but only differ in that they bear the malonyl-CoA biosynthetic genes. The three strains showed higher cumulative yield of cinnamic and coumaric acid with no 2-phenylethnol detected compared to KmJA.225 (figure 6A). However, the new modifications did not result in the detection of naringenin and phloretic acid which was detected in KmJA.225 (figure 6B).

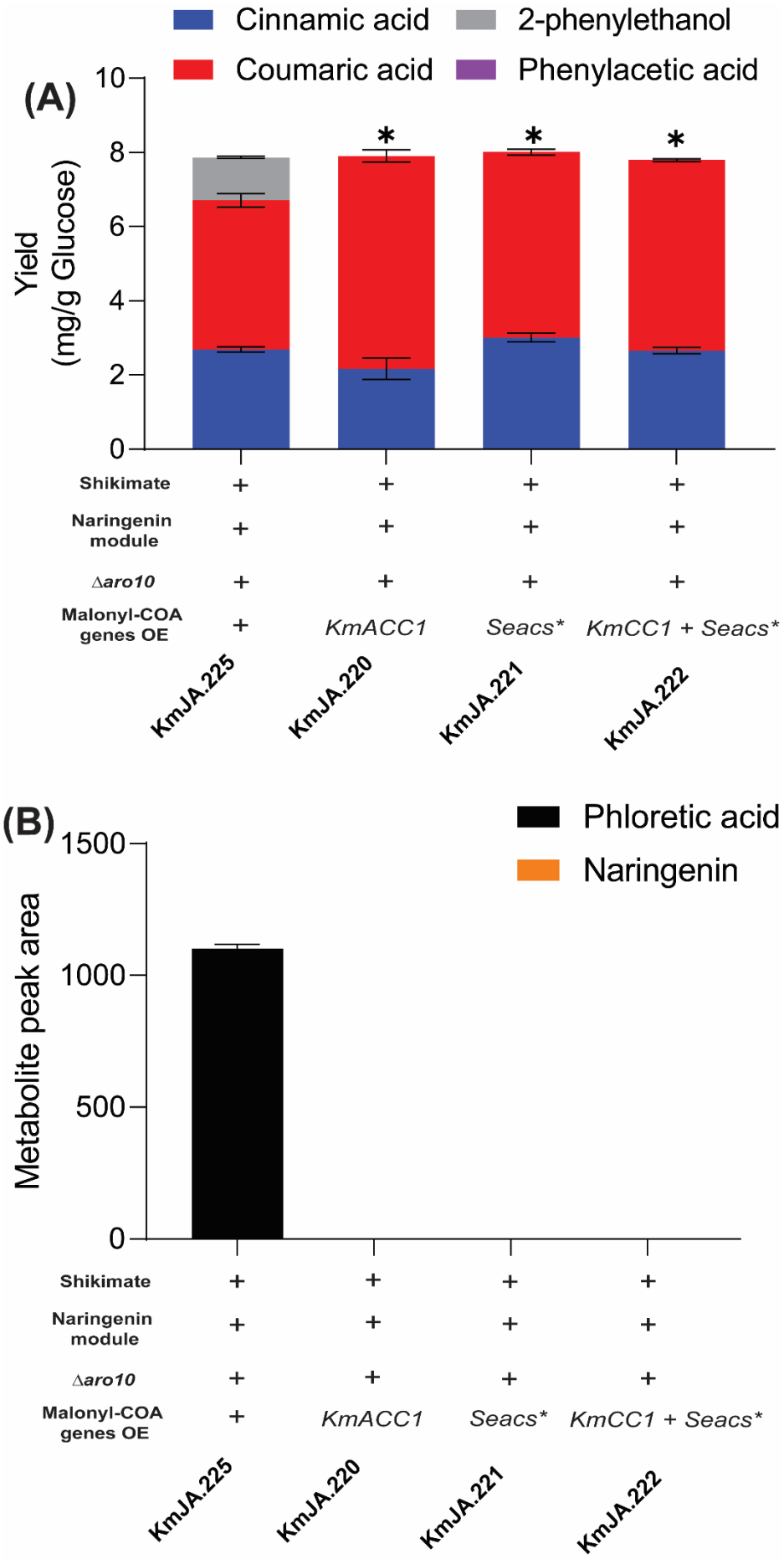

(A)
Cinnamic acid
2-phenylethanol
Coumaric acid
Phenylacetic acid
Yield (mg/g Glucose)
10
8
6
4
2
0
*
*
*
Shikimate
Naringenin module
Δaro10
Malonyl-COA genes OE
+
+
+
+
+
+
+
+
+
+
+
+
+
KmACC1
Seacs*
KmCC1 + Seacs*
KmJA.225
KmJA.220
KmJA.221
KmJA.222
(B)
Phloretic acid
Naringenin
Metabolite peak area
1500
1000
500
0
Shikimate
Naringenin module
Δaro10
Malonyl-COA genes OE
+
+
+
+
+
+
+
+
+
+
+
+
+
KmACC1
Seacs*
KmCC1 + Seacs*
KmJA.225
KmJA.220
KmJA.221
KmJA.222

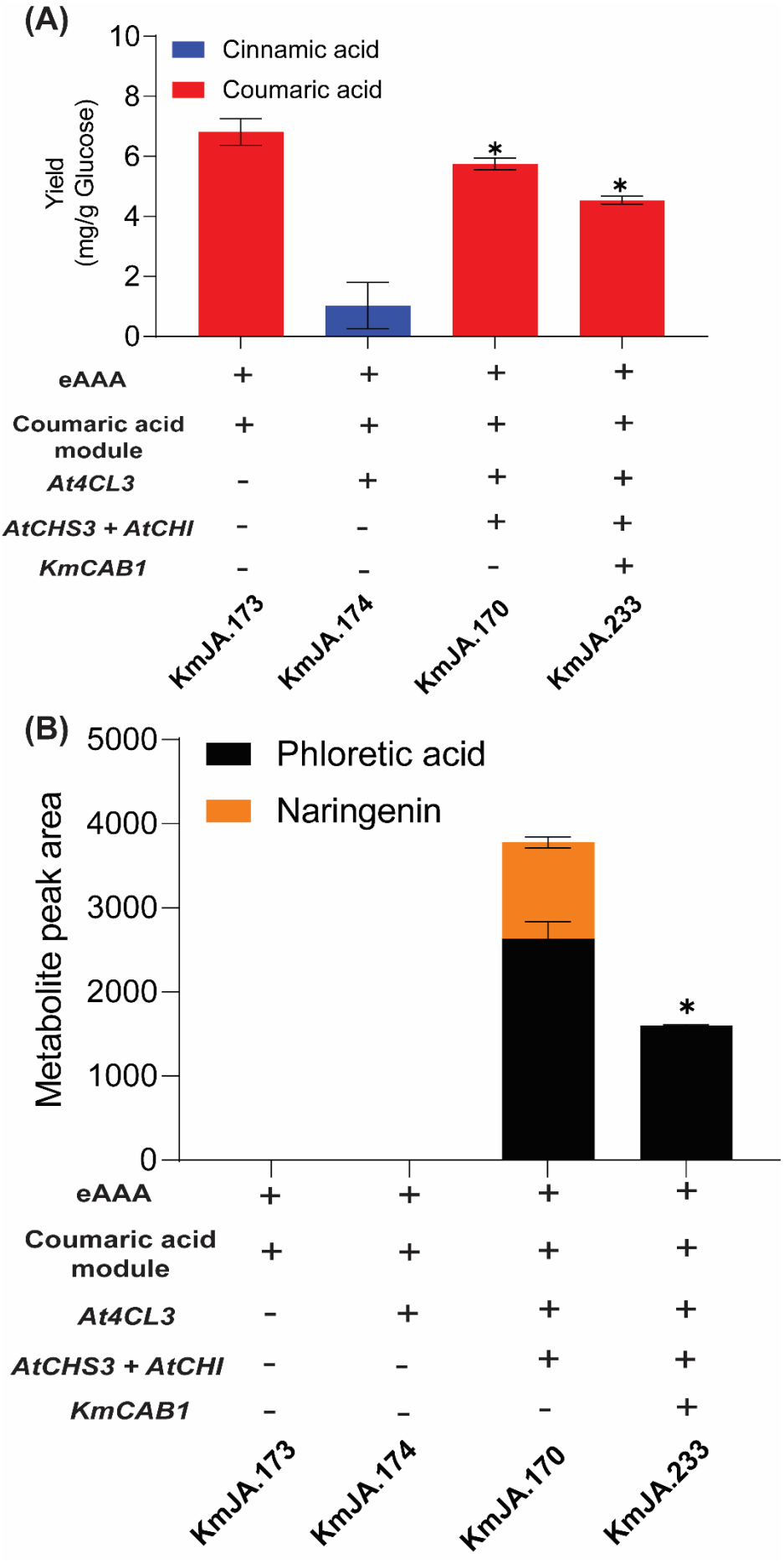
Figure 6. Overexpression of malonyl-CoA biosynthetic genes. Genes involved in malonyl-CoA biosynthesis (native *ACC1* and a mutant version of *S. enterica acs* (*Seacs**)) were integrated into the genome of KmJA.220, KmJA.221 and KmJA.222. These strains are similar to KmJA.225 except that it has no integrated malonyl-CoA biosynthetic genes. Metabolite yields were determined for the four strains during growth in minimal glucose medium (A). Peak areas are reported for phloretic acid and naringenin to reflect their production at concentrations below the limits of accurate measurements (B). Data shown are the mean ± standard deviation of three replicates. Statistically significant difference (examined by independent ttest) in coumaric acid yield by strains relative to control strain, KmJA.225 are denoted by an asterisk (*, $p<0.05$).

3.2. Coenzyme A is a limiting metabolite in naringenin biosynthesis

To determine whether coumaric acid can be converted into coumaroyl-CoA by At4-Cl3, the coumaric acid module (*PlPAL, AtC4H and AtCPR1*) was integrated into the genome of KmASR.047 to generate strain KmJA.173 which produced 6.8mg/g of coumaric acid and *At4-CL3* was further integrated into this strain to generate KmJA.174. Coumaric acid could not be detected in KmJA.174 but cinnamic acid was produced, showing that At4-Cl3 was indeed able to consume coumaric acid, presumably into coumaroyl-CoA (figure 7A). Phloretic acid which should be produced from the coumaroyl-CoA pool was, however, not detected in the strain (figure 7B). Coumaric acid, naringenin and phloretic acid were detected in KmJA.170 which is similar to KmJA174 in regard to all modifications but additionally has integrated 3 copies of *AtCHS3* and a copy of *AtCHI* (figure 7A and B). These results suggest that coenzyme A may be a limiting metabolite in naringenin production because it appears that the expression of At4-Cl3 which catalysis requires only one coenzyme A molecule efficiently activated coumaric acid, but the additional expression of *AtCHS3* which catalysis requires three coenzyme A molecules in the form of 3 molecules of malonyl-CoA, and *AtCHI* increased the need for coenzyme A, thereby compromising both reactions and resulting in an inability to efficiently deplete coumaric acid.

As illustrated in figure 7C, the biosynthesis of coenzyme A begins with the transportation of pantothenate into the cell by the plasma membrane H+-pantothenate symporter (Fen2) in *S. cerevisiae*. Intracellular pantothenate is then phosphorylated by pantothenate kinase (Cab1) to produce 4'-phosphopantothenate which can together with cysteine be converted into 4'-phosphopantothenoylcystein by 4′-phosphopantothenoylcysteine synthetase (Cab2). 4'-phosphopantothenoylcystein is then decarboxylated into 4'phosphopantetheine by a heterotrimeric enzyme called 4′-phosphopantothenoylcysteine decarboxylase for which the three constituent proteins are encoded by three separate genes, *CAB3, HAL3 and VHS3*. 4'phosphopantetheine is then adenylated by 4'phosphopantetheine adenylyltransferase (Cab4) to form dephospho-CoA which is finally phosphorylated into coenzyme A by the action of the last enzyme in the pathway, dephospho-CoA kinase (Cab5). Although the pathway is

energetically costly due to its requirement for ATP by four of the five reactions, the expression of the native genes were recently reported for successful engineering of increased coenzyme A biosynthesis in *S. cerevisiae* (Olzhausen *et al.*, 2021). In line with our hypothesis here that coenzyme could be a limiting metabolite in the production of naringenin, the overexpression of *CAB1* in KmJA.233 further reduced the yield of coumaric acid compared to KmJA.170 (figure 7A). Naringenin production was however not detected in this strain, but we observed a lower peak area for phloretic acid compared to KmJA.170 (figure 7B). Limited coenzyme A availability can be the reason why the overexpression of malonyl-CoA biosynthetic genes did not improve naringenin production above in figure 6. To test this, we integrated the native *CAB1* into the genomes KmJA.220, KmJA.221, KmJA.222 to generate KmJA.334, KmJA.335 and KmJA.336, respectively, and naringenin was detected in the three derivative strains (figure 7D).


(A)
Yield (mg/g Glucose)
Cinnamic acid
Coumaric acid
eAAA
Coumaric acid module
At4CL3
AtCHS3 + AtCHI
KmCAB1
KmJA.173
KmJA.174
KmJA.170
KmJA.233
(B)
Metabolite peak area
Phloretic acid
Naringenin
eAAA
Coumaric acid module
At4CL3
AtCHS3 + AtCHI
KmCAB1
KmJA.173
KmJA.174
KmJA.170
KmJA.233

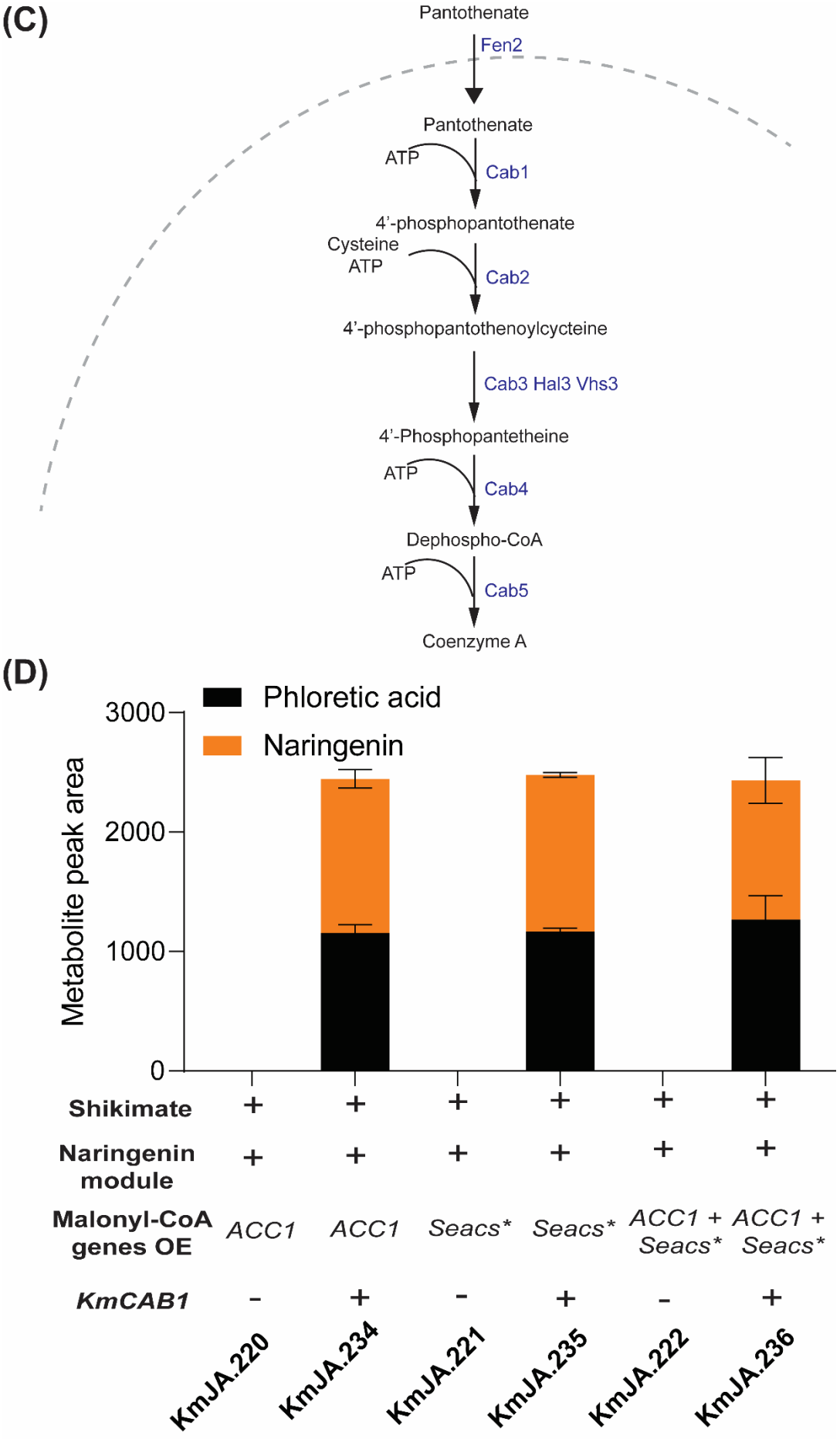

(C)
Pantothenate
Fen2
Pantothenate
ATP
Cab1
4'-phosphopantothenate
Cysteine
ATP
Cab2
4'-phosphopantothenoylcycteine
Cab3 Hal3 Vhs3
4'-Phosphopantetheine
ATP
Cab4
Dephospho-CoA
ATP
Cab5
Coenzyme A
(D)
Phloretic acid
Naringenin
Metabolite peak area
3000
2000
1000
0
Shikimate
Naringenin module
Malonyl-CoA genes OE
KmCAB1
ACC1
ACC1
Seacs*
Seacs*
ACC1 + Seacs*
ACC1 + Seacs*
KmJA.220
KmJA.234
KmJA.221
KmJA.235
KmJA.222
KmJA.236

Figure 7. Demonstration of coenzyme A as a limiting metabolite for naringenin production. (A and B) The coumaric acid module (*PlPAL, AtC4H* and *AtCPR1*) was integrated into the genome of KmASR.047 with or without *At4-CL3* to generate KmJA.173 and KmJA.174. KmASR.047 already has its aromatic amino acid biosynthetic pathway enhanced towards phenylalanine production (eAAA). *AtCHS3* and *AtCHI1* were expressed in addition to those modifications in KmJA.170 and KmJA.233 with the difference between these two strains being the overexpression of the native *CAB1.* (C) Schematic representation of the coenzyme A biosynthetic pathway. (D) The native *CAB1* was overexpressed in addition to malonyl-CoA biosynthetic genes. Metabolite yields were determined for all strains during growth in minimal glucose medium. Peak areas are reported for phloretic acid and naringenin to reflect their production at concentrations below the limits of accurate measurements. Data shown are the mean ± standard deviation of three replicates. Statistically significant difference (examined by independent ttest) in coumaric acid yield by strains relative to control strains are denoted by an asterisk (*, $p<0.05$).

3.3. Overexpression of Coenzyme A biosynthetic genes for naringenin production

To test whether the overexpression of the coenzyme A biosynthetic genes can increase naringenin production in *K. marxianus*, all the genes in the pathway were overexpressed. First, we compared KmJA.236 bearing a single copy of *KmCAB1* to KmJA.245, KmJA.246 and KmJA.254 which bear 6 copies of *KmCAB1*, single copy of *ScCAB1* and single copy of W331R mutant of *ScCAB1*, respectively. The mutant *ScCAB1* is resistant to feedback inhibition by acetyl-CoA which is a major end-product metabolite that inhibits the enzyme. Compared to KmJA.236, KmJA.245 produced lower coumaric acid, higher naringenin and no phloretic acid while KmJA.246 produced the 3 metabolites at higher yields and the KmJA.254 produced them at similar yields (figure 8 A and B). Overexpression of the remaining coenzyme A biosynthetic genes in KmJA.273 lowered coumaric acid production but neither naringenin nor phloretic acid was detected in this strain, showing that expressing the entire pathway improved coumaric acid consumption by At4-Cl but has negative effect on naringenin and phloretic acid production. The expression of only *ScCAB4* and *ScCAB5* in addition to the mutant *ScCAB1* (KmJA.274), however, increased the cumulative yield of naringenin and phloretic acid but did not reduce coumaric acid yield compared to KmJA.254.

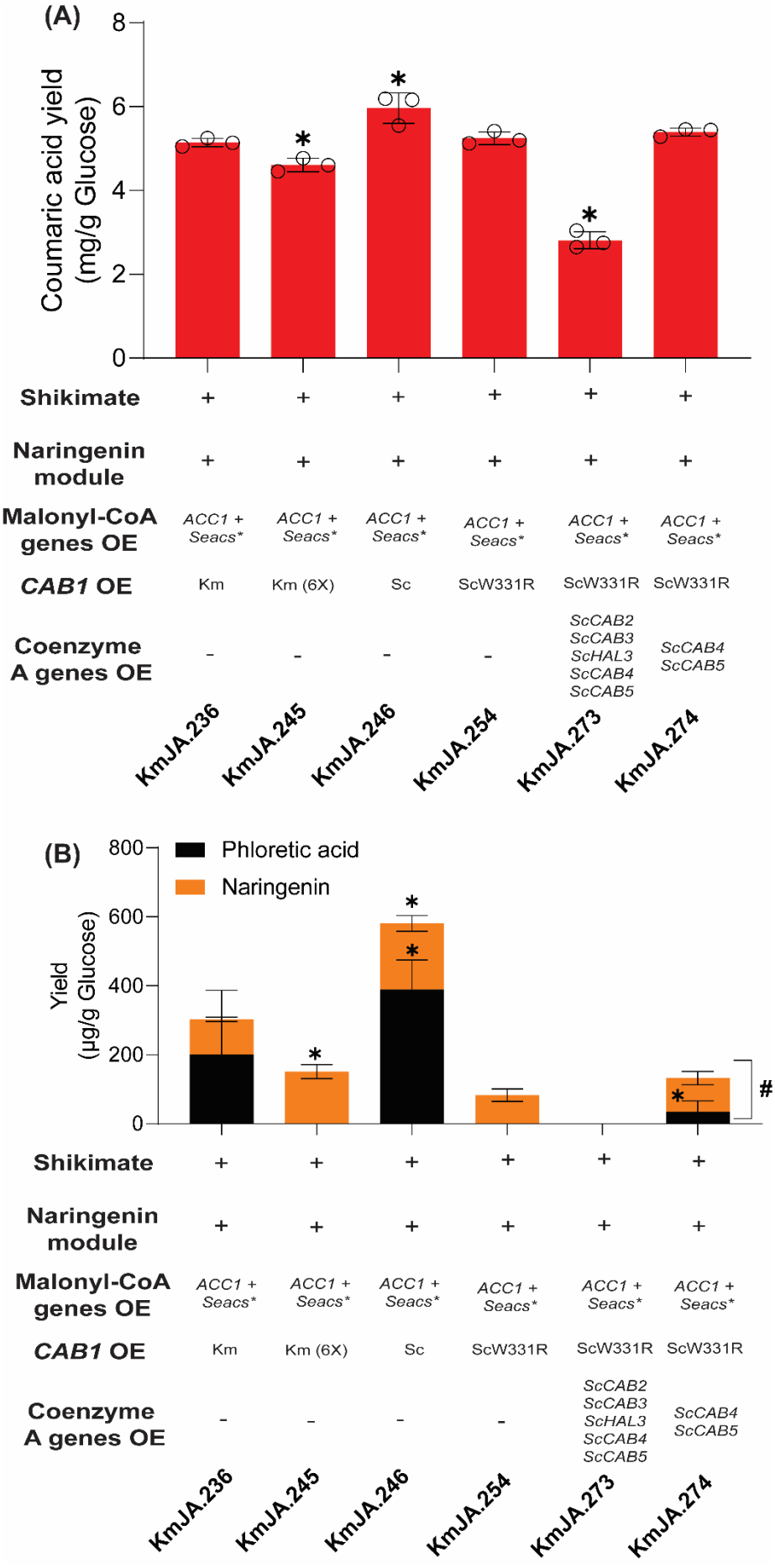

(A)
Coumaric acid yield (mg/g Glucose)
0
2
4
6
8
Shikimate + + + + + +
Naringenin module + + + + + +
Malonyl-CoA genes OE ACC1 + Seacs* ACC1 + Seacs* ACC1 + Seacs* ACC1 + Seacs* ACC1 + Seacs* ACC1 + Seacs*
CAB1 OE Km Km (6X) Sc ScW331R ScW331R ScW331R
Coenzyme A genes OE - - - - ScCAB2 ScCAB3 ScHAL3 ScCAB4 ScCAB5 ScCAB4 ScCAB5
KmJA.236 KmJA.245 KmJA.246 KmJA.254 KmJA.273 KmJA.274
(B)
Phloretic acid
Naringenin
Yield (µg/g Glucose)
0
200
400
600
800
#
Shikimate + + + + + +
Naringenin module + + + + + +
Malonyl-CoA genes OE ACC1 + Seacs* ACC1 + Seacs* ACC1 + Seacs* ACC1 + Seacs* ACC1 + Seacs* ACC1 + Seacs*
CAB1 OE Km Km (6X) Sc ScW331R ScW331R ScW331R
Coenzyme A genes OE - - - - ScCAB2 ScCAB3 ScHAL3 ScCAB4 ScCAB5 ScCAB4 ScCAB5
KmJA.236 KmJA.245 KmJA.246 KmJA.254 KmJA.273 KmJA.274

Figure 8. Engineering of coenzyme A biosynthetic pathway. The copy number of the native *CAB1* was increased from one in KmJA.236 to six in KmJA.245. Single copies of wildtype *ScCAB1* and the mutant genes were also tested in KmJA.246 and KmJA.254, respectively. In addition to the mutant *ScCAB1*, genes that encode the functional enzymes that catalyse the remaining 4 reactions, or the last two (Cab4 and Cab5) reactions of pathway were overexpressed in KmJA.273 and KmJA.274. Metabolite yields were determined for the four strains during growth in minimal glucose medium. Data shown are the mean ± standard deviation of three replicates. Statistically significant differences (examined by independent ttest) in extracellular metabolite yields by strains relative to control strains, KmJA.236 and KmJA.254 are denoted by an asterisk (*, $p<0.05$) and a hash (#, $p<0.05$), respectively. The symbols are directly placed above the bar representing the metabolites they refer to. For KmJA.274, the statistical analysis was based on the cumulative yield of naringenin and phloretic acid.

3.4. High copy number of *At4-CL3* and overexpression of coenzyme A biosynthetic genes increased naringenin production

Having demonstrated that the overexpression of coenzyme A biosynthetic genes increased the yield of naringenin and phloretic acid but that coumaric acid production remained detectable, we asked whether the 4Cl reaction can be improved by increasing the copies of *At4-CL3.* To test this, KmJA.287 was constructed by integrating 5 additional copies of the gene into the genome of KmJA.222, making a total of 6 copies. Indeed, coumaric acid was not detected in this strain (figure 9A), suggesting that the increase in *At4-CL3* copies caused coumaric acid to be more efficiently converted into coumaroyl-CoA. However, neither naringenin nor phloretic acid was detected in KmJA.287. KmJA.288 was also constructed by integrating 5 copies of *At4-CL3* into the genome of KmJA.273 which bears an integration of the entire coenzyme A pathway. Again, coumaric acid, naringenin and phloretic acid were not detected in KmJA.288. To explore the possibility that increased intracellular availability of the pantothenate precursor of coenzyme A could increase coenzyme A biosynthesis, we overexpressed *ScFEN2* in order to increase the transportation of the precursor into the cell in KmJA.273 which already has the coenzyme A biosynthetic pathway integrated to generate KmJA.275. KmJA.282 was also constructed from KmJA.273 by integrating 5 additional copies of *At4-CL3* and *ScFEN2*. While the production of the three metabolites was still not detected in KmJA.275, KmJA.282 on the other hand produced all three metabolites with lower coumaric acid production than KmJA.222. KmJA.282 produced 570mg/g of naringenin and 466mg/g of phloretic acid. To rule out the possibility that At-4Cl3 is taking cinnamic acid as substrate for conversion into cinnamoyl-CoA which could be at play especially in those strains that lost their ability to produce coumaric acid without producing detectable naringenin and phloretic acid (KmJA.287, KmJA.288 and KmJA.275), the effect of *At4-CL3* on cinnamic acid production was tested by integrating 5 copies of the genes alongside a single *ScFEN2* copy into the genome of KmJA.033 to generate KmJA.313. KmJA.033 has only *PlPAL* integrated and produces cinnamic acid at 1.2mg/g. However, the additional integration of these genes in KmJA.313 did not reduce cinnamic acid production, showing that At4-Cl3 did not deplete cinnamic acid.

Thus, it is unlikely that the high copy of *At4-CL3* depleted cinnamic acid in KmJA.287, KmJA.288 and KmJA.275.

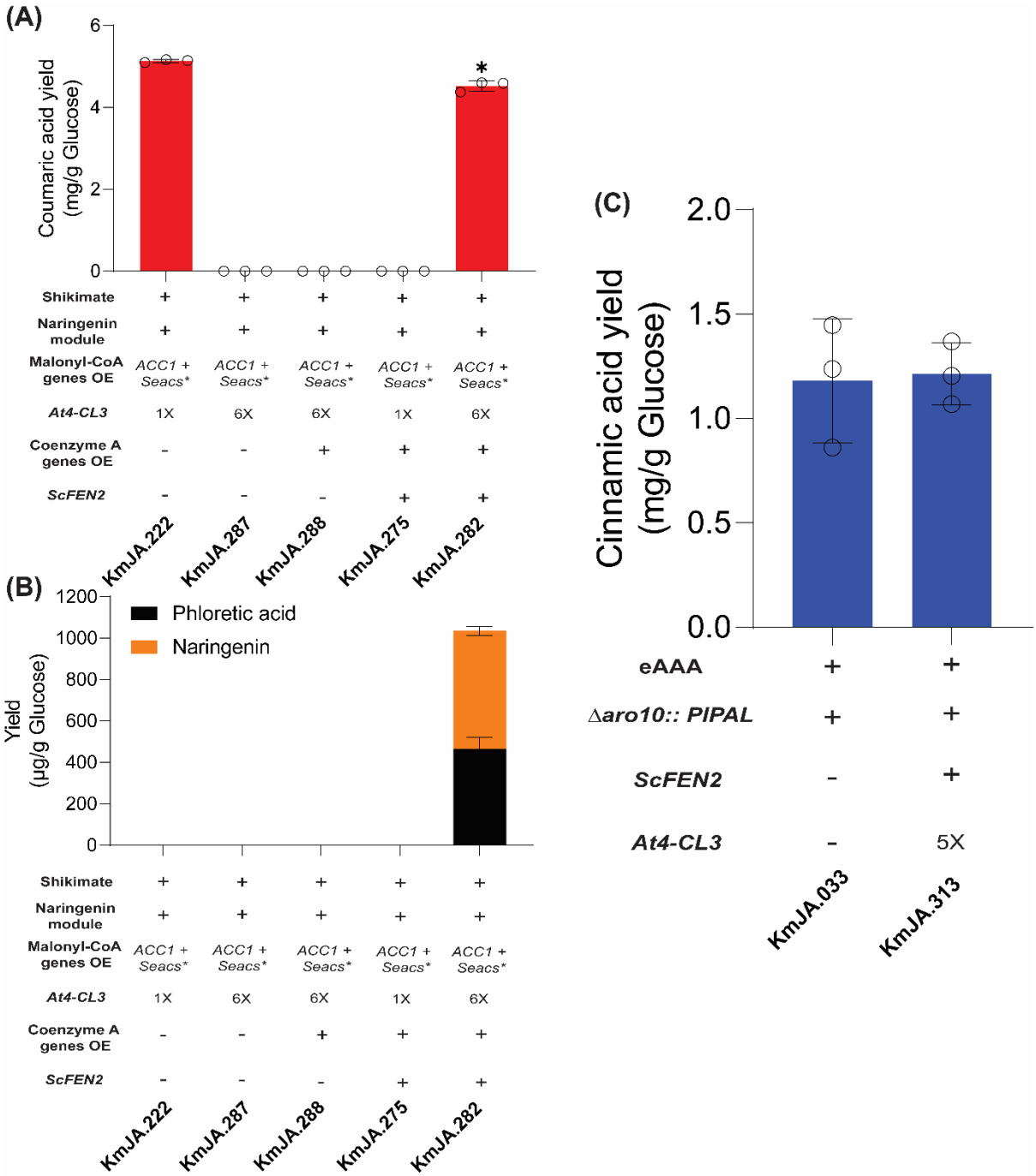


Figure 9. High copy of *At4-CL3* and expression of coenzyme A biosynthetic genes increased naringenin and phloretic acid production. Five additional copies of *At-4CL3* were integrated into the genome of KmJA.222 to make a total of 6 copies of the gene in KmJA.287. KmJA.288 also has 6 copies of *At4-CL3* but in addition to the integration of the coenzyme A pathway. *ScFEN2* was also integrated into KmJA.275 and KmJA.282 which have one and six copies of *At-4CL3*, respectively. The yields of coumaric acid (A) and naringenin and phloretic acid (B) were determined for these strains. (C) strain KmJA.033 was constructed from KmASR.047 which has enhanced aromatic amino acid biosynthetic pathway (eAAA) by the integration of only a PAL gene from *Photorhabdus luminescens* (*PlPAL*) at the *ARO10* locus, causing the gene to be disrupted. Further integration of a cassette with 5 copies of

At4CL3 and *ScFEN2* generated KmJA.313. Metabolite yields were determined for all strains during growth in minimal glucose medium. Data shown are the mean ± standard deviation of three replicates. Statistically significant difference (examined by independent ttest) in metabolite yield by strains relative to control strain, KmJA.222 are denoted by an asterisk (*, $p<0.05$).

3.5. Multi-loci integration of *AtCHS3* increased naringenin production

To further increase the copy number *AtCHS3* in KmJA.282 which produces the highest yield of naringenin so far, this strain was transformed with a CRISPR plasmid targeting the delta site sequences that are distributed around yeast's genome (pJA_M4 delta) and an integrative cassette bearing 2 copies of *AtCHS3* and a *Venus* (YFP) transcriptional unit for fluorescence based screening. The transformation generated six transformants of which one was fluorescence-positive, was called KmJA.233 and produced double the yields of naringenin and phloretic acid at 1231µg/g and 1143µg/g, respectively, compared to KmJA.282 (figure 10). To increase pantothenate precursor supply available during cultivation, media supplementation of pantothenate was increase by 50X from 1mg/L to 50mg/L. However, naringenin and phloretic acid yields were reduced for both strains in higher pantothenate concentrations.

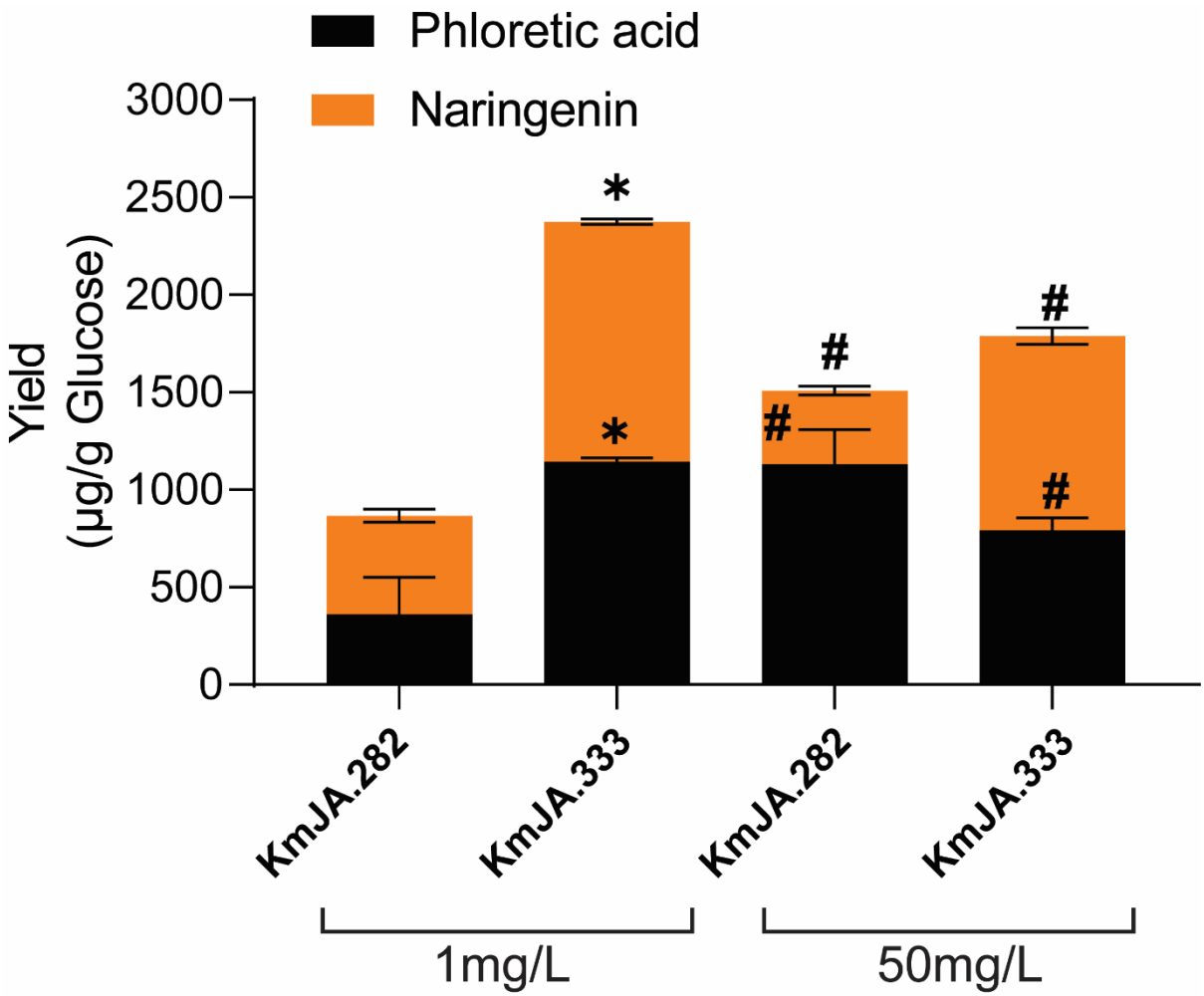


Figure 10. Multi-loci integration of *AtCHS3* at delta sites. KmJA.333 was constructed through CRISPR-mediated integration of a cassette with two copies of *AtCHS3* and a transcriptional unit for encoding a yellow fluorescent protein (*Venus*) for screening successfully integrated transformants. The yields of naringenin and phloretic acid were determined for both strains during growth in minimal glucose medium supplemented with 1mg/L or 50mg/L of pantothenate. Data shown are the mean ± standard deviation of three replicates. Statistically significant difference (examined by independent ttest) in metabolite yields by KmJA.333 relative to control strain, KmJA.282 are denoted by an asterisk (*, $p<0.05$) while difference in yields in 50mg/L pantothenate compared to in 1mg/L is denoted by a hash (#, $p<0.05$). The symbols are directly placed above the bars representing the metabolites they refer to.

Supplementary materials

Table S1. Non-yeast native genes expressed in this study.

Gene name	Source organism	GenBank accession number
At4-CL3	*Arabidopsis thaliana*	AY376730
AtCHS3	*Arabidopsis thaliana*	BT000596
AtCHI	*Arabidopsis thaliana*	BT005528
AtC4H	*Arabidopsis thaliana*	AY065145
PlPAL	*Photorhabdus luminescens*	BX57186
AtCPR1	*Arabidopsis thaliana*	NM_118585

Table S2. Plasmids used in this study.

Name	Description	Source
Level I plasmids		
JA_NM L1.9	Level I plasmid *Photorhabdus luminescens PAL* codon optimised for *S. cerevisiae*	Work in chapter 3
JA_NM L1.10	Level I plasmid of *Arabidopsis thaliana CPR1* codon optimised for *S. cerevisiae*	Work in chapter 3
JA_NM L1.6	Level I plasmid of *Arabidopsis thaliana C4H* codon optimised for *S. cerevisiae*	This work
JA_NM L1.31	Level I plasmid of *K. marxianus GLN1*	Work in chapter 3
JA_NM L1.3	Level I plasmid of *Arabidopsis thaliana 4-CL3* codon optimised for *S. cerevisiae*	This work
JA_NM L1.4	Level I plasmid of *Arabidopsis thaliana CHS3* codon optimised for *S. cerevisiae*	This work
JA_NM L1.5	Level I plasmid of *Arabidopsis thaliana CHI* codon optimised for *S. cerevisiae*	This work
JA_NM L1.35	Level I plasmid of *K. marxianus CAB1*	This work
JA_NM L1.36	Level I plasmid of *S. cerevisiae CAB1*	This work

JA_NM L1.37	Level I plasmid of *S. cerevisiae CAB2*	This work
JA_NM L1.38	Level I plasmid of *S. cerevisiae CAB3*	This work
JA_NM L1.39	Level I plasmid of *S. cerevisiae HAL3*	This work
JA_NM L1.40	Level I plasmid of *S. cerevisiae CAB4*	This work
JA_NM L1.41	Level I plasmid of *S. cerevisiae CAB5*	This work
JA_NM L1.42	Level I plasmid of *K. marxianus TRP2(S64R)*	This work
JA_NM L1.43	Level I plasmid of *K. marxianus TRP2(S74L)*	This work
JA_NM L1.44	Level I plasmid of *K. marxianus TRP2(S64R; S74L)*	This work
pA10	Level I plasmid of *K. marxianus ENO1*	(Rajkumar and Morrissey, 2020)
pA9	Level I plasmid of *K. marxianus TKL1*	(Rajkumar and Morrissey, 2020)
pA11	Level I plasmid of *K. marxianus TAL1*	(Rajkumar and Morrissey, 2020)
pA1f	Level I plasmid of *K. marxianus ARO4(K221L)*	(Rajkumar and Morrissey, 2020)

Level-II Dropout plasmids		
JA_LII RFP-DO 1	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 1 of level III (multigene) plasmids	Work in chapter 3
JA_LII RFP-DO 2	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 2 of level III (multigene) plasmids	Work in chapter 3
JA_LII RFP-DO 3	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 3 of level III (multigene) plasmids	Work in chapter 3
JA_LII RFP-DO 4	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 4 of level III (multigene) plasmids	Work in chapter 3
JA_LII RFP-DO 5	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for position 5 of level III (multigene) plasmids	Work in chapter 3
JA_LII RFP-DO 6	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 6 of level III (multigene) plasmids	Work in chapter 3
JA_LII RFP-DO 7	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 2 of level III (multigene) plasmids	Work in chapter 3
JA_LII RFP-DO 8	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 3 of level III (multigene) plasmids	Work in chapter 3

JA_LII RFP-DO 9	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 4 of level III (multigene) plasmids	Work in chapter 3
JA_LII RFP-DO 10	Level II plasmid with RFP dropout for constructing level II (gene transcriptional unit) plasmid intended for the terminal transcriptional unit at position 5 of level III (multigene) plasmids	Work in chapter 3

Level-II plasmids		
pJA_VM L2.24	pJA_LII RFP-DO 1 with insert *KmTEF1pr-KmTKL1-KmINU1t*	This work
P1A11T5	*KmPDC1pr-KmTKL1-KMXK_A03020t* for position II TU in level III plasmids	(Rajkumar and Morrissey, 2020)
pJA_VM L2.23	pJA_LII RFP-DO 3 with insert *KmSSA2pr-KmTKL1-ScADH1t*	This work
pJA_VM L2.20	pJA_LII RFP-DO 6 with insert *KmPDC1pr-KmTRP2(S64R)-ScPGK1*	This work
pJA_VM L2.21	pJA_LII RFP-DO 6 with insert *KmPDC1pr-KmTRP2(S74R) -ScPGK1*	This work
pJA_VM L2.22	pJA_LII RFP-DO 6 with insert *KmPDC1pr-KmTRP2-(S64R;S74R)-ScPGK1*	This work
pJA_NM L2.85	JA_LII RFP-DO 5 with insert *KmPDC1pr-KmGLN1-ScPGK1t*	Work in chapter 3
pJA_NM L2.15	pJA_LII RFP-DO 1 with insert *KmTDH3pr-At4-CL3-KmPDC1t*	This work
pJA_NM L2.16	pJA_LII RFP-DO 2 with insert *KmTEF1pr-AtCHS3-KMXK_A03020t*	This work
pJA_NM L2.17	pJA_LII RFP-DO 3 with insert *KmPDC1pr-AtCHI1-KmINU1t*	This work
pJA_NM L2.6	pJA_LII RFP-DO 4 with insert *KmENO1pr-AtC4H-KmPGK1t*	This work
pJA_NM L2.12	pJA_LII RFP-DO 10 with insert *KmENO1pr-AtC4H-KmPGK1t*	This work
pJA_NM L2.26	pJA_LII RFP-DO 1 with insert *KmTEF1pr-AtCHS3-KMXK_A03020t*	This work
pJA_NM L2.27	pJA_LII RFP-DO 7 with insert *KmTEF1pr-AtCHS3-KMXK_A03020t*	This work
pJA_NM L2.33	pJA_LII RFP-DO 8 with insert *KmTEF1pr-AtCHS3-KMXK_A03020t*	This work
pJA_NM L2.69	pJA_LII RFP-DO 3 with insert *KmTEF1pr-AtCHS3-KMXK_A03020t*	This work
pJA_NM L2.90	pJA_LII RFP-DO 9 with insert *KmPDC1pr-KmCAB1-ScPGK1*	This work
JA_NM L2. 18	pJA_LII RFP-DO 1 with insert *KmPDC1pr-PlPAL-ScPGK1t*	Work in chapter 3
JA_NM L2. 20	pJA_LII RFP-DO 2 with insert *KmENO1pr-AtC4H-KmPGK1t*	Work in chapter 3
JA_NM L2. 21	pJA_LII RFP-DO 8 with insert *KmENO1pr-AtCPR1-KmPGK1t*	Work in chapter 3
pJA_NM L2.70	pJA_LII RFP-DO 9 with insert *KmPDC1pr-KmACC1-ScPGK1t*	This work
pJA_NM L2.71	pJA_LII RFP-DO 9 with insert *KmPDC1pr-Seacs(L641P)-ScPGK1t*	This work
pJA_NM L2.72	pJA_LII RFP-DO 4 with insert *KmPDC1pr-KmACC1-ScPGK1t*	This work
pJA_NM L2.71	pJA_LII RFP-DO 10 with insert *KmPDC1pr-Seacs*(L641P)*-ScPGK1t*	This work
pJA_NM L2.98	pJA_LII RFP-DO 1 with insert *KmPDC1pr-ScCAB1(W331R)-ScPGK1t*	This work
pJA_NM L2.110	pJA_LII RFP-DO 2 with insert *KmPDC1pr-ScCAB2-ScPGK1t*	This work
pJA_NM L2.111	pJA_LII RFP-DO 3 with insert *KmPDC1pr-ScCAB3-ScPGK1t*	This work
pJA_NM L2.112	pJA_LII RFP-DO 4 with insert *KmPDC1pr-ScHAL3-ScPGK1t*	This work
pJA_NM L2.113	pJA_LII RFP-DO 5 with insert *KmPDC1pr-ScCAB4-ScPGK1t*	This work
pJA_NM L2.114	pJA_LII RFP-DO 6 with insert *KmPDC1pr-ScCAB5-ScPGK1t*	This work
pJA_NM L2.48	pJA_LII RFP-DO 2 with insert *KmTDH3pr-At4-CL3-KmPDC1t*	This work

pJA_NM L2.49	pJA_LII RFP-DO 3 with insert *KmTDH3pr-At4-CL3-KmPDC1t*	This work
pJA_NM L2.50	pJA_LII RFP-DO 4 with insert *KmTDH3pr-At4-CL3-KmPDC1t*	This work
pJA_NM L2.51	pJA_LII RFP-DO 5 with insert *KmTDH3pr-At4-CL3-KmPDC1t*	This work
pJA_NM L2.52	pJA_LII RFP-DO 6 with insert *KmTDH3pr-At4-CL3-KmPDC1t*	This work
pJA_NM L2.91	pJA_LII RFP-DO 1 with insert *KmPDC1-KmCAB1-ScPGK1*	This work
pJA_NM L2.92	pJA_LII RFP-DO 2 with insert *KmPDC1-KmCAB1-ScPGK1*	This work
pJA_NM L2.93	pJA_LII RFP-DO 3 with insert *KmPDC1-KmCAB1-ScPGK1*	This work
pJA_NM L2.94	pJA_LII RFP-DO 4 with insert *KmPDC1-KmCAB1-ScPGK1*	This work
pJA_NM L2.95	pJA_LII RFP-DO 5 with insert *KmPDC1-KmCAB1-ScPGK1*	This work
pJA_NM L2.96	pJA_LII RFP-DO 6 with insert *KmPDC1-KmCAB1-ScPGK1*	This work
pJA_NM L2.118	pJA_LII RFP-DO1 with insert *KmPDC1pr-FEN1-ScPGK1t*	This work

MTU-DO plasmids		
pI3 MTU-DO-G418	Integrative plasmid with GFP dropout targeting locus between *KmSWF1* and *KmARO1*; KanMX marker	Addgene ID. 160023
pI1 MTU-DO-HIS	Integrative plasmid with GFP dropout targeting *KmLAC4* locus; *HIS3* marker	Addgene ID. 160142
pI4 MTU-DO-G418	Integrative plasmid with GFP dropout targeting locus downstream of *KmHSP104*; KanMX marker	Addgene ID. 160215
pI3 MTU-DO-LEU	Integrative plasmid with GFP dropout targeting locus between *KmSWF1* and *KmARO1*; ScLEU2 marker	Addgene ID. 160176
pI3 MTU-DO-Hph	Integrative plasmid with GFP dropout targeting locus between *KmSWF1* and *KmARO1*; *HphMX* marker	Addgene ID. 160176
pI3 MTU-DO-TRP1	Integrative plasmid with GFP dropout targeting locus between *KmSWF1* and *KmARO1*; ScTRP1 marker	This work

CRIPSR plasmids used for gene deletions		
pJA_U *KmARO10*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmARO10}$-HDV-ScCYC1t HphMX- KanR*	Work in chapter 3 and 4
pJA_U *KmHIS3*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmHIS3}$-HDV-ScCYC1t HphMX- KanR*	Work in chapter 2
pJA_U *KmLEU2*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmLEU2}$-HDV-ScCYC1t HphMX- KanR*	Work in chapter 2
pJA_U *KmTRP1*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmAro9}$-HDV-ScCYC1t HphMX- KanR*	This work
pJA_ M3 *UHL*	*KmTEF1pr-SpCas9-KmINU1t ScTDH3pr-HH-gRNA$_{KmURA3}$-HDV-ScCYC1t ScTDH3pr-HH-gRNA$_{KmHIS3}$-HDV-ScCYC1t ScTDH3pr-HH-gRNA$_{KmLEU2}$-HDV-ScCYC1t HphMX- KanR*	Work in chapter 2
pJA_M4 delta	*KmTEF1pr-SpCas9-KmINU1t* 4 X (*ScTDH3pr-HH-gRNA$_{delta}$ -HDV-ScCYC1t*) *hphMX- KanR*	This work

Table S3. Primers used in this study.

Name	Sequence	Purpose
ASR_K001F	TTTGCTGGCCTTTTGCTC	General colony PCR forward primer, priming at the 3' end of ColE1 in all of plasmid backbones; used for confirmation of plasmid assembly and sequencing of level I plasmid inserts
ASR_K002R	CATCTGGATTTGTTCAGAACG	Colony PCR reverse primer, priming downstream of the BsmBI cloning site in YTK001; used for confirmation of plasmid assembly and sequencing of level I plasmid inserts

ASR_K001R	ATTGGTAACTGTCAGACCAAGTTTA	Colony PCR reverse primer, priming at the 5' end of AmpR; used for confirmation of level II plasmid assembly
ASR_K007R	CAGTCATCGGTATGATCTG	Colony PCR reverse primer, priming in ConRE; used for confirmation of I7/ARO10 MTU-DO plasmid assembly
ASR_I1US_F	CATTAGAACCTTTTTCAACACTC	Forward primer for checking integration at I1/*LAC4* integration site; primes 92bp upstream of the site.
ASR_I1DS_R	CTTAGTGGTTGTGAAGGTTT	Reverse primer for checking integration at I1/*LAC4* integration site; primes 90bp downstream of the site.
ASR_I3_US_F	ATTCCAGACACAGGTGGAA	Forward primer for checking integration at I3 integration site; primes 113bp upstream of the site.
ASR_I3_DS_R	AAAGGAAGGATTCGGCGTTA	Reverse primer for checking integration at I3 integration site; primes 114bp downstream of the site.
ASR_I5_US_F	AGTAGTGAGTGACAGACAC	Forward primer for checking integration at I4 integration site; primes 62bp upstream of the site.
ASR_I5_DS_R	GCAGTTTCTGTGCAGAAAA	Reverse primer for checking integration at I4 integration site; primes 102bp downstream of the site.
ASR_I6_US_F	AGCCTGTCTACGTACAAC	Forward primer for checking integration at I6/*ARO3* integration site; primes 107bp upstream of the site.
ASR_I6_DS_R	CGTTCCAAGGAAACCACTATA	Reverse primer for checking integration at I6/*ARO3* integration site; primes 98bp upstream of the site.
JA063	AAACAGCAGAGTTGTTAAGATTAGTTGAGGTTTTGGGTCCATATATCTGTCTATTGAAGACACA TGTAGATATCTTGGAGGATTTCAGCTTTGAGTTTGGGACAACAATA	Forward primer for *KmURA3* repair DNA
JA064	GATCTCTTCCCTTTGCGAAAAGACCTCTACCAACAATAATGATGTCTGATCCACCGGCAACAA CTTCATCCACAGTTCTGTATTGTTGTCCCAAACTCAAAGCTGAAATC	Reverse primer for *KmURA3* repair DNA
JA059	ATGACATACCCAGAAAGGAAAGCTTTTGTGTCTAGAATAACAAATGAGACAAAAATTCAGATA GCCATATCCTTACATGGGGGGCATATCTCAATGATTTGGGTTTAAAA	Forward primer for *KmHIS3* repair DNA
JA060	AAAGTTACTCTTGCAGCTTCCGCAAATGACTCCAAGAAATGTGGTATCATTTCACATGAAAGAT CACCGATTTTTTCTCTTTTTAAACCCAAATCATTGAGATATGCCCC	Reverse primer for *KmHIS3* repair DNA
JA061	ATCAAAGTTTTGAACGCTATTTCAGAAGCACGCCCTTCTATCAAGTTTAATTTTGAACATCATTT GATTGGTGGTGCTGCAATTGATGCTACTGGTGGCATCATCTAGAT	Forward primer for *KmLEU2* repair DNA
JA062	AGCGGAATCGATCAATTGATGTTGGACAGTCAATTGAGGGAACTCATTCTTGATGGTTTCTTC TACTGTTTTTCTCCACAATCTAGATGATGCCACCAGTAGCATCAATT	Reverse primer for *KmLEU2* repair DNA
JA080	AGCAGCTGCTCATAGAAGT	Forward diagnostic primers for detecting *KmURA3* deletion
JA081	GGATCTGCCTACTCTCTTCAA	Reverse diagnostic primers for detecting *KmURA3* deletion
JA065	CCAGAAAGGAAAGCTTTTGTGTCTAGAATAAC	Forward diagnostic primers for detecting *KmHIS3* deletion
JA319	GCTGACTCACTTCTGTGATGATCATTAAAGC	Reverse diagnostic primers for detecting *KmHIS3* deletion
JA320	TCAAAGAATATCGTTGTCTTACCAG	Forward diagnostic primers for detecting *KmLEU2* deletion

JA321	TGCCAAAATCTCCTTTACAGC	Reverse diagnostic primers for detecting *KmLEU2* deletion
JA311	CCCAGGTGTATAATAGTTTAATATCAATCGAGTAATATACGGTTTTCTCAGGAATAACAACTGC CTTAGTCTAGTCCACAGACATGCTCGTCAAGGGGCCAGCAGCTCGA	Forward diagnostic primers for detecting *KmTRP1* deletion
JA312	TATCTGGGGTGAGTCCTCCCGCGAGTATGAAGCGGGTGGTATTACCTTGGTTCTGACACCAG CTGGCCATGCCGTTCCAGTCGAGCTGCTGGCCCCTTGACGAGCATGTC	Reverse diagnostic primers for detecting *KmTRP1* deletion
JA313	CACCATCCAAAAGAAGGACATCA	Forward diagnostic primers for detecting *KmTRP1* deletion
JA314	TCCATTGACAAGTGGTTCTTCAT	Reverse diagnostic primers for detecting *KmTRP1* deletion
JA198	CTTGAATAGTCATTTGAGCAGCACCATCACCTTCACACAAGATCAACTTTGGAACATAGTCTG CTGGGACAGCAGATTGGTTGATAATGTGAGATTTCTAGCAAAGGCAA	Forwards primer of overlapping primer set for producing *KmARO10* repair DNA
JA199	TCACTCTTGGTCGTTATGTGTTCGAGAGATTGCTCAACTGTGGTTCCAAGACCATTTTCGGTG TTCCAGGTGACTTCAACTTGCCTTTGCTAGAAATCTCACATTATCAA	Reverse primer of overlapping primer set for producing *KmARO10* repair DNA
JA204	AGTGTAACCATTGTTGTTGAA	Forward diagnostic primer to amplify and confirm deletion at *KmARO10* locus
JA205	CCAGTAGTTCTAGACGATAAATC	Reverse diagnostic primer to amplify and confirm deletion at *KmARO10* locus

Table S4. CRIPSR targets for gene deletions.

Gene	**Target sequence**
KmURA3	AGGAAGTTACGAAAGAACCT
KmHIS3	AGCGAAGCACTCTGGTTGGT
KmLEU2	TTGAAACCTGAGTACGCCAA
K. marxianus Delta sites	AAAGCATTTGATAATTGAAT
KmTRP1	AATCTCACAACTGGTTCACC
KmARO10	GTCCAAAGAAGACTTAGCCA

Chapter 6. General discussion and recommendations for future research

Joel A. Akinola

1. Summary of results in this thesis

The objective of this thesis was to develop tools that can contribute to rapid construction of *K. marxianus* strains for synthetic biology (SynBio) applications and use them to build strains that produce plant aromatic compounds. These objectives were chosen because (1) while *S. cerevisiae* is commonly used for yeast synthetic biology, *K. marxianus* has specific physiological properties that makes it industrially attractive such as the fact that it is thermotolerant and can metabolise and grow on many substrates, thus allowing the use of cheap feedstock for production. (2) having a tool for rapid modification of this yeast will drive its use for SynBio applications. This will encourage its development into products that can be commercialised. (3) aromatic compounds are profitable due to their highly valuable applications in many industries including food, pharmaceutical, nutraceutical and agrotech industries. To achieve these objectives, rapid CRISPR-based *K. marxianus* genome editing tools were developed and applied in combination with SynBio tools developed by others for engineering the yeast to produce important members of a large group of plant aromatic molecules called phenylpropanoids.

First, genome engineering was addressed in terms of the limited number of selectable genes available for *K. marxianus* by developing the *Aspergillus niger* acetamidase gene (*amdS*) as a selection marker to increase the number of markers available for this yeast, and by CRISPR-mediated marker free integration to do away with selection marker requirements for genome editing in the yeast (Chapter 2). *amdS* was chosen because it has been demonstrated to be functional for this purpose in another Kluyveromyces yeast, *K. lactis,* and it is counter-selectable in *S. cerevisiae*. Unlike these two yeasts, *K. marxianus* can metabolise and grow on acetamide, and homology search revealed that four putative acetamidase genes are present in *K. marxianus*. Phenotypic characterisation of strains that have these genes disrupted showed that only two of them are required for growth on acetamide. *amdS* was successfully used as a marker for efficient genomic integration of cassettes in a *K. marxianus* strain that was unable grow on acetamide due to loss of acetamidase activity. However, *amdS* marker recycling was not successful in our work here due to evidence that fluoroacetamide which is used to counter-select for *amdS* in *S. cerevsiae* was unable to counter-select for the gene in *K. marxianus.* Marker-free genomic integration systems provide an alternative to the search for more selectable markers in this thesis. To achieve this, a CRISPR platform that allowed the integration of cassettes without the need for markers was developed. The platform was demonstrated to be efficient for integrating up to three expression cassettes simultaneously and for multi loci cassette integration at specific gene loci and at delta site sequences distributed across the yeast's genome.

In chapter 3, available SynBio tools developed by others were used to overexpress the shikimate pathway in the yeast to increase the biosynthesis of aromatic amino acids which serve as biological precursors of aromatic compounds. A strain with enhanced aromatic amino acid production was then used as platform for additional engineering to produce coumaric acid at a maximum yield of close to 50mg/g glucose via phenylalanine by expressing *PlPAL, AtC4H* and *AtCPR1* under the control of strong constitutive promoters. One of the strategies that was necessary for attaining coumaric acid production at that yield was the attenuation of phenylpyruvate decarboxylase activity to prevent the degradation of the phenylalanine precursors. This was achieved by disrupting the genes (*ARO10* and *PDC5*) suspected contribute this activity in *K. marxianus* based on homology to the protein that contributes the activity in *S. cerevisiae*, Aro10. However, detailed enzymatic and functional characterization in chapter 4 later revealed that *PDC5* does not have phenylpyruvate activity but that it regulates the transcription of thiamine biosynthetic genes. Furthermore, bioinformatic and synteny analysis identified it as the *THI3* of *K. marxianus* (Chapter 4). Thus, it was concluded that that this gene should not be disrupted as a strategy for attenuating phenylpyruvate decarboxylase activity. *ARO10* and *PDC1* are the only 2-oxo acid decarboxylases in *K. marxianus* and functional characterization showed that *PDC1* contributes all the native pyruvate decarboxylase activity while *ARO10* contributes most of the phenylpyruvate decarboxylase activity with some contributions from *PDC1*. Thus, the disruption of *ARO10* does not guarantee the eradication of phenylpyruvate decarboxylase activity. Contrary to reports by others that PDC-null *K. marxianus* and *K. lactis* display no growth defect, it was found in this thesis that this was not the case because *pdc1* strain of *K. marxianus* showed poorer glucose consumption and formed lower biomass than wildtype (Chapter 4).

The focus of chapter 5 was on the derivatisation of coumaric acid into naringenin, which is the precursor of many plant flavonoids, by the action of *At4-CL3, AtCHS3* and *AtCHI*. Importantly, the derivatization requires 4 molecules of coenzyme A and 3 molecules of malonyl-CoA. Malonyl-CoA engineering through the overexpression of the native acetyl-CoA carboxylase (*ACC1*) and a mutant acetyl-CoA synthase of *Salmonella enterica* did not improve naringenin production. Although the expression of *At4-CL3,* the enzyme immediately below coumaric acid, depleted coumaric acid, the additional expression of *AtCHS3* and *AtCHI* resulted its poor depletion into downstream metabolites, suggesting that insufficient coenzyme A could be limiting the derivatization. Indeed, the overexpression of enzymes of the coenzyme A biosynthetic pathway increased naringenin production both in strains expressing malonyl-CoA biosynthetic genes and those not, and higher copy number of *At4-CL3* further increased the yield of naringenin. The multi-loci integration tool developed in chapter 2 was then applied for

the integration of multiple copies of *AtCHS3* at delta site, resulting in a strain that produced the highest naringenin yield of 1231µg/g.

2. Future perspectives on the work on developing *amdS* as selectable marker for *K. marxianus*

Although it was shown in this thesis that orthologues of the two genes that contribute to the acetamidase activities in *K. marxianus* are found in *K. lactis,* but while *K. marxianus* grows on acetamide *K. lactis* does not. Attempt to look for differences in the sequence of these proteins showed that conserved residues in amidases such as those that form the catalytic triad remain conserved in the *K. lactis* proteins. Further analysis of the proteins to determine why the *K. lactis* orthologues are not functional in the yeast is required. It is yet to be ascertained whether the inability of these proteins to mediate the growth of *K. lactis* on acetamide is because they are not functional in the yeast and may be in other yeasts such as *K. lactis.* Thus, an easy methodology to kick off this analysis is the expression of the orthologues in the *K. marxianus* strain without functional acetamidase activity that was constructed in chapter 2 (KmJA.272). This will inform on whether these proteins are at all functional as acetamidases. The protein fold can also be compared with their *K. marxianus* orthologues and other functional acetamidases. It is interesting that *K. marxianus* was not counter-selectable by fluoroacetamide as was demonstrated for *S. cerevisiae* elsewhere (Solis-Escalante *et al.*, 2013). Evidence showed that the sensitivity of a *K. marxianus* that is constitutively expressing *amdS* was not clear at the concentration that a similar *S. cerevisiae* strain is clearly sensitive to fluoroacetamide (2.3g/L) until 10 times this concentration was used. Nonetheless *K. marxianus* showed lower sensitivity at this higher fluoroacetamide concentration than *S. cerevisiae.* Thus, *amdS* was not recyclable by fluoroacetamide counter-selection with the direct repeat construct in this thesis. An alternative is to counter-select by an inducible lethal gene system in which the gene is placed beside the marker as was demonstrated for efficient seamless gene deletion in *S. cerevisiae* with the aid of galactose inducible cell wall protein 1 (Cwp1) gene (Hu *et al.*, 2019). However, this will require the identification of a functional and inducible lethal gene in *K. marxianus.* To that effect, the lethality of two lethal bacterial genes that have been previously demonstrated to function in *S. cerevisiae*, Zeta toxin of *Streptococcus pyogenes* toxin-antitoxin system (Zielenkiewicz *et al.*, 2009) and *nucA* from *Serratia marcescens* (Balan and Schenberg, 2005) was briefly evaluated in *K. marxianus*. It was shown that galactose induced expression of *nucA* prevents the growth of *K. marxianus* (Appendix 1, figure 1).

3. Indication of cytosolic acetyl-CoA supply mediators in PDC-null *K. marxianus*

It is interesting that PDC-null *K. marxianus* and *K. lactis* can grow in glucose as reported in chapter 4. This is in contrast to the *S. cerevisiae* mutant which does not grow at all in glucose due to its inability to supply cytosolic acetyl-CoA (Flikweert *et al.*, 1996; Pronk, Steensma and Van Dijken, 1996). Therefore, the question of how cytosolic acetyl-CoA which is necessary for normal cell function is supplied in both Kluyveromyces yeasts is yet to be answered. The possible sources are illustrated in figure 1. As follow up from chapter 4 in which the objective was to characterise the 2-oxo acid decarboxylase genes of *K. marxianus*, the genes encoding potential mediators of cytosolic acetyl-CoA supply in *pdc1* can be disrupted in this strain and the ability of derivative strains to grow on glucose can be evaluated to indicate the mediators that contribute to cytosolic acetyl-CoA supply. As shown in figure 1 of Appendix 2, among the mediators disrupted, only *pdc1* (KmJA.002) displayed growth defects and this strain lost its growth in the presence of the respiration blocker, antimycin A, in complex medium. This suggests that the source of cytosolic acetyl-CoA for supporting PDC-null growth is mitochondrial in origin. Thus, we sought to test whether the additional disruption of these mediators in KmJA.002 increased the strain's growth defect. Only the disruption of threonine aldolase Gly1 orthologues of *K. marxianus* of which there are two increased the growth defect of *pdc1* (Appendix 2, figure 2). This data is interesting because the overexpression of *ScGLY1* was reported to alleviate the phenotypes of PDC-null *S. cerevisiae* (Van Maris *et al.*, 2003), but the route as a possible source of cytosolic acetyl-CoA in *S. cerevisiae* was later excluded because of the low affinity of ScGly1 for threonine, which has relatively low intracellular concentration (Van Maris *et al.*, 2004). Since the *pdc1gly1A/B* strain maintains its ability to grow, it is clear that other mediators are at play in the *pdc1* and further investigations to obtain a wholistic knowledge of how cytosolic acetyl-CoA is supplied in *K. marxianus* could benefit the application of the yeast as cell factory. Given that the yeast has respiratory physiology, and that pyruvate is directed towards the TCA cycle to a greater extent than in *S. cerevisiae* (Sakihama *et al.*, 2019), knowledge of important mediators to overexpress for cytosolic acetyl-CoA supply from the mitochondrial pool can improve the production of acetyl-CoA derived molecules like naringenin.

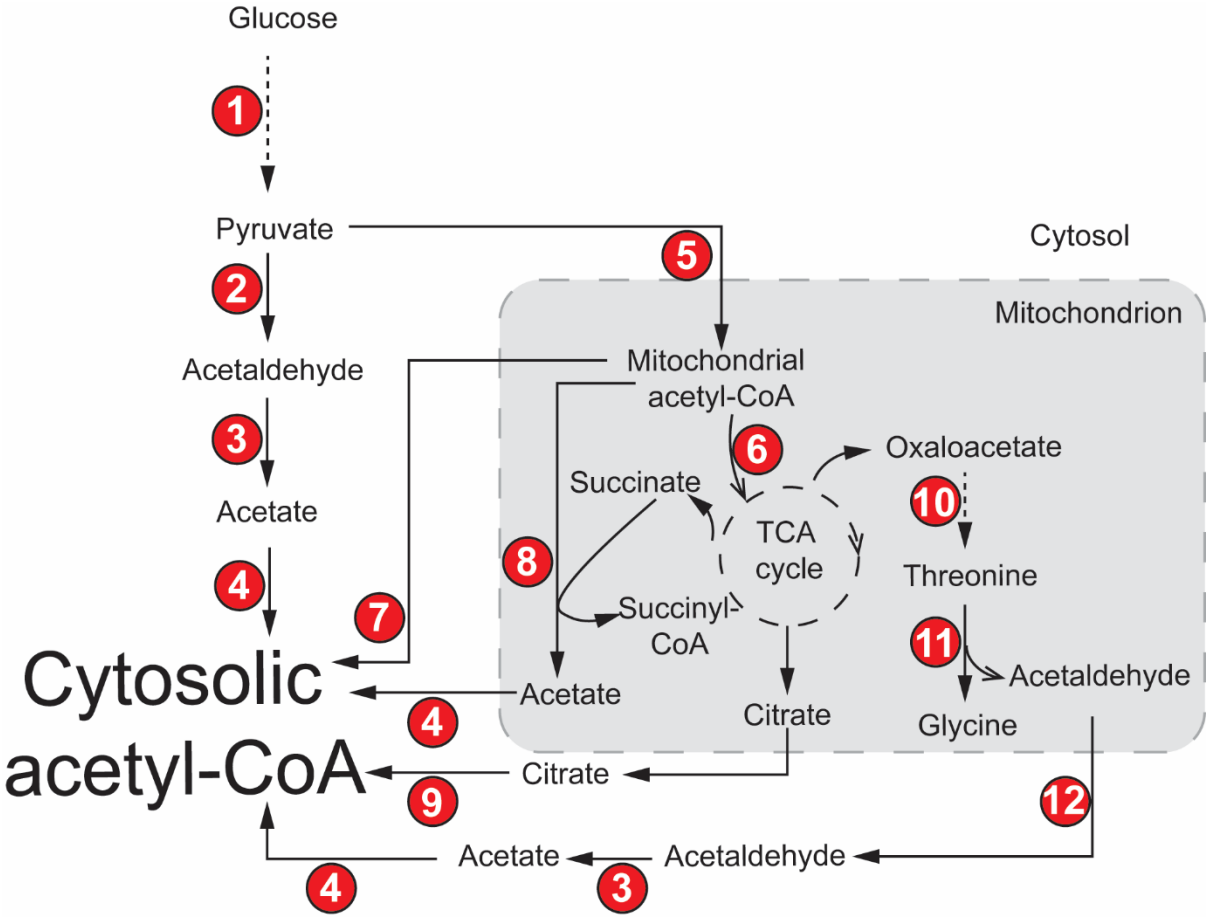


Figure 1. Potential sources of cytosolic acetyl-CoA in fungi. glucose is catabolised via glycolysis into pyruvate (1). Pyruvate is decarboxylated by pyruvate decarboxylase (2) into acetaldehyde from which acetate can be produced by the action of aldehyde dehydrogenase (3). Acetyl-CoA synthetase (4) can then convert acetate into acetyl-CoA. Pyruvate is alternatively taken into the mitochondrion to be acted upon by pyruvate dehydrogenase complex (5), forming mitochondrial acetyl-CoA. This mitochondrial acetyl-CoA can be taken up by the TCA cycle (6), exported into the cytosol in the acetyl-CoA form by carnitine dependent transport (7) or acted upon by acetyl-CoA hydroxylase (8) to transfer the CoA group onto succinate for cytosolic transport in the acetyl form. TCA cycle intermediates also serve as sources of cytosolic acetyl-CoA through cytosolic transportation of citrate for conversion into cytosolic acetyl-CoA by ATP-citrate lyase (9), but this reaction is absent in most budding yeasts. Oxaloacetate serves as precursor for threonine biosynthesis (10) and cleavage into glycine and acetaldehyde by threonine aldolase (11). Acetaldehyde may be released into the cytosol (12) for biosynthesis of cytosolic acetyl-CoA by aldehyde dehydrogenase (3) and acetyl-CoA synthase (4). These mediators can be disrupted in order to test their contribution to cytosolic acetyl-CoA supply by disrupting the genes encoding them. These genes were identified by bioinformatics. For carnitine shuttles, their requirement for carnitine means that this compound must be supplied in the media or produced by the yeast for the shuttle to function. Unlike *S. cerevsiae* which cannot produce carnitine due to the absence of the genes encoding the biosynthetic enzymes, the complete biosynthetic pathway was found to be present in *K. marxianus*. Thus, the shuttle can be made inactive by disrupting the first enzyme (tri-methyl-lysine dioxygenase (TMLD)) of the carnitine pathway because the strain becomes unable to produce carnitine.

4. Future perspectives on engineering *K. marxianus* for flavonoids production

Although the successful derivatisation of coumaric acid into naringenin was reported in Chapter 5, the depletion of coumaric acid into downstream metabolites in the pathway is still

inefficient because even the best naringenin producers maintained high coumaric acid production. The expression of *At4-CL3* without *AtCHS3* and *AtCHI* resulted in complete loss of coumaric acid production. However, coumaric acid biosynthesis emerged when *AtCHS3* and *AtCHI* were additionally expressed. These results indicated that coenzyme A could be limiting, and the expression of coenzyme A biosynthetic genes proved to be major strategy for derivatising coumaric acid into naringenin. The configuration used to express the coenzyme A genes may however be suboptimal as the same constitutive *KmPDC1* promoter and *ScPGK1* terminator were used for all 7 genes. To optimise the expression of these genes and identify the configuration that allows optimum coenzyme A availability for naringenin production, combinatorial assembly based on the yeast toolkit (YTK) standard used throughout this thesis can be developed for further engineering the coumaric acid/naringenin producing strains that were constructed in this thesis. Below coumaroyl-CoA, phloretic acid biosynthesis by enoyl-CoA reductase Tsc13 competes against naringenin biosynthesis by chalcone synthase and the naringenin producers in chapter 5 consequently also produced phloretic acid, making Tsc13 a target for further engineering. This can involve the repression of the native *TSC13* expression or the replacement of this gene with plant variants that have low specificity for coumaroyl-CoA but can still reduce enoyl-CoA as was demonstrated in *S. cerevisiae* (Lehka *et al.*, 2017). Given the industrially attractive properties of *K.* marxianus, future work should also evaluate the ability of the aromatics producing strains to grow in industrially relevant conditions such as high temperatures and in various feedstocks.

References

Aditiya, H. B. *et al.* (2016) 'Second generation bioethanol production: A critical review', *Renewable and Sustainable Energy Reviews*. Elsevier, 66, pp. 631–653. doi: 10.1016/j.rser.2016.07.015.

Akama-Garren, E. H. *et al.* (2016) 'A Modular Assembly Platform for Rapid Generation of DNA Constructs', *Scientific Reports*. Nature Publishing Group, 6, pp. 1–9. doi: 10.1038/srep16836.

Alani, E., Cao, L. and Kleckner, N. (1987) 'A Method for Gene Disruption That Allows Repeated Use of URA3 Selection in the Construction of Multiply Disrupted Yeast Strains', *Genetics*, 116(4), pp. 541–545. doi: 10.1093/genetics/116.4.541.

Asunis, F. *et al.* (2020) 'The dairy biorefinery: Integrating treatment processes for cheese whey valorisation', *Journal of Environmental Management*. Elsevier Ltd, 276(June), p. 111240. doi: 10.1016/j.jenvman.2020.111240.

Atsumi, S., Hanai, T. and Liao, J. C. (2008) 'Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels', *Nature*, 451(7174), pp. 86–89. doi: 10.1038/nature06450.

Ayora, S. *et al.* (2011) 'Double-strand break repair in bacteria: A view from Bacillus subtilis', *FEMS Microbiology Reviews*, 35(6), pp. 1055–1081. doi: 10.1111/j.1574-6976.2011.00272.x.

Balan, A. and Schenberg, A. C. G. (2005) 'A conditional suicide system for Saccharomyces cerevisiae relying on the intracellular production of the Serratia marcescens nuclease', *Yeast*, 22(3), pp. 203–212. doi: 10.1002/yea.1203.

Bang, H. B. *et al.* (2016) 'Metabolic engineering of Escherichia coli for the production of cinnamaldehyde', *Microbial Cell Factories*. BioMed Central, 15(1), pp. 1–12. doi: 10.1186/s12934-016-0415-9.

Baskaran, S. M. *et al.* (2021) 'Valorization of biodiesel side stream waste glycerol for rhamnolipids production by Pseudomonas aeruginosa RS6', *Environmental Pollution*, 276. doi: 10.1016/j.envpol.2021.116742.

Becker, J. V. W. *et al.* (2003) 'Metabolic engineering of Saccharomyces cerevisiae for the synthesis of the wine-related antioxidant resveratrol', *FEMS Yeast Research*, 4(1), pp. 79–85. doi: 10.1016/S1567-1356(03)00157-0.

Beekwilder, J. *et al.* (2006) 'Production of resveratrol in recombinant microorganisms', *Applied and Environmental Microbiology*, 72(8), pp. 5670–5672. doi: 10.1128/AEM.00609-06.

Bensen, E. S., Yeung, B. G. and Payne, G. S. (2001) 'Ric1p and the Ypt6p GTPase function in a common pathway required for localization of trans-Golgi network membrane proteins', *Molecular Biology of the Cell*, 12(1), pp. 13–26. doi: 10.1091/mbc.12.1.13.

Bertacchi, S. *et al.* (2022) 'Interdependence between lignocellulosic biomasses, enzymatic hydrolysis and yeast cell factories in biorefineries', *Microbial Biotechnology*, 15(3), pp. 985–995. doi: 10.1111/1751-7915.13886.

Bever, D., Wheeldon, I. and Da Silva, N. (2022) 'RNA polymerase II-driven CRISPR-Cas9 system for efficient non-growth-biased metabolic engineering of Kluyveromyces marxianus', *Metabolic Engineering Communications*. Elsevier B.V., 15(September), p. e00208. doi: 10.1016/j.mec.2022.e00208.

Bianchi, M. M., Tizzani, L. and Frontal, L. (1996) 'The "petite-negative" yeast Kluyveromyces lactis has a single gene expressing pyruvate decarboxylase activity', 19, pp. 27–36. Available at: https://pubmed.ncbi.nlm.nih.gov/8821934/.

Bilal, M. *et al.* (2022) 'Bioprospecting Kluyveromyces marxianus as a Robust Host for Industrial Biotechnology', *Frontiers in Bioengineering and Biotechnology*, 10(April), pp. 1–18. doi: 10.3389/fbioe.2022.851768.

Binder, A. *et al.* (2014) 'A modular plasmid assembly kit for multigene expression, gene silencing and silencing rescue in plants', *PLoS ONE*, 9(2). doi: 10.1371/journal.pone.0088218.

Bisquert, R. *et al.* (2022) 'Metabolic engineering of Saccharomyces cerevisiae for hydroxytyrosol overproduction directly from glucose', *Microbial Biotechnology*, 15(5), pp. 1499–1510. doi: 10.1111/1751-7915.13957.

Blount, J. W. *et al.* (2000) 'Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway', *Plant Physiology*, 122(1), pp. 107–116. doi: 10.1104/pp.122.1.107.

Bogdanove, A. J. and Voytas, D. F. (2011) 'TAL effectors: Customizable proteins for DNA targeting', *Science*, 333(6051), pp. 1843–1846. doi: 10.1126/science.1204094.

Bowman, S. B. *et al.* (1992) 'Positive regulation of the LPD1 gene of Saccharomyces cerevisine by the HAP2/HAP3/HAP4 activation system', *MGG Molecular & General Genetics*, 231(2), pp. 296–303. doi: 10.1007/BF00279803.

Brachmann, C. B. *et al.* (1998) 'Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications', *Yeast*, 14(2), pp. 115–132. doi: 10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2.

Brown, J. F. and Dawes, I. W. (1990) 'Regulation of chorismate mutase in Saccharomyces cerevisiae', *MGG Molecular & General Genetics*, 220(2), pp. 283–288. doi: 10.1007/BF00260495.

Burgess, S. M. and Kleckner, N. (1999) 'Collisions between yeast chromosomal loci in vivo are governed by three layers of organization', *Genes and Development*, 13(14), pp. 1871–1883. doi: 10.1101/gad.13.14.1871.

Byrne, K. P. and Wolfe, K. H. (2005) 'The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species', *Genome Research*, 15(10), pp. 1456–1461. doi: 10.1101/gr.3672305.

Caballero, R. *et al.* (1995) 'Evaluation of Kluyveromyces marxianus as baker's yeast', *Food Research International*, 28(1), pp. 37–41. doi: 10.1016/0963-9969(95)93329-S.

Cai, P. *et al.* (2021) 'Recombination machinery engineering facilitates metabolic engineering of the industrial yeast Pichia pastoris', *Nucleic Acids Research*. Oxford University Press, 49(13), pp. 7791–7805. doi: 10.1093/nar/gkab535.

Cai, P., Gao, J. and Zhou, Y. (2019) 'CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications.', *Microbial cell factories*, 18(1), p. 63. doi: 10.1186/s12934-019-1112-2.

Cai, W. and Zhang, W. (2018) 'Engineering modular polyketide synthases for production of biofuels and industrial chemicals', *Current Opinion in Biotechnology*. Elsevier Ltd, 50, pp. 32–38. doi: 10.1016/j.copbio.2017.08.017.

Camakaris, J. and Pittard, J. (1974) 'Purification and Properties of 3-Deoxy-D-Arabinoheptulosonic Acid-7-Phosphate Synthetase (trp) from Escherichia coli', 120(2), pp. 590–597. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC245816/.

Campero-Basaldua, C. *et al.* (2017) 'Diversification of the kinetic properties of yeast NADP-glutamate-dehydrogenase isozymes proceeds independently of their evolutionary origin', *MicrobiologyOpen*, 6(2), pp. 1–18. doi: 10.1002/mbo3.419.

Cao, M. *et al.* (2020) 'Building microbial factories for the production of aromatic amino acid pathway derivatives: From commodity chemicals to plant-sourced natural products', *Metabolic Engineering*. Elsevier Inc., 58(August), pp. 94–132. doi: 10.1016/j.ymben.2019.08.008.

Cardenas, J. and Da Silva, N. A. (2014) 'Metabolic engineering of Saccharomyces cerevisiae for the production of triacetic acid lactone', *Metabolic Engineering*. Elsevier, 25, pp. 194–203. doi: 10.1016/j.ymben.2014.07.008.

del Carmen Vasallo, M. *et al.* (2006) 'Cell wall proteins of Kluyveromyces fragilis: Surface and emulsifying properties', *Lwt*, 39(7), pp. 729–739. doi: 10.1016/j.lwt.2005.06.003.

Casini, A. *et al.* (2014) 'One-pot DNA construction for synthetic biology : the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy', 42(1), pp. 1–13. doi: 10.1093/nar/gkt915.

Chandran, H. *et al.* (2020) 'Plant tissue culture as a perpetual source for production of industrially important bioactive compounds', *Biotechnology Reports*. Elsevier B.V., 26, p. e00450. doi: 10.1016/j.btre.2020.e00450.

Chandran, S. S. *et al.* (2003) 'Phosphoenolpyruvate availability and the biosynthesis of shikimic acid', *Biotechnology Progress*, 19(3), pp. 808–814. doi: 10.1021/bp025769p.

Chang, I. *et al.* (2009) 'Proteomic profiling of tandem affinity purified 14-3-3 protein complexes in Arabidopsis thaliana', pp. 2967–2985. doi: 10.1002/pmic.200800445.

Chattoo, B. B. *et al.* (1979) 'Selection of lys2 Mutants of the Yeast Saccharomyces cerevisiae by the Utilization of alpha-AMINOADIPATE', *Genetics*, pp. 51–65. Available at: https://pubmed.ncbi.nlm.nih.gov/17248969/.

Chen, K. *et al.* (2012) 'Deletion of the aroK gene is essential for high shikimic acid accumulation through the shikimate pathway in E . coli', *BIORESOURCE TECHNOLOGY*. Elsevier Ltd, 119, pp. 141–147. doi: 10.1016/j.biortech.2012.05.100.

Chen, L. *et al.* (2018) 'Discovery of feed-forward regulation in L-tryptophan biosynthesis and its use in metabolic engineering of E. coli for efficient tryptophan bioproduction', *Metabolic Engineering*. Elsevier Inc., 47(May), pp. 434–444. doi: 10.1016/j.ymben.2018.05.001.

Chen, R. *et al.* (2022) 'Engineering cofactor supply and recycling to drive phenolic acid biosynthesis in yeast', *Nature Chemical Biology*, 18(5), pp. 520–529. doi: 10.1038/s41589-022-01014-6.

Chen, Y. *et al.* (2015) 'Ach1 is involved in shuttling mitochondrial acetyl units for cytosolic C2 provision in Saccharomyces cerevisiae lacking pyruvate decarboxylase', *FEMS Yeast Research*, 15(3), pp. 1–8. doi: 10.1093/femsyr/fov015.

Chen, Y., Li, F. and Nielsen, J. (2022) 'Genome-scale modeling of yeast metabolism: retrospectives and perspectives', *FEMS Yeast Research*, 22(1), pp. 1–9. doi: 10.1093/femsyr/foac003.

Chen, Y., Siewers, V. and Nielsen, J. (2012) 'Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in Saccharomyces cerevisiae', *PLoS ONE*, 7(8). doi: 10.1371/journal.pone.0042475.

Chia, M. *et al.* (2012) 'Triacetic acid lactone as a potential biorenewable platform chemical', *Green Chemistry*, 14(7), pp. 1850–1853. doi: 10.1039/c2gc35343a.

Cho, M. H. *et al.* (2007) 'Phenylalanine biosynthesis in Arabidopsis thaliana: Identification and characterization of arogenate dehydratases', *Journal of Biological Chemistry*. Â© 2007 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 282(42), pp. 30827–30835. doi: 10.1074/jbc.M702662200.

Choo, J. H. *et al.* (2018) 'Molecular and functional characterization of two pyruvate decarboxylase genes, PDC1 and PDC5, in the thermotolerant yeast Kluyveromyces marxianus', *Applied Microbiology and Biotechnology*. Applied Microbiology and Biotechnology, 102(8), pp. 3723–3737. doi: 10.1007/s00253-018-8862-3.

Ciamponi, F. E. *et al.* (2022) 'Multi-omics network model reveals key genes associated with p-coumaric acid stress response in an industrial yeast strain', *Scientific Reports*. Nature Publishing Group UK, 12(1), pp. 1–17. doi: 10.1038/s41598-022-26843-2.

Courdavault, V. *et al.* (2021) 'Metabolic engineering for plant natural products biosynthesis: New procedures, concrete achievements and remaining limits', *Natural Product Reports*. Royal Society of Chemistry, 38(12), pp. 2145–2153. doi: 10.1039/d0np00092b.

Curran, K. A. *et al.* (2013) 'Metabolic engineering of muconic acid production in Saccharomyces cerevisiae', *Metabolic Engineering*. Elsevier, 15(1), pp. 55–66. doi: 10.1016/j.ymben.2012.10.003.

Dai, J. *et al.* (2020) 'Sc3 . 0 : revamping and minimizing the yeast genome'. Genome Biology, pp. 1–4.

David, F. and Siewers, V. (2015) 'Advances in yeast genome engineering.', *FEMS yeast research*, 15(1), pp. 1–14. doi: 10.1111/1567-1364.12200.

Delneri, D. *et al.* (2000) 'Exploring redundancy in the yeast genome: An improved strategy for use of the cre-loxP system', *Gene*, 252(1–2), pp. 127–135. doi: 10.1016/S0378-1119(00)00217-1.

DeLuna, A. *et al.* (2001) 'NADP-glutamate dehydrogenase isoenzymes of Saccharomyces cerevisiae: Purification, kinetic properties, and physiological roles', *Journal of Biological Chemistry*, 276(47), pp. 43775–43783. doi: 10.1074/jbc.M107986200.

Dicarlo, J. E. *et al.* (2013) 'Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems', *Nucleic Acids Research*, 41(7), pp. 4336–4343. doi: 10.1093/nar/gkt135.

Donzella, L. *et al.* (2021) 'Identification of novel pentose transporters in Kluyveromyces marxianus using a new screening platform', *FEMS Yeast Research*, 21(4), pp. 1–11. doi: 10.1093/femsyr/foab026.

Doong, R. L., Ganson, R. J. and Jensen, R. A. (1993) 'Plastid-localized 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DS-Mn): the early-pathway target of sequential feedback inhibition in higher plants', *Plant, Cell & Environment*, 16(4), pp. 393–402. doi: 10.1111/j.1365-3040.1993.tb00885.x.

Doong, R. Lou *et al.* (1992) 'The cytosolic isoenzyme of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase in Spinacia oleracea and other higher plants : extreme substrate ambiguity and other properties'.

Duportet, X. *et al.* (2014) 'A platform for rapid prototyping of synthetic gene networks in mammalian cells', 42(21), pp. 13440–13451. doi: 10.1093/nar/gku1082.

Dyer, W. E. *et al.* (1989) 'Wounding induces the first enzyme of the shikimate pathway in Solanaceae', *Proceedings of the National Academy of Sciences*, 86(19), pp. 7370–7373. doi: 10.1073/pnas.86.19.7370.

Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Elani, Y. (2021) 'Interfacing Living and Synthetic Cells as an Emerging Frontier in Synthetic Biology', *Angewandte Chemie*, 133(11), pp. 5662–5671. doi: 10.1002/ange.202006941.

Elledge, S. J. *et al.* (1991) 'λYES: A multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and Escherichia coli mutations', *Proceedings of the National Academy of Sciences of the United States of America*, 88(5), pp. 1731–1735. doi: 10.1073/pnas.88.5.1731.

Engler, C. *et al.* (2009) 'Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes', *PLoS ONE*, 4(5). doi: 10.1371/journal.pone.0005553.

Engler, C. *et al.* (2014) 'A Golden Gate Modular Cloning Toolbox for Plants'.

Entus, R., Poling, M. and Herrmann, K. M. (2002) 'Redox Regulation of Arabidopsis 3-Deoxy-D - arabino - Heptulosonate 7-Phosphate Synthase 1', 129(August), pp. 1866–1871. doi: 10.1104/pp.002626.1866.

Evans, S. G., Kelley, L. C. and Potts, M. D. (2015) 'The potential impact of second-generation biofuel landscapes on at-risk species in the US', *GCB Bioenergy*, 7(2), pp. 337–348. doi: 10.1111/gcbb.12131.

Figueroa Pires, A. *et al.* (2021) 'Dairy By-Products: A Review on the Valorization of Whey and Second Cheese Whey', *Foods*, 10, p. 1067. Available at: https://www.mdpi.com/journal/foods.

Fleck, C. B. and Brock, M. (2009) 'Re-characterisation of Saccharomyces cerevisiae Ach1p: Fungal CoA-transferases are involved in acetic acid detoxification', *Fungal Genetics and Biology*. Elsevier Inc., 46(6–7), pp. 473–485. doi: 10.1016/j.fgb.2009.03.004.

Flikweert, M. T. *et al.* (1996) 'Pyruvate Decarboxylase:An Indispensable Enzyme for Growth of Saccharomyces cerevisiae on Glucose', 12, pp. 247–257.

Flikweert, M. T. *et al.* (1999) 'Growth requirements of pyruvate-decarboxylase-negative Saccharomyces cerevisiae', *FEMS Microbiology Letters*, 174(1), pp. 73–79. doi: 10.1016/S0378-1097(99)00124-X.

Fraczek, M. G., Naseeb, S. and Delneri, D. (2018) 'History of genome editing in yeast', *Yeast*. John Wiley and Sons Ltd, 35(5), pp. 361–368. doi: 10.1002/yea.3308.

Fraser, C. M. and Chapple, C. (2011) 'The phenylpropanoid pathway in Arabidopsis.', *The arabidopsis book*, 9, p. e0152. doi: 10.1199/tab.0152.

Fu, J. *et al.* (2016) 'Development of a Candida glabrata dominant nutritional transformation marker utilizing the Aspergillus nidulans acetamidase gene (amdS)', *FEMS Yeast Research*, 16(3), pp. 1–9. doi: 10.1093/femsyr/fow023.

Fuentes, P., Armarego-Marriott, T. and Bock, R. (2018) 'Plastid transformation and its application in metabolic engineering', *Current Opinion in Biotechnology*. Elsevier Ltd, 49, pp. 10–15. doi: 10.1016/j.copbio.2017.07.004.

Gaertner, F. H. and DeMoss, J. A. (1969) 'Purification and characterization of a multienzyme complex in the tryptophan pathway of Neurospora crassa.', *Journal of Biological Chemistry*. Â© 1969 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 244(10), pp. 2716–2725. doi: 10.1016/s0021-9258(18)83455-x.

Gao, M. *et al.* (2017) 'Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics', *ACS Synthetic Biology*, 6(1), pp. 29–38. doi: 10.1021/acssynbio.6b00132.

Gao, S. *et al.* (2016) 'Multiplex gene editing of the Yarrowia lipolytica genome using the CRISPR-Cas9 system', *Journal of Industrial Microbiology and Biotechnology*. Springer Berlin Heidelberg, 43(8), pp. 1085–1093. doi: 10.1007/s10295-016-1789-8.

Gao, S., Xu, X., *et al.* (2020) 'Efficient Biosynthesis of (2 S)-Eriodictyol from (2 S)-Naringenin in Saccharomyces cerevisiae through a Combination of Promoter Adjustment and Directed Evolution', *ACS Synthetic Biology*, 9(12), pp. 3288–3297. doi: 10.1021/acssynbio.0c00346.

Gao, S., Gao, S., *et al.* (2020) 'Promoter-Library-Based Pathway Optimization for Efficient (2 S)-Naringenin Production from p-Coumaric Acid in Saccharomyces cerevisiae', *Journal of Agricultural and Food Chemistry*, 68(25), pp. 6884–6891. doi: 10.1021/acs.jafc.0c01130.

García, S. *et al.* (2017) 'The Role of the ydiB Gene , Which Encodes Quinate / Shikimate Dehydrogenase , in the Production of Quinic , Dehydroshikimic and Shikimic Acids in a PTS – Strain of Escherichia coli', 2016, pp. 11–21. doi: 10.1159/000450611.

Geddes, B. A., Mendoza-suárez, M. A. and Poole, P. S. (2019) 'A Bacterial Expression Vector Archive (BEVA) for Flexible Modular Assembly of Golden Gate-Compatible Vectors', 9(January), pp. 1–14. doi: 10.3389/fmicb.2018.03345.

Gibson, D. G. *et al.* (2008) 'Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome', 319(February), pp. 1215–1220.

Gibson, D. G. *et al.* (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', 6(5), pp. 12–16. doi: 10.1038/NMETH.1318.

Gibson, D. G. *et al.* (2010) 'Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome', 329(July), pp. 52–56.

Gorlach, J. *et al.* (1995) 'Temporally distinct accumulation of transcripts encoding enzymes of the prechorismate pathway in elicitor-treated , cultured tomato cells', 92(April), pp. 3166–3170.

Gottardi, M., Knudsen, J. D., *et al.* (2017) 'De novo biosynthesis of trans-cinnamic acid derivatives in Saccharomyces cerevisiae', *Applied Microbiology and Biotechnology*. Applied Microbiology and Biotechnology, 101(12), pp. 4883–4893. doi: 10.1007/s00253-017-8220-x.

Gottardi, M., Grün, P., *et al.* (2017) 'Optimisation of trans-cinnamic acid and hydrocinnamyl alcohol production with recombinant Saccharomyces cerevisiae and identification of cinnamyl methyl ketone as a by-product', *FEMS Yeast Research*, 17(8), pp. 1–10. doi: 10.1093/femsyr/fox091.

Graf, R., Mehmann, B. and Braus, G. H. (1993) 'Analysis of feedback-resistant anthranilate synthases from Saccharomyces cerevisiae', *Journal of Bacteriology*, 175(4), pp. 1061–1068. doi: 10.1128/jb.175.4.1061-1068.1993.

Gu, Y., Ma, J., Zhu, Y., Ding, X., *et al.* (2020) 'Engineering Yarrowia lipolytica as a chassis for de novo synthesis of five aromatic-derived natural products and chemicals', *bioRxiv*, pp. 1–27. doi: 10.1101/2020.04.04.025288.

Gu, Y., Ma, J., Zhu, Y. and Xu, P. (2020) 'Refactoring Ehrlich Pathway for High-Yield 2-Phenylethanol Production in Yarrowia lipolytica', *ACS Synthetic Biology*, 9(3), pp. 623–633. doi: 10.1021/acssynbio.9b00468.

Gueldener, U. *et al.* (2002) 'A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast', *Nucleic acids research*, 30(6), p. 23. doi: 10.1093/nar/30.6.e23.

Guo, W. *et al.* (2019) 'Rational engineering of chorismate-related pathways in Saccharomyces cerevisiae for improving tyrosol production', *Frontiers in Bioengineering and Biotechnology*, 7(JUL), pp. 1–7. doi: 10.3389/fbioe.2019.00152.

Guo, W. *et al.* (2020) 'Rewiring central carbon metabolism for tyrosol and salidroside production in Saccharomyces cerevisiae', *Biotechnology and Bioengineering*, 117(8), pp. 2410–2419. doi: 10.1002/bit.27370.

Hamilton, M. *et al.* (2020) 'Identification of a Yarrowia lipolytica acetamidase and its use as a yeast genetic marker', *Microbial Cell Factories*. BioMed Central, 19(1), pp. 1–9. doi: 10.1186/s12934-020-1292-9.

Hang, Y. D., Woodams, E. E. and Hang, L. E. (2003) 'Utilization of corn silage juice by Klyuveromyces marxianus', *Bioresource Technology*, 86(3), pp. 305–307. doi: 10.1016/S0960-8524(02)00170-0.

Hartmann, M. *et al.* (2003) 'Evolution of feedback-inhibited □ ¯␣ barrel isoenzymes by gene duplication and a single mutation', 100(3). Available at: https://pubmed.ncbi.nlm.nih.gov/12540830/.

Hasegawa, T. *et al.* (2015) 'Consequence of Climate Mitigation on the Risk of Hunger', *Environmental Science and Technology*, 49(12), pp. 7245–7253. doi: 10.1021/es5051748.

Hassing, E. J. *et al.* (2019) 'Connecting central carbon and aromatic amino acid metabolisms to improve de novo 2-phenylethanol production in Saccharomyces cerevisiae', *Metabolic Engineering*. Elsevier Inc., 56(July), pp. 165–180. doi: 10.1016/j.ymben.2019.09.011.

Hassing, E. J. *et al.* (2021) 'Elimination of aromatic fusel alcohols as by-products of Saccharomyces cerevisiae strains engineered for phenylpropanoid production by 2-oxo-acid decarboxylase replacement', *Metabolic Engineering Communications*. Elsevier B.V., 13(July), p. e00183. doi: 10.1016/j.mec.2021.e00183.

Hazelwood, L. A. *et al.* (2008) 'The Ehrlich pathway for fusel alcohol production: A century of research on Saccharomyces cerevisiae metabolism', *Applied and Environmental Microbiology*, 74(8), pp. 2259–2266. doi: 10.1128/AEM.02625-07.

Herrmann, K. M. (1995) 'The shikimate pathway: Early steps in the biosynthesis of aromatic compounds', *Plant Cell*, 7(7), pp. 907–919. doi: 10.2307/3870046.

Herrmann, K. M. and Weaver, L. M. (1999) 'The shikimate pathway', *Annual Review of Plant Biology*, 50, pp. 473–503. doi: 10.1146/annurev.arplant.50.1.473.

Horwitz, A. A. *et al.* (2015) 'Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas', *Cell Systems*. The Authors, 1(1), pp. 88–96. doi: 10.1016/j.cels.2015.02.001.

Houfani, A. A. *et al.* (2020) 'Insights from enzymatic degradation of cellulose and hemicellulose to fermentable sugars– a review', *Biomass and Bioenergy*. Elsevier Ltd, 134(February), p. 105481. doi: 10.1016/j.biombioe.2020.105481.

Hu, Y. *et al.* (2019) 'A new strategy for seamless gene editing and marker recycling in Saccharomyces cerevisiae using lethal effect of Cwp1', *MicrobiologyOpen*, 8(6), pp. 1–7. doi: 10.1002/mbo3.750.

Hua, Y. *et al.* (2019) 'Release of glucose repression on xylose utilization in Kluyveromyces marxianus to enhance glucose-xylose co-utilization and xylitol production from corncob hydrolysate', *Microbial Cell Factories*. BioMed Central, 18(1), pp. 1–18. doi: 10.1186/s12934-019-1068-2.

Hughes, E. H. *et al.* (2004) 'Metabolic engineering of the indole pathway in Catharanthus roseus hairy roots and increased accumulation of tryptamine and serpentine', *Metabolic Engineering*, 6(4), pp. 268–276. doi: 10.1016/j.ymben.2004.03.002.

Hutchison, A. C. *et al.* (2016) 'Design and synthesis of a minimal bacterial genome', 6253(June). doi: 10.1126/science.aad6253.

Hwang, E. Il *et al.* (2003) 'Production of plant-specific flavanones by Escherichia coli containing an artificial gene cluster', *Applied and Environmental Microbiology*, 69(5), pp. 2699–2706. doi: 10.1128/AEM.69.5.2699-2706.2003.

Itaya, M. *et al.* (2008) 'Bottom-up genome assembly using the Bacillus subtilis genome vector', 5(1), pp. 2001–2003. doi: 10.1038/NMETH1143.

Jain, A. *et al.* (2022) 'Bioenergy and bio-products from bio-waste and its associated modern circular economy: Current research trends, challenges, and future outlooks', *Fuel*. Elsevier Ltd, 307(April 2021), p. 121859. doi: 10.1016/j.fuel.2021.121859.

Jakočiūnas, T. *et al.* (2015) 'Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae.', *Metabolic engineering*, 28, pp. 213–222. doi: 10.1016/j.ymben.2015.01.008.

Jansen, G. *et al.* (2005) 'Drag&Drop cloning in yeast', *Gene*, 344, pp. 43–51. doi: 10.1016/j.gene.2004.10.016.

Jinek, M. *et al.* (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.

Jinek, M. *et al.* (2013) 'RNA-programmed genome editing in human cells', *eLife*. doi: 10.7554/eLife.00471.

Juergens, H. *et al.* (2018) 'Genome editing in Kluyveromyces and Ogataea yeasts using a broad-host-range Cas9/gRNA co-expression plasmid', *FEMS Yeast Research*, 18(3), pp. 1–16. doi: 10.1093/femsyr/foy012.

Juhas, M. and Ajioka, J. W. (2017) 'Critical Reviews in Biotechnology High molecular weight DNA assembly in vivo for synthetic biology applications High molecular weight DNA assembly in vivo for synthetic biology applications', 8551. doi: 10.3109/07388551.2016.1141394.

Juminaga, D. *et al.* (2012) 'Modular engineering of L-tyrosine production in Escherichia coli', *Applied and Environmental Microbiology*, 78(1), pp. 89–98. doi: 10.1128/AEM.06017-11.

Jung, J. Y. *et al.* (2016) '13C metabolite profiling to compare the central metabolic flux in two yeast strains', *Biotechnology and Bioprocess Engineering*, 21(6), pp. 814–822. doi: 10.1007/s12257-016-0536-3.

Kádár, Z., Szengyel, Z. and Réczey, K. (2004) 'Simultaneous saccharification and fermentation (SSF) of industrial wastes for the production of ethanol', *Industrial Crops and Products*, 20(1), pp. 103–110. doi: 10.1016/j.indcrop.2003.12.015.

Kanaris, M. *et al.* (2022) 'Elevated tyrosine results in the cytosolic retention of 3-deoxy-d-arabino-heptulosonate 7-phosphate synthase in Arabidopsis thaliana', *Plant Journal*, 109(4), pp. 789–803. doi: 10.1111/tpj.15590.

Kaneko, S. *et al.* (2009) 'Marker-disruptive gene integration and URA3 recycling for multiple gene manipulation in Saccharomyces cerevisiae', *Applied Microbiology and Biotechnology*, 83(4), pp. 783–789. doi: 10.1007/s00253-009-2038-0.

Käufer, N. F. *et al.* (1980) 'Cycloheximide resistance in yeast: the gene and its protein', *Methods*, 12(21), pp. 8235–8251.

Keesey, J. K. and Demoss, J. A. (1982) 'Cloning of the trp-I gene from Neurospora crassa by complementation of a trpC mutation in Escherichia coli', *Journal of Bacteriology*, 152(2), pp. 954–958. Available at: https://pubmed.ncbi.nlm.nih.gov/6290461/.

Keith, B. *et al.* (1991) 'Differential induction of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase genes in Arabidopsis thaliana by wounding and pathogenic attack', *Proceedings of the National Academy of Sciences of the United States of America*, 88(19), pp. 8821–8825. doi: 10.1073/pnas.88.19.8821.

Kelly, J. M. and Hynes, M. J. (1985) 'Transformation of Aspergillus niger by the amdS gene of Aspergillus nidulans.', *The EMBO journal*, 4(2), pp. 475–479. doi: 10.1002/j.1460-2075.1985.tb03653.x.

Kelly, M. J. and Hynes, J. M. (1974) 'Transformation of Asoergillus niger by the amdS gene of Aspergillus nidulans Joan', *EMBO J*, 4(2), pp. 447–510. doi: 10.1007/978-1-4899-1710-2_26.

Kiers, J. *et al.* (1998) 'Regulation of alcoholic fermentation in batch and chemostat cultures of Kluyveromyces lactis CBS 2359', *Yeast*, 14(5), pp. 459–469. doi: 10.1002/(SICI)1097-0061(19980330)14:5<459::AID-YEA248>3.0.CO;2-O.

Kim, T. Y., Lee, S. W. and Oh, M. K. (2014) 'Biosynthesis of 2-phenylethanol from glucose with genetically engineered Kluyveromyces marxianus', *Enzyme and Microbial Technology*. Elsevier Inc., 61–62, pp. 44–47. doi: 10.1016/j.enzmictec.2014.04.011.

Knogge, W., Schmelzer, E. and Weissenböck, G. (1986) 'The role of chalcone synthase in the regulation of flavonoid biosynthesis in developing oat primary leaves', *Archives of Biochemistry and Biophysics*, 250(2), pp. 364–372. doi: 10.1016/0003-9861(86)90738-1.

Ko, Y. S. *et al.* (2020) 'Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production', *Chemical Society Reviews*. Royal Society of Chemistry, 49(14), pp. 4615–4636. doi: 10.1039/d0cs00155d.

Kobayashi, M. *et al.* (1997) 'Identification of active sites in amidase: Evolutionary relationship between amide bond- and peptide bond-cleaving enzymes', *Proceedings of the National Academy of Sciences of the United States of America*, 94(22), pp. 11986–11991. doi: 10.1073/pnas.94.22.11986.

Kogure, T. *et al.* (2016) 'Metabolic engineering of Corynebacterium glutamicum for shikimate overproduction by growth-arrested cell reaction', *Metabolic Engineering*. Elsevier, 38, pp. 204–216. doi: 10.1016/j.ymben.2016.08.005.

Koopman, F. *et al.* (2012) 'De novo production of the flavonoid naringenin in engineered Saccharomyces cerevisiae', *Microbial Cell Factories*, 11(155), pp. 1–15. Available at: https://pubmed.ncbi.nlm.nih.gov/23216753/.

Kos, A. *et al.* (1985) 'Isolation and characterization of the Aspergillus niger trpC gene', *Gene*, 39(2–3), pp. 231–238. doi: 10.1016/0378-1119(85)90317-8.

Kowalski, Z. *et al.* (2022) 'Second-generation biofuel production from the organic fraction of municipal solid waste', *Frontiers in Energy Research*, 10(August), pp. 1–15. doi: 10.3389/fenrg.2022.919415.

Kuivanen, J. *et al.* (2021) 'Engineering of Saccharomyces cerevisiae for anthranilate and methyl anthranilate production', *Microbial Cell Factories*. BioMed Central, 20(1), pp. 1–12. doi: 10.1186/s12934-021-01532-3.

Kumar, S. *et al.* (2018) 'MEGA X: Molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution*, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Lane, M. M. *et al.* (2011) 'Physiological and metabolic diversity in the yeast Kluyveromyces marxianus', *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 100(4), pp. 507–519. doi: 10.1007/s10482-011-9606-x.

Lane, M. M. and Morrissey, J. P. (2010) 'Kluyveromyces marxianus: A yeast emerging from its sister's shadow', *Fungal Biology Reviews*. Elsevier Ltd, 24(1–2), pp. 17–26. doi: 10.1016/j.fbr.2010.01.001.

Längle-rouault, F. and Jacobs, E. (1995) 'A method for performing precise alterations in the yeast genome using a recyclable selectable marker', *Nucleic Acids Research*, 23(15), pp. 3079–3081. doi: 10.1093/nar/23.15.3079.

Larroude, M. *et al.* (2019) 'A modular Golden Gate toolkit for Yarrowia lipolytica synthetic biology', *Microbial Biotechnology*, 12(6), pp. 1249–1259. doi: 10.1111/1751-7915.13427.

Larson, M. H. *et al.* (2013) 'CRISPR interference (CRISPRi) for sequence-specific control of gene expression', *Nature Protocols*. Nature Publishing Group, 8(11), pp. 2180–2196. doi: 10.1038/nprot.2013.132.

Last, R. L. *et al.* (1991) 'Tryptophan mutants in Arabidopsis: The consequences of duplicated tryptophan synthase β genes', *Plant Cell*, 3(4), pp. 345–358. doi: 10.1105/tpc.3.4.345.

Lee, M. E. *et al.* (2015) 'A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly', *ACS Synthetic Biology*, 4(9), pp. 975–986. doi: 10.1021/sb500366v.

Lee, M. Y., Hung, W. P. and Tsai, S. H. (2017) 'Improvement of shikimic acid production in Escherichia coli with growth phase-dependent regulation in the biosynthetic pathway from glycerol', *World Journal of Microbiology and Biotechnology*. Springer Netherlands, 33(2), pp. 1–8. doi: 10.1007/s11274-016-2192-3.

Lee, Y. J. *et al.* (2012) 'Involvement of GDH3-encoded NADP+-dependent glutamate dehydrogenase in yeast cell resistance to stress-induced apoptosis in stationary phase cells', *Journal of Biological Chemistry*, 287(53), pp. 44221–44233. doi: 10.1074/jbc.M112.375360.

Lehka, B. J. *et al.* (2017) 'Improving heterologous production of phenylpropanoids in Saccharomyces cerevisiae by tackling an unwanted side reaction of Tsc13, an endogenous double-bond reductase', *FEMS Yeast Research*, 17(1). doi: 10.1093/femsyr/fox004.

Lertwattanasakul, N. *et al.* (2015) 'Genetic basis of the highly efficient yeast Kluyveromyces marxianus: Complete genome sequence and transcriptome analyses', *Biotechnology for Biofuels*, 8(1), pp. 1–14. doi: 10.1186/s13068-015-0227-x.

Levisson, M. *et al.* (2018) 'Engineering de novo anthocyanin production in Saccharomyces cerevisiae', *Microbial Cell Factories*. BioMed Central, 17(1), pp. 1–16. doi: 10.1186/s12934-018-0951-6.

Li, G. *et al.* (2020) 'Enhanced Biosynthesis of Dihydromyricetin in Saccharomyces cerevisiae by Coexpression of Multiple Hydroxylases', *Journal of Agricultural and Food Chemistry*, 68(48), pp. 14221–14229. doi: 10.1021/acs.jafc.0c05261.

Li, J. *et al.* (1995) 'Isolation of cDNAs encoding the tryptophan pathway enzyme indole-3-glycerol phosphate synthase from Arabidopsis thaliana', *Plant Physiology*, 108(2), pp. 877–878. doi: 10.1104/pp.108.2.877.

Li, M. *et al.* (2015) 'De novo production of resveratrol from glucose or ethanol by engineered Saccharomyces cerevisiae', *Metabolic Engineering*. Elsevier, 32, pp. 1–11. doi: 10.1016/j.ymben.2015.08.007.

Li, M. *et al.* (2016) 'Engineering yeast for high-level production of stilbenoid antioxidants', *Scientific Reports*. Nature Publishing Group, 6, pp. 1–8. doi: 10.1038/srep36827.

Li, W., Huang, C. and Chen, J. (2022) 'The application of CRISPR /Cas mediated gene editing

in synthetic biology: Challenges and optimizations', *Frontiers in Bioengineering and Biotechnology*, 10(August), pp. 1–13. doi: 10.3389/fbioe.2022.890155.

Li, Y. *et al.* (2020) 'De Novo Biosynthesis of Caffeic Acid from Glucose by Engineered Saccharomyces cerevisiae', *ACS Synthetic Biology*, 9(4), pp. 756–765. doi: 10.1021/acssynbio.9b00431.

Lim, Y. P., Go, M. K. and Yew, W. S. (2016) 'Exploiting the biosynthetic potential of type III polyketide synthases', *Molecules*, 21(6), pp. 1–37. doi: 10.3390/molecules21060806.

Liu, H. *et al.* (2022) ' High-Level Production of Hydroxytyrosol in Engineered Saccharomyces cerevisiae ', *ACS Synthetic Biology*. doi: 10.1021/acssynbio.2c00316.

Liu, L. *et al.* (2019) 'Engineering the Biosynthesis of Caffeic Acid in Saccharomyces cerevisiae with Heterologous Enzyme Combinations', *Engineering*. Chinese Academy of Engineering, 5(2), pp. 287–295. doi: 10.1016/j.eng.2018.11.029.

Liu, M. *et al.* (2022) 'Remodelling metabolism for high-level resveratrol production in Yarrowia lipolytica', *Bioresource Technology*. Elsevier Ltd, 365(September), p. 128178. doi: 10.1016/j.biortech.2022.128178.

Liu, Q. *et al.* (2019) 'Rewiring carbon metabolism in yeast for high level production of aromatic chemicals', *Nature Communications*. Springer US, 10(1), pp. 1–13. doi: 10.1038/s41467-019-12961-5.

Liu, Q. *et al.* (2020) 'Current state of aromatics production using yeast: achievements and challenges', *Current Opinion in Biotechnology*, 65, pp. 65–74. doi: 10.1016/j.copbio.2020.01.008.

Liu, W., Zhang, B. and Jiang, R. (2017) 'Improving acetyl-CoA biosynthesis in Saccharomyces cerevisiae via the overexpression of pantothenate kinase and PDH bypass', *Biotechnology for Biofuels*. BioMed Central, 10(1), pp. 1–9. doi: 10.1186/s13068-017-0726-z.

Liu, Y. and Nielsen, J. (2019) 'Recent trends in metabolic engineering of microbial chemical factories', *Current Opinion in Biotechnology*. Elsevier Ltd, 60, pp. 188–197. doi: 10.1016/j.copbio.2019.05.010.

Liu, Z., Zhang, Y. and Nielsen, J. (2019) 'Synthetic Biology of Yeast', *Biochemistry*, 58(11), pp. 1511–1520. doi: 10.1021/acs.biochem.8b01236.

Löbs, A. K., Schwartz, C. and Wheeldon, I. (2017) 'Genome and metabolic engineering in non-conventional yeasts: Current advances and applications', *Synthetic and Systems Biotechnology*, 2(3), pp. 198–207. doi: 10.1016/j.synbio.2017.08.002.

Lowry, O. H. *et al.* (1951) 'Protein measurement with the Folin phenol reagent.', *The Journal of biological chemistry*. Â© 1951 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 193(1), pp. 265–275. doi: 10.1016/s0021-9258(19)52451-6.

Lütke-Eversloh, T. and Stephanopoulos, G. (2005) 'Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of Escherichia coli: Generation and characterization of tyrosine-insensitive mutants', *Applied and Environmental Microbiology*, 71(11), pp. 7224–7228. doi: 10.1128/AEM.71.11.7224-7228.2005.

Lv, Y. *et al.* (2019) 'Optimizing Oleaginous Yeast Cell Factories for Flavonoids and Hydroxylated Flavonoids Biosynthesis', *ACS Synthetic Biology*, 8(11), pp. 2514–2523. doi: 10.1021/acssynbio.9b00193.

Lynch, J. H. (2022) 'Revisiting the dual pathway hypothesis of Chorismate production in plants', *Horticulture Research*, 9(March). doi: 10.1093/hr/uhac052.

Lyu, X. *et al.* (2017) 'Enhancement of Naringenin Biosynthesis from Tyrosine by Metabolic Engineering of Saccharomyces cerevisiae', *Journal of Agricultural and Food Chemistry*, 65(31), pp. 6638–6646. doi: 10.1021/acs.jafc.7b02507.

MacDonald, M. C. *et al.* (2016) 'Rhodotorula glutinis phenylalanine/tyrosine ammonia lyase

enzyme catalyzed synthesis of the methyl ester of para-hydroxycinnamic acid and its potential antibacterial activity', *Frontiers in Microbiology*, 7(MAR), pp. 1–11. doi: 10.3389/fmicb.2016.00281.

Madzak, C. (2021) 'Yarrowia lipolytica strains and their biotechnological applications: How natural biodiversity and metabolic engineering could contribute to cell factories improvement', *Journal of Fungi*, 7(7). doi: 10.3390/jof7070548.

Maeda, H. *et al.* (2010) 'RNAi suppression of Arogenate Dehydratase1 reveals that phenylalanine is synthesized predominantly via the arogenate pathway in Petunia petals', *Plant Cell*, 22(3), pp. 832–849. doi: 10.1105/tpc.109.073247.

Maeda, H. and Dudareva, N. (2012) 'The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants'. doi: 10.1146/annurev-arplant-042811-105439.

Maeda, H., Yoo, H. and Dudareva, N. (2011) 'Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate', *Nature Chemical Biology*, 7(1), pp. 19–21. doi: 10.1038/nchembio.485.

Mahmood, H. *et al.* (2019) 'Recent advances in the pretreatment of lignocellulosic biomass for biofuels and value-added products', *Current Opinion in Green and Sustainable Chemistry*. Elsevier Ltd, 20, pp. 18–24. doi: 10.1016/j.cogsc.2019.08.001.

Malcı, K., Walls, L. E. and Rios-Solis, L. (2020) 'Multiplex Genome Engineering Methods for Yeast Cell Factory Development', *Frontiers in Bioengineering and Biotechnology*, 8(October), pp. 1–21. doi: 10.3389/fbioe.2020.589468.

Mankar, A. R. *et al.* (2021) 'Pretreatment of lignocellulosic biomass: A review on recent advances', *Bioresource Technology*. Elsevier Ltd, 334(March), p. 125235. doi: 10.1016/j.biortech.2021.125235.

Mannhaupt, G. *et al.* (1989) 'Characterization of the prephenate dehydrogenase-encoding gene, TYR1, from Saccharomyces cerevisiae', *Gene*, 85(2), pp. 303–311. doi: 10.1016/0378-1119(89)90422-8.

Mans, R. *et al.* (2015) 'CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae', *FEMS Yeast Research*, 15(2), pp. 1–15. doi: 10.1093/femsyr/fov004.

Mara, P. *et al.* (2018) 'The pleiotropic effects of the glutamate dehydrogenase (GDH) pathway in Saccharomyces cerevisiae', *Microbial Cell Factories*. BioMed Central, 17(1), pp. 1–13. doi: 10.1186/s12934-018-1018-4.

Marcet-Houben, M. and Gabaldón, T. (2015) 'Beyond the whole-genome duplication: Phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage', *PLoS Biology*, 13(8), pp. 1–26. doi: 10.1371/journal.pbio.1002220.

Van Maris, A. J. A. *et al.* (2003) 'Overproduction of threonine aldolase circumvents the biosynthetic role of pyruvate decarboxylase in glucose-limited chemostat cultures of Saccharomyces cerevisiae', *Applied and Environmental Microbiology*, 69(4), pp. 2094–2099. doi: 10.1128/AEM.69.4.2094-2099.2003.

Van Maris, A. J. A. *et al.* (2004) 'Directed Evolution of Pyruvate Decarboxylase-Negative Saccharomyces cerevisiae, Yielding a C2-Independent, Glucose-Tolerant, and Pyruvate-Hyperproducing Yeast', *Applied and Environmental Microbiology*, 70(1), pp. 159–166. doi: 10.1128/AEM.70.1.159-166.2004.

Markham, K. A. *et al.* (2018) 'Rewiring yarrowia lipolytica toward triacetic acid lactone for materials generation', *Proceedings of the National Academy of Sciences of the United States of America*, 115(9), pp. 2096–2101. doi: 10.1073/pnas.1721203115.

Matsumura, F. and O'Brien, R. D. (1963) 'A comparative study of the modes of action of fluoroacetamide and fluoroacetate in the mouse and american cockroach', *Biochemical Pharmacology*, 12(10), pp. 1201–1205. doi: 10.1016/0006-2952(63)90095-9.

McKenna, R. *et al.* (2014) 'Rational and combinatorial approaches toengineering styrene

production by Saccharomyces cerevisiae', *Microbial Cell Factories*, 13(1), pp. 1–12. doi: 10.1186/s12934-014-0123-2.

McTaggart, T. L. *et al.* (2019) 'Synthesis of polyketides from low cost substrates by the thermotolerant yeast Kluyveromyces marxianus', *Biotechnology and Bioengineering*, 116(7), pp. 1721–1730. doi: 10.1002/bit.26976.

Metcalf, R. L. (1987) 'Plant volatiles as insect attractants', *Critical Reviews in Plant Sciences*, 5(3), pp. 251–301. doi: 10.1080/07352688709382242.

Michel, G. *et al.* (2003) 'Structures of Shikimate Dehydrogenase AroE and Its Paralog YdiB', 278(21), pp. 19463–19472. doi: 10.1074/jbc.M300794200.

Milke, L. *et al.* (2018) 'Production of plant-derived polyphenols in microorganisms : current state and perspectives'. Applied Microbiology and Biotechnology, pp. 1575–1585.

Mitsui, R., Yamada, R. and Ogino, H. (2019) 'CRISPR system in the yeast Saccharomyces cerevisiae and its application in the bioproduction of useful chemicals.', *World journal of microbiology & biotechnology*, 35(7), p. 111. doi: 10.1007/s11274-019-2688-8.

Mobley, E. M., Kunkel, B. N. and Keith, B. (1999) 'Identification, characterization and comparative analysis of a novel chorismate mutase gene in Arabidopsis thaliana', *Gene*, 240(1), pp. 115–123. doi: 10.1016/S0378-1119(99)00423-0.

Monteiro Espíndola, K. M. *et al.* (2019) 'Chemical and pharmacological aspects of caffeic acid and its activity in hepatocarcinoma', *Frontiers in Oncology*, 9(JUN), pp. 3–5. doi: 10.3389/fonc.2019.00541.

Montini, N. *et al.* (2022) 'Identification of a novel gene required for competitive growth at high temperature in the thermotolerant yeast Kluyveromyces marxianus', *Microbiology (United Kingdom)*, 168(3), pp. 1–13. doi: 10.1099/mic.0.001148.

Moore, S. J. *et al.* (2016) 'EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology', *ACS Synthetic Biology*, 5(10), pp. 1059–1069. doi: 10.1021/acssynbio.6b00031.

Morrissey, J. P. *et al.* (2015) 'Cell factory applications of the yeast Kluyveromyces marxianus for the biotechnological production of natural flavour and fragrance molecules', *Yeast*, (December 2014), pp. 32: 3–16. doi: 10.1002/yea.3054.

Mousdale, D. M. and Coggins, J. R. (1985) 'Planta of the common shikimate-pathway enzymes in Pisum sativum L .', 163(2), pp. 241–249. Available at: https://pubmed.ncbi.nlm.nih.gov/24249346/.

Mullaney, E. J. *et al.* (1985) 'Primary structure of the trpC gene from Aspergillus nidulans divergent from the sequence of the functionally analogous', pp. 37–45. Available at: https://pubmed.ncbi.nlm.nih.gov/3158796/.

Mustafa, R. N. and Verpoorte, R. (2005) 'Chorismate derived C6C1 compounds in plants', *Planta*, 222(1), pp. 1–5. doi: 10.1007/S00425-005-1554-0.

Mutschler, H. *et al.* (2019) 'Special Issue on Bottom-Up Synthetic Biology', *ChemBioChem*, 20(20), pp. 2533–2534. doi: 10.1002/cbic.201900507.

Nambu-Nishida, Y. *et al.* (2017) 'Development of a comprehensive set of tools for genome engineering in a cold- A nd thermo-tolerant Kluyveromyces marxianus yeast strain', *Scientific Reports*. Springer US, 7(1), pp. 1–7. doi: 10.1038/s41598-017-08356-5.

Neelam, Khatkar, A. and Sharma, K. K. (2020) 'Phenylpropanoids and its derivatives: biological activities and its role in food, pharmaceutical and cosmetic industries', *Critical Reviews in Food Science and Nutrition*. Taylor & Francis, 60(16), pp. 2655–2675. doi: 10.1080/10408398.2019.1653822.

Nichols, B. P. *et al.* (1980) 'Nucleotide sequences of the trpG regions of Escherichia coli, Shigella dysenteriae, Salmonella typhimurium and Serratia marcescens', *Journal of Molecular Biology*, 142(4), pp. 503–517. doi: 10.1016/0022-2836(80)90260-0.

Nielsen, J. and Keasling, J. D. (2016) 'Engineering Cellular Metabolism', *Cell*. Elsevier Ltd,

164(6), pp. 1185–1197. doi: 10.1016/j.cell.2016.02.004.

Niyogi, K. K. *et al.* (1993) 'Suppressors of trp1 fluorescence identify a new arabidopsis gene, TRP4, encoding the anthranilate synthase β subunit', *Plant Cell*, 5(9), pp. 1011–1027. doi: 10.1105/tpc.5.9.1011.

Niyogi, K. K. and Fink, G. R. (1992) 'Two anthranilate synthase genes in arabidopsis: Defense-related regulation of the tryptophan pathway', *Plant Cell*, 4(6), pp. 721–733. doi: 10.1105/tpc.4.6.721.

Noel, J. P., Austin, M. B. and Bomati, E. K. (2005) 'Structure-function relationships in plant phenylpropanoid biosynthesis', *Current Opinion in Plant Biology*, 8(3 SPEC. ISS.), pp. 249–253. doi: 10.1016/j.pbi.2005.03.013.

Nosaka, K. (2006) 'Recent progress in understanding thiamin biosynthesis and its genetic regulation in Saccharomyces cerevisiae', *Applied Microbiology and Biotechnology*, 72(1), pp. 30–40. doi: 10.1007/s00253-006-0464-9.

Olzhausen, J. *et al.* (2021) 'Increased biosynthesis of acetyl-CoA in the yeast Saccharomyces cerevisiae by overexpression of a deregulated pantothenate kinase gene and engineering of the coenzyme A biosynthetic pathway', *Applied Microbiology and Biotechnology*. Springer Berlin Heidelberg, 105(19), pp. 7321–7337. doi: 10.1007/s00253-021-11523-4.

Ooyen, A. J. J. Van *et al.* (2006) 'Heterologous protein production in the yeast Kluyveromyces lactis', 6, pp. 381–392. doi: 10.1111/j.1567-1364.2006.00049.x.

Oud, B. *et al.* (2012) 'An internal deletion in MTH1 enables growth on glucose of pyruvate-decarboxylase negative, non-fermentative Saccharomyces cerevisiae', pp. 1–10. Available at: https://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-11-131.

Ouedraogo, J. P. and Tsang, A. (2020) 'CRISPR_Cas systems for fungal research', *Fungal Biology Reviews*. Elsevier Ltd, 34(4), pp. 189–201. doi: 10.1016/j.fbr.2020.10.002.

Ouyang, J., Shao, X. and Li, J. (2000) 'Indole-3-glycerol phosphate, a branchpoint of indole-3-acetic acid biosynthesis from the tryptophan biosynthetic pathway in Arabidopsis thaliana', *Plant Journal*, 24(3), pp. 327–334. doi: 10.1046/j.1365-313X.2000.00883.x.

Ozcan, S. and Johnston, M. (1995) 'Three different regulatory mechanisms enable yeast hexose transporter (HXT) genes to be induced by different levels of glucose', *Molecular and Cellular Biology*, 15(3), pp. 1564–1572. doi: 10.1128/mcb.15.3.1564.

de Paiva, L. C. *et al.* (2019) 'Genomic analysis and lactose transporter expression in Kluyveromyces marxianus CCT 7735', *Fungal Biology*, 123(9), pp. 687–697. doi: 10.1016/j.funbio.2019.06.004.

Palmer, C. M. *et al.* (2020) 'Engineering 4-coumaroyl-CoA derived polyketide production in Yarrowia lipolytica through a β-oxidation mediated strategy', *Metabolic Engineering*. Elsevier Inc., 57(November 2019), pp. 174–181. doi: 10.1016/j.ymben.2019.11.006.

Panina, E. . *et al.* (1999) 'Regulation of Aromatic Amino Acid Biosynthesis in Gamma-Proteobacteria', *Zentralbl. Bakteriol. `Orig. A]..*, 220, pp. 298–300. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC103779/.

Paravicini, G., Schmidhetni, T. and Braus, G. (1989) 'Purification and properties of the 3-deoxy- ~ -arabino-heptulosonate-7-phosphate synthase (phen ylalanine-inhibitable) of Saccharomyces cerevisiae', 366, pp. 361–366.

Passi, A. *et al.* (2022) 'Genome-scale metabolic modeling enables in-depth understanding of big data', *Metabolites*, 12(1). doi: 10.3390/metabo12010014.

Patnaik, R. and Liao, J. C. (1994) 'Engineering of Escherichia coli central metabolism for aromatic metabolite production with near theoretical yield', *Applied and Environmental Microbiology*, 60(11), pp. 3903–3908. doi: 10.1128/aem.60.11.3903-3908.1994.

Patra, P. *et al.* (2021) 'Recent advances in systems and synthetic biology approaches for developing novel cell-factories in non-conventional yeasts', *Biotechnology Advances*. Elsevier

Inc., 47(September 2020), p. 107695. doi: 10.1016/j.biotechadv.2021.107695.

Paul, B. and Montoya, G. (2020) 'CRISPR-Cas12a: Functional overview and applications', *Biomedical Journal*. Elsevier Ltd, 43(1), pp. 8–17. doi: 10.1016/j.bj.2019.10.005.

Paulsen, B. S. *et al.* (2017) 'Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR-Cas9 genome editing', *Nature Biomedical Engineering*. Springer US, 1(11), pp. 878–888. doi: 10.1038/s41551-017-0145-2.

Peralta-Yahya, P. P. *et al.* (2012) 'Microbial engineering for the production of advanced biofuels', *Nature*, 488(7411), pp. 320–328. doi: 10.1038/nature11478.

Piva, L. C. *et al.* (2018) 'Acetamidase as a dominant recyclable marker for Komagataella phaffii strain engineering', *Applied Microbiology and Biotechnology*. Applied Microbiology and Biotechnology, 102(6), pp. 2753–2761. doi: 10.1007/s00253-018-8824-9.

Pohnert, G. *et al.* (1999) 'Regulation of phenylalanine biosynthesis. Studies on the mechanism of phenylalanine binding and feedback inhibition in the Escherichia coli P- protein', *Biochemistry*, 38(38), pp. 12212–12217. doi: 10.1021/bi991134w.

Ponnusamy, V. K. *et al.* (2019) 'A review on lignin structure, pretreatments, fermentation reactions and biorefinery potential', *Bioresource Technology*. Elsevier, 271(September 2018), pp. 462–472. doi: 10.1016/j.biortech.2018.09.070.

Postma, E. D. *et al.* (2021) 'A supernumerary designer chromosome for modular in vivo pathway assembly in Saccharomyces cerevisiae'. Oxford University Press, 49(3), pp. 1769–1783. doi: 10.1093/nar/gkaa1167.

Poulsen, C., Verpoorte, R. and Bongaerts, R. (1993) 'Purification and characterization of anthranilate synthase from Catharanthus roseus', 440, pp. 431–440.

Pronk, J. T., Steensma, H. Y. and Van Dijken, J. P. (1996) 'Pyruvate metabolism in Saccharomyces cerevisiae', *Yeast*, 12(16), pp. 1607–1633. doi: 10.1002/(SICI)1097-0061(199612)12:16<1607::AID-YEA70>3.0.CO;2-4.

Puchta, H., Dujon, B. and Hohn, B. (1993) 'Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease', *Nucleic Acids Research*, 21(22), pp. 5034–5040. doi: 10.1093/nar/21.22.5034.

Quan, J. and Tian, J. (2009) 'Circular polymerase extension cloning of complex gene libraries and pathways', *PLoS ONE*, 4(7). doi: 10.1371/journal.pone.0006441.

Radwanski, E. R., Zhao, J. and Last, R. L. (1995) 'Arabidopsis thaliana tryptophan synthase alpha: Gene cloning, expression, and subunit interaction', *Mgg Molecular & General Genetics*, 248(6), pp. 657–667. doi: 10.1007/BF02191705.

Rainha, J. *et al.* (2022) 'Curcumin biosynthesis from ferulic acid by engineered Saccharomyces cerevisiae', *Biotechnology Journal*, 17(3), pp. 1–8. doi: 10.1002/biot.202100400.

Rajkumar, A. S. *et al.* (2019) 'Biological Parts for Kluyveromyces marxianus Synthetic Biology', *Frontiers in Bioengineering and Biotechnology*, 7. doi: 10.3389/fbioe.2019.00097.

Rajkumar, A. S. and Morrissey, J. P. (2020) 'Rational engineering of Kluyveromyces marxianus to create a chassis for the production of aromatic products', *Microbial Cell Factories*. BioMed Central, 19(1), pp. 1–19. doi: 10.1186/s12934-020-01461-7.

Rajkumar, A. S. and Morrissey, J. P. (2022) 'Protocols for marker-free gene knock-out and knock-down in Kluyveromyces marxianus using CRISPR/Cas9', *FEMS Yeast Research*, 22(1), pp. 1–16. doi: 10.1093/femsyr/foab067.

Ramani, S., Patil, N. and Jayabaskaran, C. (2010) 'UV-B induced transcript accumulation of DAHP synthase in suspension-cultured Catharanthus roseus cells', *Journal of Molecular Signaling*, 5, pp. 1–9. doi: 10.1186/1750-2187-5-13.

Read, J. D. *et al.* (2007) 'Acetamide selection of Kluyveromyces lactis cells transformed with

an integrative vector leads to high-frequency formation of multicopy strains', *Applied and Environmental Microbiology*, 73(16), pp. 5088–5096. doi: 10.1128/AEM.02253-06.

Resnick, M. A. and Martin, P. (1976) 'The repair of double-strand breaks in the nuclear DNA of Saccharomyces cerevisiae and its genetic control', *MGG Molecular & General Genetics*, 143(2), pp. 119–129. doi: 10.1007/BF00266917.

Ribeiro, O. *et al.* (2007) 'Application of the Cre-loxP system for multiple gene disruption in the yeast Kluyveromyces marxianus', *Journal of Biotechnology*, 131(1), pp. 20–26. doi: 10.1016/j.jbiotec.2007.05.027.

Richardson, S. M. *et al.* (2017) 'Design of a synthetic yeast genome', 1044(March), pp. 1040–1044.

Rienzo, M. *et al.* (2021) 'High-throughput screening for high-efficiency small-molecule biosynthesis', *Metabolic Engineering*. Elsevier Inc., 63(September 2020), pp. 102–125. doi: 10.1016/j.ymben.2020.09.004.

Robak, K. and Balcerek, M. (2018) 'Review of second generation bioethanol production from residual biomass', *Food Technology and Biotechnology*, 56(2), pp. 174–187. doi: 10.17113/ftb.56.02.18.5428.

Rodriguez, A. *et al.* (2015) 'Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis', *Metabolic Engineering*. Elsevier, 31, pp. 181?188-188. doi: 10.1016/j.ymben.2015.08.003.

Rodriguez, A., Chen, Y., *et al.* (2017) 'Comparison of the metabolic response to over-production of p-coumaric acid in two yeast strains', *Metabolic Engineering*. Elsevier Inc., 44(October), pp. 265–272. doi: 10.1016/j.ymben.2017.10.013.

Rodriguez, A., Strucko, T., *et al.* (2017) 'Metabolic engineering of yeast for fermentative production of flavonoids', *Bioresource Technology*. The Author(s), 245, pp. 1645–1654. doi: 10.1016/j.biortech.2017.06.043.

Van Roermund, C. W. T. *et al.* (1995) 'The membrane of peroxisomes in Saccharomyces cerevisiae is impermeable to NAD(H) and acetyl-CoA under in vivo conditions', *EMBO Journal*, 14(14), pp. 3480–3486. doi: 10.1002/j.1460-2075.1995.tb07354.x.

Romagnoli, G. *et al.* (2012) 'Substrate specificity of thiamine pyrophosphate-dependent 2-oxo-acid decarboxylases in Saccharomyces cerevisiae', *Applied and Environmental Microbiology*, 78(21), pp. 7538–7548. doi: 10.1128/AEM.01675-12.

Romagnoli, G. (2015) 'Deletion of the Saccharomyces cerevisiae ARO8 gene, encoding an aromatic amino acid transaminase, enhances phenylethanol production from glucose', *Yeast*, (May 2014), pp. 191–198. doi: 10.1002/yea.

Rosebrock, A. P. (2017) 'Synchronization and arrest of the budding yeast cell cycle using chemical and genetic methods', *Cold Spring Harbor Protocols*, 2017(1), pp. 47–52. doi: 10.1101/pdb.prot088724.

Rouet, P., Smih, F. and Jasin, M. (1994) 'Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells', *Proceedings of the National Academy of Sciences of the United States of America*, 91(13), pp. 6064–6068. doi: 10.1073/pnas.91.13.6064.

Royer, J. C. *et al.* (1999) 'Deletion of the Trichodiene synthase gene of Fusarium venenatum: Two systems for repeated gene deletions', *Fungal Genetics and Biology*, 28(1), pp. 68–78. doi: 10.1006/fgbi.1999.1162.

Rubin, J. L. and Jensen, R. A. (1979) ' Enzymology of l-Tyrosine Biosynthesis in Mung Bean (Vigna radiata [L.] Wilczek) ', *Plant Physiology*, 64(5), pp. 727–734. doi: 10.1104/pp.64.5.727.

Sáez-Sáez, J. *et al.* (2020) 'Engineering the oleaginous yeast Yarrowia lipolytica for high-level resveratrol production', *Metabolic Engineering*, 62(August), pp. 51–61. doi: 10.1016/j.ymben.2020.08.009.

Sakihama, Y. *et al.* (2019) 'Increased flux in acetyl-CoA synthetic pathway and TCA cycle of Kluyveromyces marxianus under respiratory conditions', *Scientific Reports*, 9(1), pp. 1–7. doi: 10.1038/s41598-019-41863-1.

Sarrion-Perdigones, A. *et al.* (2013) 'Goldenbraid 2.0: A comprehensive DNA assembly framework for plant synthetic biology', *Plant Physiology*, 162(3), pp. 1618–1631. doi: 10.1104/pp.113.217661.

Saunders, L. P. *et al.* (2015) 'Triacetic acid lactone production in industrial Saccharomyces yeast strains', *Journal of Industrial Microbiology and Biotechnology*, 42(5), pp. 711–721. doi: 10.1007/s10295-015-1596-7.

Schmid, J. *et al.* (1990) 'The in-vitro synthesized tomato shikimate kinase precursor is enzymatically active and is imported and processed to the mature enzyme by chloroplasts', 2(1 992), pp. 375–383. Available at: https://pubmed.ncbi.nlm.nih.gov/1338949/.

Schmidt, C. L. *et al.* (1990) 'Shikimate Kinase from Spinach Chloroplasts', *Plant Physiology*, 93(2), pp. 758–766. Available at: https://pubmed.ncbi.nlm.nih.gov/16667533/.

Schmidt, S. *et al.* (2011) 'Identification of a Saccharomyces cerevisiae glucosidase that hydrolyzes flavonoid glucosides', *Applied and Environmental Microbiology*, 77(5), pp. 1751–1757. doi: 10.1128/AEM.01125-10.

Schnappauf, G. *et al.* (1998) 'The two 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase isoenzymes from Saccharomyces cerevisiae show different kinetic modes of inhibition', *Archives of Microbiology*, 169(6), pp. 517–524. doi: 10.1007/s002030050605.

Schuster, A. *et al.* (2012) 'A versatile toolkit for high throughput functional genomics with Trichoderma reesei', *Biotechnology for Biofuels*, 5, pp. 1–10. doi: 10.1186/1754-6834-5-1.

Seeboth, P. G., Bohnsack, K. and Hollenberg, C. P. (1990) 'pdc1(0) mutants of Saccharomyces cerevisiae give evidence for an additional structural PDC gene: cloning of PDC5, a gene homologous to PDC1.', *Journal of bacteriology*, 172(2), pp. 678–685. doi: 10.1128/jb.172.2.678-685.1990.

Shao, S. *et al.* (2017) 'Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing Saccharomyces cerevisiae Rad52', *International Journal of Biochemistry and Cell Biology*. Elsevier, 92(January), pp. 43–52. doi: 10.1016/j.biocel.2017.09.012.

Sharma, H. K., Xu, C. and Qin, W. (2019) 'Biological Pretreatment of Lignocellulosic Biomass for Biofuels and Bioproducts: An Overview', *Waste and Biomass Valorization*. Springer Netherlands, 10(2), pp. 235–251. doi: 10.1007/s12649-017-0059-y.

Sheng, H. *et al.* (2020) 'Metabolic Engineering of Microorganisms for the Production of Flavonoids', *Frontiers in Bioengineering and Biotechnology*, 8(October), pp. 1–15. doi: 10.3389/fbioe.2020.589069.

Shi, S. *et al.* (2016) 'A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in Saccharomyces cerevisiae', *Metabolic Engineering*. Elsevier, 33, pp. 19–27. doi: 10.1016/j.ymben.2015.10.011.

Shin, S. Y. *et al.* (2011) 'Production of resveratrol from p-coumaric acid in recombinant Saccharomyces cerevisiae expressing 4-coumarate:coenzyme A ligase and stilbene synthase genes', *Enzyme and Microbial Technology*. Elsevier Inc., 48(1), pp. 48–53. doi: 10.1016/j.enzmictec.2010.09.004.

Siewers, V. (2014) 'An overview on selection marker genes for transformation of Saccharomyces cerevisiae.', *Methods in molecular biology (Clifton, N.J.)*, 1152, pp. 3–15. doi: 10.1007/978-1-4939-0563-8_1.

Singh, A. and Sherman, F. (1974) 'Association of methionine requirement with methyl mercury resistant mutants of yeast', *Nature*, 244, pp. 227–229. Available at: https://pubmed.ncbi.nlm.nih.gov/4589576/.

Singh, S. A. and Christendat, D. (2006) 'Structure of Arabidopsis Dehydroquinate

Dehydratase-Shikimate Dehydrogenase and Implications for Metabolic Channeling in the Shikimate Pathway', pp. 7787–7796.

Singh, S., Lewis, N. G. and Towers, G. H. N. (1998) 'Nitrogen recycling during phenylpropanoid metabolism in sweet potato tubers', *Journal of Plant Physiology*, 153(3–4), pp. 316–323. doi: 10.1016/S0176-1617(98)80157-0.

Smart, C. C. and Amrhein, N. (1987) 'Planta Ultrastructural localisation by protein A-gold immunocytochemistry', pp. 1–6. Available at: https://pubmed.ncbi.nlm.nih.gov/24232834/.

Solis-Escalante, D. *et al.* (2013) 'amdSYM, A new dominant recyclable marker cassette for Saccharomyces cerevisiae', *FEMS Yeast Research*, 13(1), pp. 126–139. doi: 10.1111/1567-1364.12024.

Solis-Escalante, D. *et al.* (2015) 'The genome sequence of the popular hexose-transport-deficient Saccharomyces cerevisiae strain EBY.VW4000 reveals LoxP/Cre-induced translocations and gene loss', *FEMS yeast research*, 15(2), pp. 1–12. doi: 10.1093/femsyr/fou004.

Sternberg, S. H. *et al.* (2014) 'DNA interrogation by the CRISPR RNA-guided endonuclease Cas9', *Nature*. Nature Publishing Group, 507(7490), pp. 62–67. doi: 10.1038/nature13011.

Stompor-Gorący, M. and Machaczka, M. (2021) 'Recent advances in biological activity, new formulations and prodrugs of ferulic acid', *International Journal of Molecular Sciences*, 22(23). doi: 10.3390/ijms222312889.

Storch, M. *et al.* (2015) 'BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology', pp. 4–10. doi: 10.1021/sb500356d.

Suástegui, M. *et al.* (2016) 'Investigating strain dependency in the production of aromatic compounds in Saccharomyces cerevisiae', *Biotechnology and Bioengineering*, 113(12), pp. 2676–2685. doi: 10.1002/bit.26037.

Suástegui, M. *et al.* (2017) 'Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in Saccharomyces cerevisiae for high production of polymer and drug precursors', *Metabolic Engineering*. Elsevier Inc., 42(June), pp. 134–144. doi: 10.1016/j.ymben.2017.06.008.

Suzuki, N., Sakuta, M. and Shimizu, S. (1996) 'Purification and Characterization of a Cytosolic Isozyme of from Cultured Carrot Cells', *Journal of Plant Physiology*. Gustav Fischer Verlag, Stuttgart, 149(1–2), pp. 19–22. doi: 10.1016/S0176-1617(96)80167-2.

Swarts, D. C. and Jinek, M. (2018) 'Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing', *Wiley Interdisciplinary Reviews: RNA*, 9(5), pp. 1–19. doi: 10.1002/wrna.1481.

Sydor, T., Schaffer, S. and Boles, E. (2010) 'Considerable Increase in Resveratrol Production by Recombinant Industrial Yeast Strains with Use of Rich Medium', *Applied and Environmental Microbiology*, 76(10), pp. 3361–3363. doi: 10.1128/AEM.02796-09.

Tang, Y., Sieg, A. and Trotter, P. J. (2011) '13C-metabolic enrichment of glutamate in glutamate dehydrogenase mutants of Saccharomyces cerevisiae', *Microbiological Research*. Elsevier GmbH., 166(7), pp. 521–530. doi: 10.1016/j.micres.2010.10.004.

Tarasov, D., Leitch, M. and Fatehi, P. (2018) 'Lignin-carbohydrate complexes: Properties, applications, analyses, and methods of extraction: A review', *Biotechnology for Biofuels*. BioMed Central, 11(1), pp. 1–28. doi: 10.1186/s13068-018-1262-1.

Teshiba, S. *et al.* (1986) 'Cloning of the ARO3 gene of Saccharomyces cerevisiae and its regulation', *MGG Molecular & General Genetics*, 205(2), pp. 353–357. doi: 10.1007/BF00430450.

Tohge, T. *et al.* (2013) 'Shikimate and phenylalanine biosynthesis in the green lineage', *Frontiers in Plant Science*, 4(MAR), pp. 1–13. doi: 10.3389/fpls.2013.00062.

Tozawa, Y. *et al.* (2001) 'Characterization of rice anthranilate synthase α-subunit genes OASA1 and OASA2. Tryptophan accumulation in transgenic rice expressing a feedback-insensitive mutant of OASA1', *Plant Physiology*, 126(4), pp. 1493–1506. doi: 10.1104/pp.126.4.1493.

Tran, N. T. *et al.* (2019) 'Enhancement of Precise Gene Editing by the Association of Cas9 With Homologous Recombination Factors', *Frontiers in Genetics*, 10(APR), pp. 1–13. doi: 10.3389/fgene.2019.00365.

Trenchard, I. J. and Smolke, C. D. (2015) 'Engineering strategies for the fermentative production of plant alkaloids in yeast', *Metabolic Engineering*. Elsevier, 30, pp. 96–104. doi: 10.1016/j.ymben.2015.05.001.

Tribe, D. E., Camakaris, H. and Pittard, J. (1976) 'Constitutive and repressible enzymes of the common pathway of aromatic biosynthesis in Escherichia coli K-12: Regulation of enzyme synthesis at different growth rates', *Journal of Bacteriology*, 127(3), pp. 1085–1097. doi: 10.1128/jb.127.3.1085-1097.1976.

Trubitsyna, M. *et al.* (2014) 'PaperClip : rapid multi-part DNA assembly from existing libraries', 42(20), pp. 0–7. doi: 10.1093/nar/gku829.

Tutino, M. L. *et al.* (1993) 'Tryptophan biosynthesis genes trpEGC in the thermoacidophilic archaebacterium Sulfolobus solfataricus', *Journal of Bacteriology*, 175(1), pp. 299–302. doi: 10.1128/jb.175.1.299-302.1993.

Tzin, V. *et al.* (2012) 'Expression of a bacterial feedback-insensitive 3-deoxy- D -arabino-heptulosonate 7-phosphate synthase of the shikimate pathway in Arabidopsis elucidates potential metabolic bottlenecks between primary and secondary metabolism', pp. 430–439. Available at: https://pubmed.ncbi.nlm.nih.gov/22296303/.

Tzin, V. *et al.* (2013) 'Tomato fruits expressing a bacterial feedback-insensitive 3-deoxy- d -arabino-heptulosonate 7-phosphate synthase of the shikimate pathway possess enhanced levels of multiple specialized metabolites and upgraded aroma', 64(14), pp. 4441–4452. doi: 10.1093/jxb/ert250.

Tzin, V. and Galili, G. (2010) ' The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in Arabidopsis thaliana ', *The Arabidopsis Book*, 8(i), p. e0132. doi: 10.1199/tab.0132.

Urlacher, V. B. and Girhard, M. (2019) 'Cytochrome P450 Monooxygenases in Biotechnology and Synthetic Biology', *Trends in Biotechnology*. Elsevier Ltd, 37(8), pp. 882–897. doi: 10.1016/j.tibtech.2019.01.001.

Urnov, F. D. *et al.* (2010) 'Genome editing with engineered zinc finger nucleases', *Nature Reviews Genetics*. Nature Publishing Group, 11(9), pp. 636–646. doi: 10.1038/nrg2842.

Urrestarazu, A. *et al.* (1998) 'Phenylalanine- and tyrosine-auxotrophic mutants of Saccharomyces cerevisiae impaired in transamination', *Molecular and General Genetics*, 257(2), pp. 230–237. doi: 10.1007/s004380050643.

Varela, J. A. *et al.* (2017) 'Polymorphisms in the LAC12 gene explain lactose utilisation variability in Kluyveromyces marxianus strains', *FEMS Yeast Research*, 17(3), pp. 1–13. doi: 10.1093/femsyr/fox021.

Varela, J. A. *et al.* (2019) 'Expansion and diversification of MFS transporters in Kluyveromyces marxianus', *Frontiers in Microbiology*, 10(JAN), pp. 1–15. doi: 10.3389/fmicb.2018.03330.

Velvizhi, G. *et al.* (2022) 'Valorisation of lignocellulosic biomass to value-added products: Paving the pathway towards low-carbon footprint', *Fuel*. Elsevier Ltd, 313(November 2021), p. 122678. doi: 10.1016/j.fuel.2021.122678.

Verduyn, C. *et al.* (1992) 'Effect of Benzoic Acid on Metabolic Fluxes in Yeasts':, *Yeast*, 8(1992), pp. 501–517. Available at: https://onlinelibrary.wiley.com/doi/10.1002/yea.320080703.

Verhe, R. *et al.* (2022) 'Production of Bio-Ethanol from the Organic Fraction of Municipal Solid Waste and Refuse-Derived Fuel', *Biomass*, 2(4), pp. 224–236. doi: 10.3390/biomass2040015.

Vickery, C. R. *et al.* (2018) 'A coupled in vitro/in vivo approach for engineering a heterologous type III PKS to enhance polyketide biosynthesis in Saccharomyces cerevisiae', *Biotechnology and Bioengineering*, 115(6), pp. 1394–1402. doi: 10.1002/bit.26564.

Vuralhan, Z. *et al.* (2005) 'Physiological characterization of the ARO10-dependent, broad-substrate- specificity 2-oxo acid decarboxylase activity of Saccharomyces cerevisiae', *Applied and Environmental Microbiology*, 71(6), pp. 3276–3284. doi: 10.1128/AEM.71.6.3276-3284.2005.

Wang, L. *et al.* (2018) 'Efficient CRISPR-Cas9 mediated multiplex genome editing in yeasts', *Biotechnology for Biofuels*. BioMed Central, 11(1), pp. 1–16. doi: 10.1186/s13068-018-1271-0.

Wang, M. and Maeda, H. A. (2018) 'Aromatic amino acid aminotransferases in plants', *Phytochemistry Reviews*. Springer Netherlands, 17(1), pp. 131–159. doi: 10.1007/s11101-017-9520-6.

Wang, Y. *et al.* (2011) 'Stepwise increase of resveratrol biosynthesis in yeast Saccharomyces cerevisiae by metabolic engineering', *Metabolic Engineering*. Elsevier, 13(5), pp. 455–463. doi: 10.1016/j.ymben.2011.04.005.

Wang, Y., Chen, H. and Yu, O. (2014) 'A plant malonyl-CoA synthetase enhances lipid content and polyketide yield in yeast cells', *Applied Microbiology and Biotechnology*, 98(12), pp. 5435–5447. doi: 10.1007/s00253-014-5612-z.

Watts, K. T., Lee, P. C. and Schmidt-Dannert, C. (2004) 'Exploring recombinant flavonoid biosynthesis in metabolically engineered Escherichia coli', *ChemBioChem*, 5(4), pp. 500–507. doi: 10.1002/cbic.200300783.

Webby, C. J. *et al.* (2010) 'Synergistic Allostery , a Sophisticated Regulatory Network for the Control of Aromatic Amino Acid Biosynthesis in Mycobacterium tuberculosis * □', 285(40), pp. 30567–30576. doi: 10.1074/jbc.M110.111856.

Weber, A. P. M., Schwacke, R. and Flügge, U. I. (2005) 'Solute transporters of the plastid envelope membrane', *Annual Review of Plant Biology*, 56, pp. 133–164. doi: 10.1146/annurev.arplant.56.032604.144228.

Weber, E. *et al.* (2011) 'A modular cloning system for standardized assembly of multigene constructs', *PLoS ONE*, 6(2). doi: 10.1371/journal.pone.0016765.

Wei, T., Cheng, B. Y. and Liu, J. Z. (2016) 'Genome engineering Escherichia coli for L-DOPA overproduction from glucose', *Scientific Reports*. Nature Publishing Group, 6(June), pp. 1–9. doi: 10.1038/srep30080.

Westfall, C. S., Xu, A. and Jez, J. M. (2014) 'Structural evolution of differential amino acid effector regulation in plant chorismate mutases', *Journal of Biological Chemistry*, 289(41), pp. 28619–28628. doi: 10.1074/jbc.M114.591123.

Weusthuis, R. A. *et al.* (1994) 'Effects of oxygen limitation on sugar metabolism in yeasts: A continuous-culture study of the Kluyver effect', *Microbiology*, 140(4), pp. 703–715. doi: 10.1099/00221287-140-4-703.

Woodley, J. M. (2017) 'Bioprocess intensification for the effective production of chemical products', *Computers and Chemical Engineering*. Elsevier Ltd, 105, pp. 297–307. doi: 10.1016/j.compchemeng.2017.01.015.

Wu, Z. *et al.* (2020) 'Amidase as a versatile tool in amide-bond cleavage : From molecular features to biotechnological applications', *Biotechnology Advances*. Elsevier, 43(January), p. 107574. doi: 10.1016/j.biotechadv.2020.107574.

Xie, G. and Peng, L. (2011) 'Genetic Engineering of Energy Crops: A Strategy for Biofuel Production in China', *Journal of Integrative Plant Biology*, 53(2), pp. 143–150. doi: 10.1111/j.1744-7909.2010.01022.x.

Xie, S. *et al.* (2014) 'SgRNAcas9: A software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites', *PLoS ONE*, 9(6), pp. 1–9. doi:

10.1371/journal.pone.0100448.

Xu, S. *et al.* (2022) 'Efficient production of anthocyanins in Saccharomyces cerevisiae by introducing anthocyanin transporter and knocking out endogenous degrading enzymes', *Frontiers in Bioengineering and Biotechnology*, 10(August), pp. 1–13. doi: 10.3389/fbioe.2022.899182.

Yamada, T. *et al.* (2008) 'Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of phenylalanine and tryptophan', *Plant Cell*, 20(5), pp. 1316–1329. doi: 10.1105/tpc.107.057455.

Yamashiro, C. T., Yarden, O. and Yanofsky, C. (1992) 'A dominant selectable marker that is meiotically stable in Neurospora crassa: the amdS gene of Aspergillus nidulans', *MGG Molecular & General Genetics*, 236(1), pp. 121–124. doi: 10.1007/BF00279650.

Yin, R. *et al.* (2010) 'Plants contain two distinct classes of functional tryptophan synthase beta proteins', *Phytochemistry*. Elsevier Ltd, 71(14–15), pp. 1667–1672. doi: 10.1016/j.phytochem.2010.07.006.

Yin, R. *et al.* (2012) 'Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-O-glycosylation', *Journal of Experimental Botany*, 63(7), pp. 2465–2478. doi: 10.1093/jxb/err416.

Yokoyama, R. *et al.* (2021) 'The entry reaction of the plant shikimate pathway is subjected to highly complex metabolite-mediated regulation', pp. 671–696. doi: 10.1093/plcell/koaa042.

Yoo, H. *et al.* (2013) 'An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase', *Nature Communications*. Nature Publishing Group, 4. doi: 10.1038/ncomms3833.

Yu, S. in *et al.* (2019) 'Post-translational and transcriptional regulation of phenylpropanoid biosynthesis pathway by Kelch repeat F-box protein SAGL1', *Plant Molecular Biology*, 99(1–2), pp. 135–148. doi: 10.1007/s11103-018-0808-8.

Yuan, J. *et al.* (2022) 'Shikimate Kinase Plays Important Roles in Anthocyanin Synthesis in Petunia'. Available at: https://pubmed.ncbi.nlm.nih.gov/36555606/.

Zelle, R. M. *et al.* (2010) 'Phosphoenolpyruvate carboxykinase as the sole anaplerotic enzyme in saccharomyces cerevisiae', *Applied and Environmental Microbiology*, 76(16), pp. 5383–5389. doi: 10.1128/AEM.01077-10.

Zhang, Q. *et al.* (2021) 'Systematically Engineered Fatty Acid Catabolite Pathway for the Production of (2 S)-Naringenin in Saccharomyces cerevisiae', *ACS Synthetic Biology*, 10(5), pp. 1166–1175. doi: 10.1021/acssynbio.1c00002.

Zhang, R. *et al.* (2008) 'Arabidopsis indole synthase, a homolog of tryptophan synthase alpha, is an enzyme involved in the Trp-independent indole-containing metabolite biosynthesis', *Journal of Integrative Plant Biology*, 50(9), pp. 1070–1077. doi: 10.1111/j.1744-7909.2008.00729.x.

Zhang, Yueping *et al.* (2019) 'A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in Saccharomyces cerevisiae', *Nature Communications*. doi: 10.1038/s41467-019-09005-3.

Zhang, Z., Moo-Young, M. and Chisti, Y. (1996) 'Plasmid stability in recombinant Saccharomyces cerevisiae.', *Biotechnology advances*, 14(4), pp. 401–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14540156 (Accessed: 10 August 2019).

Zhu, L. *et al.* (2013) 'Cloning, expression and characterization of phenylalanine ammonia-lyase from Rhodotorula glutinis', *Biotechnology Letters*, 35(5), pp. 751–756. doi: 10.1007/s10529-013-1140-7.

Zhu, L. *et al.* (2021) 'Improvement of 2-phenylethanol production in Saccharomyces cerevisiae by evolutionary and rational metabolic engineering', *PLoS ONE*, 16(October), pp. 1–20. doi: 10.1371/journal.pone.0258180.

Zielenkiewicz, U. *et al.* (2009) ‘In vivo interactions between toxin-antitoxin proteins epsilon and zeta of streptococcal plasmid pSM19035 in Saccharomyces cerevisiae’, *Journal of Bacteriology*, 191(11), pp. 3677–3684. doi: 10.1128/JB.01763-08.

Zotta, T. *et al.* (2020) ‘Valorization of cheese whey using microbial fermentations’, *Applied Microbiology and Biotechnology*, 104(7), pp. 2749–2764. doi: 10.1007/s00253-020-10408-2.

Appendix 1

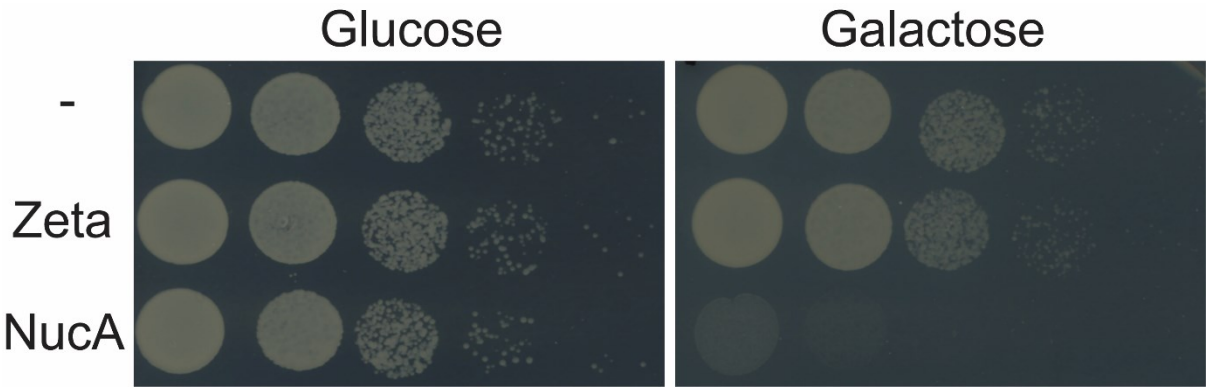


Figure 1. Evaluation of lethal genes in *K. marxianus*. Codon optimised genes encoding *Streptococcus pyogenes* Zeta toxin and *Serratia marcescens* NucA were inserted into a *K. marxianus* functional plasmid for galactose inducible expression under transcriptional control of *ScGAL1* promoter. *K. marxianus* strains transformed with the plasmid backbone or plasmids bearing the *zeta* or *nucA* genes were precultured in minimal medium with xylose for 24 hours after which the cultures were diluted in fresh media to A 600nm of 0.5, further cultured and 3µL of mid-exponential growth phase cultures diluted to A 600nm 1, 0.1, 0.01 and 0.001 were spotted onto the glucose and galactose plates. Xylose and not glucose was used for preculturing to prevent transcription leakage based on available knowledge that the *ScGAL1* promoter in *K. marxianus* is slightly active with glucose.

Appendix 2

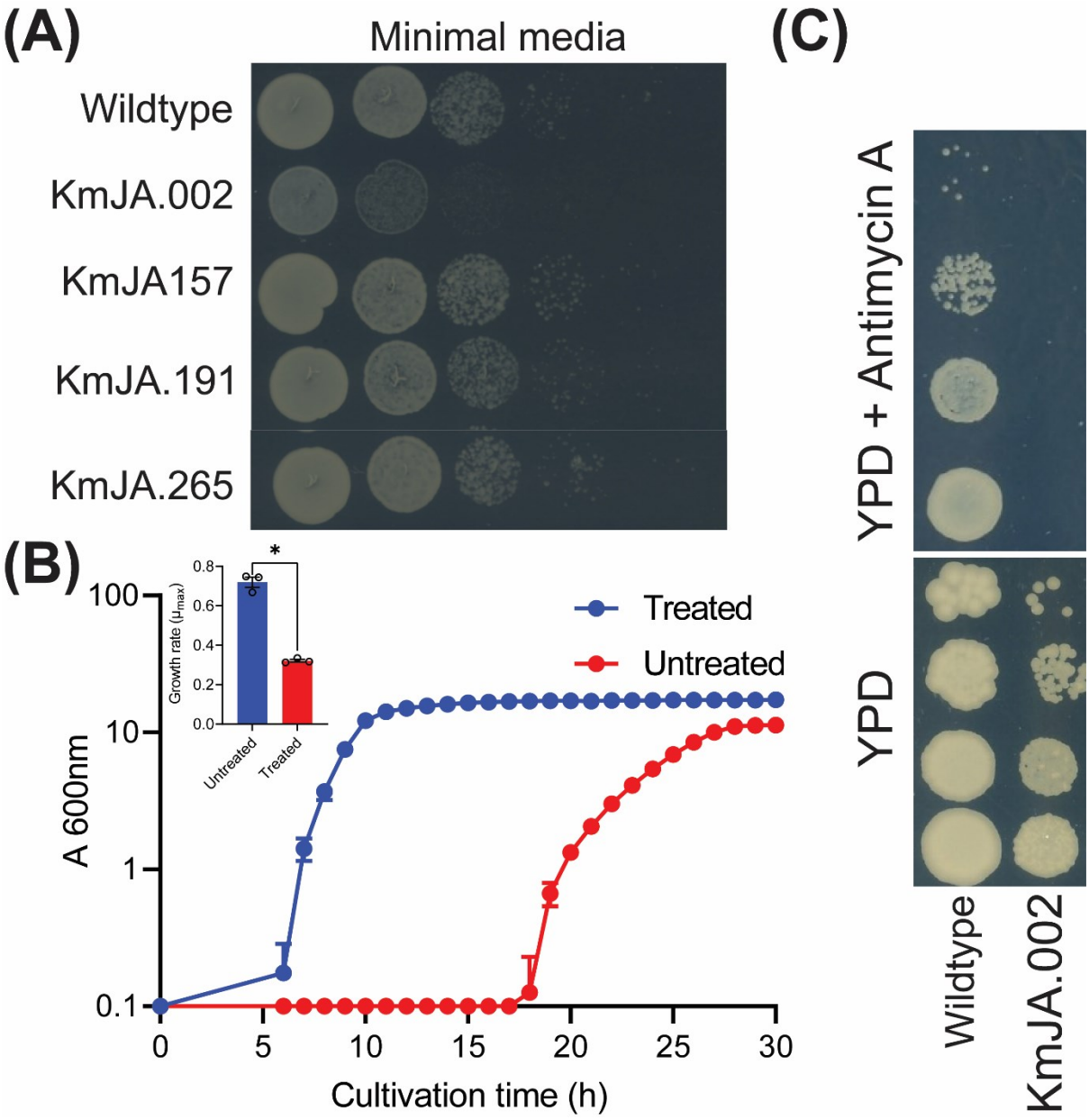


Figure 1. Pyruvate decarboxylase, acetyl-CoA hydrolase, carnitine shuttle and threonine aldolase were disrupted individually by the deletion of *PDC1* (KmJA.002), *ACH1* (KmJA.157) *TMLD* (KmJA.191) and *GLY1A/B* (KmJA.265) in order to determine whether the disruption of any of these mediators of cytosolic acetyl-CoA supply results in growth defect on glucose. Spotting of strains harbouring these deletions onto solid minimal glucose medium showed that only *PDC1* deletion in KmJA.002 caused growth defect, showing that alcoholic fermentation was supporting *K. marxianus* growth in this condition (A). The growth of the wildtype strain in liquid culture showed that respiration blocker increased lag phase time and significantly reduced growth rate by 56%, showing that respiration also plays a role in the growth (B). With the growth defect of KmJA.002 was also observable on YPD complex medium, the strain lost completely its ability to grow in antimycin A-supplemented YPD (C). For the drop test, 3µL of mid-exponential growth phase cultures diluted to A 600nm 1, 0.1, 0.01 and 0.001 were spotted onto glucose plates. Data shown are the mean ± standard deviation of three replicates. Statistically significant difference (examined by independent ttest) in growth rates are denoted by an asterisk (*, $p<0.05$).

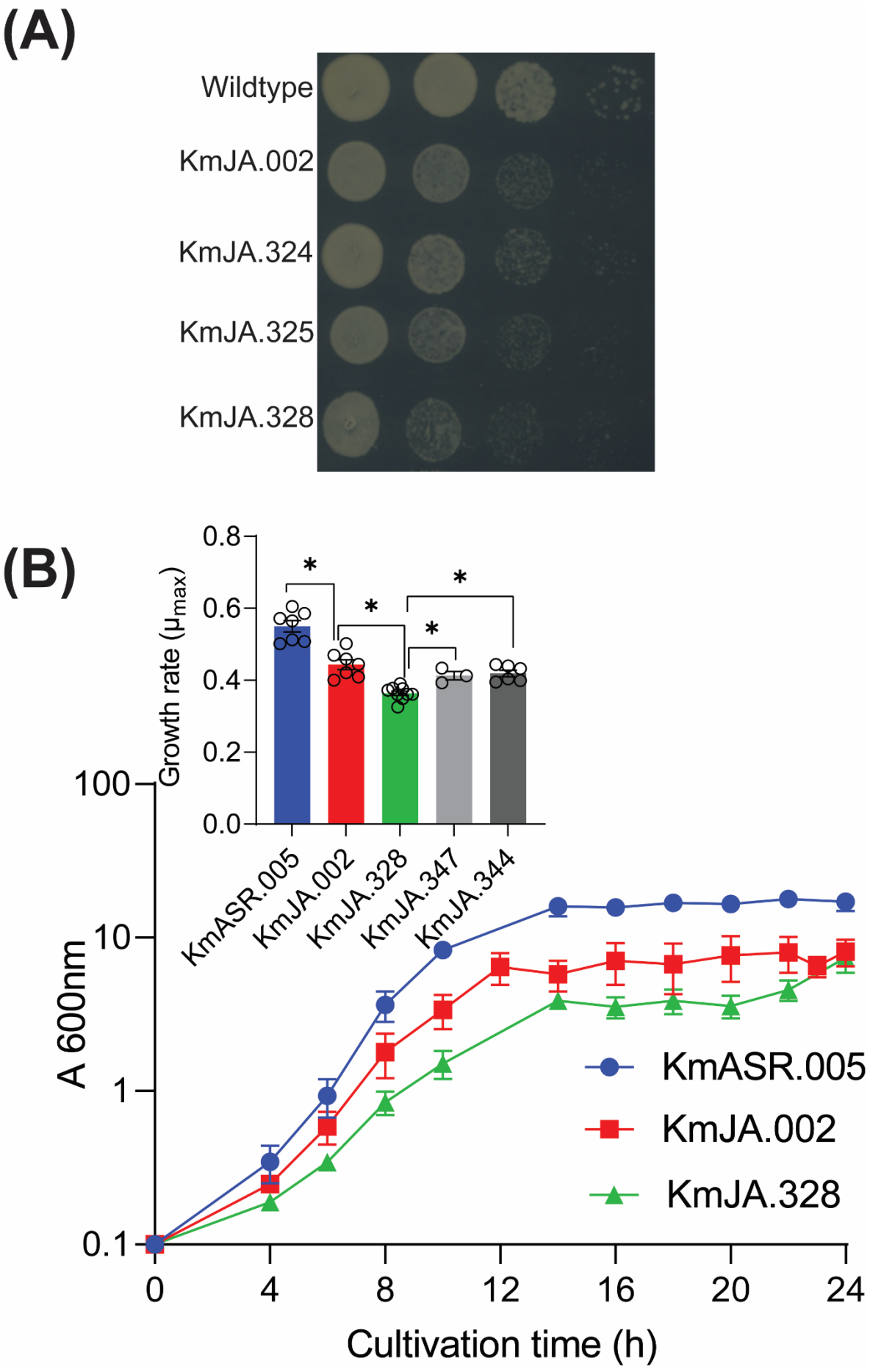

(A)
Wildtype
KmJA.002
KmJA.324
KmJA.325
KmJA.328
(B)
Growth rate (μmax)
0.8
0.6
0.4
0.2
0.0
*
*
*
*
KmASR.005
KmJA.002
KmJA.328
KmJA.347
KmJA.344
A 600nm
100
10
1
0.1
0
4
8
12
16
20
24
Cultivation time (h)
KmASR.005
KmJA.002
KmJA.328

Figure 2. Acetyl-CoA hydrolase *ACH1* (KmJA.324), carnitine shuttle *TMLD* (KmJA.325) and threonine aldolase *GLY1A/B* (KmJA.328) were disrupted in KmJA.002 (*pdc1*) in order to determine whether the disruption of any of these mediators of cytosolic acetyl-CoA supply results in growth defect of this strain on glucose. Spotting of strains harbouring these deletions onto solid minimal glucose medium showed that only the disruption *Gly1A/B* in KmJA.328 intensified the growth defect of *pdc1* (A) which was confirmed in liquid cultures and the genes complemented this phenotype (B). This data suggests that threonine aldolase partially mediates cytosolic acetyl-CoA supply in *pdc1.* For the drop test, 3µL of mid-exponential growth phase cultures diluted to A 600nm 1, 0.1, 0.01 and 0.001 were spotted onto glucose plates. Data shown are the mean ± standard deviation of at least three replicates. Statistically significant difference (examined by independent ttest) in growth rates are denoted by an asterisk (*, $p<0.05$).