Whole genome engineering, evolution, and high-throughput screening for propionic acid production and tolerance in yeast

Xin Xu

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

Department of Molecular Sciences

Macquarie University, NSW, Australia

February 2019

Table of Contents

List of figures

List of tables

List of supplementary files

Acknowledgements

Firstly, I would like to thank my principal supervisor Professor Ian Paulsen for all the guidance and support throughout my PhD study. Thank you for taking the time to discuss my research problems and provide guidance on my thesis writing. I am grateful for all these valuable suggestions. Ian also supported me to present my research in many leading conferences. I feel lucky for these opportunities.

I would like to thank my co-supervisor Dr. Thomas Williams. Tom was very patient to teach me the experiment techniques from the beginning. He gave me significant help in every project and in scientific writing. When experiments went to troubles, he always helped me trouble-shooting and coming up with several solutions. I really appreciate of all the guidance Tom has provided.

I hope to extend my gratitude to Professor Sakkie Pretorius. Back in 2014, through his great talk, I knew about synthetic biology and Sc2.0. He gave me the opportunity to work in our group and being very supportive through my study. I also want to thank Dr. Ren Yi for encouraging me to apply to Macquarie University and providing support during my candidature.

During my PhD study, I am very grateful to have met all the colleagues and friends in the synthetic biology group. I feel lucky to work with so many amazing scientists in Yeast 2.0 group, Alex, Heinrich, John, Niel, Tom, Hugh, Monica, Briardo, Roy, Liz, Kai, Elizabeth, Ben, and Patrick. I wish to especially thank Heinrich, Hugh, Niel and Liz for giving so many constructive advices on my project and giving me so many encouragements. I would like to thank Professor Robert Willows, Kerstin, Dominik, and Alexander for their assistance on

HPLC, thank Professor Christina Divne and Dominic for their help in protein modelling, and thank Kai for the suggestions on Chapter 4. I also appreciate of all the support from Iniga, for arranging all the meetings and processing all the forms.

I am thankful to have met members from Paulsen's group-especially Sheemal, Tao, Hasinika and Liping. They are always very encouraging, and I will cherish our friendship. Sheemal and Tao also gave me significant help in English writing. I really appreciate of that.

Last but not least, I would like to thank my parents for their deepest love and unconditional support. With their support, I could focus on my study and go through all the difficulties. I also want to thank my boyfriend Chang for being so encouraging and caring. Lastly, a big thank you to my friends, Yan, Huiyang, Zizhu, Qingqing, Ping, Penny, Hangrui, Lun, Shan, Li, and Yulong. They have brought so much happiness and made my life in Sydney more colourful.

Declaration

I declare the work presented in this thesis was conducted by me under the direct supervision of Professor Ian Paulsen. None of the work presented has been previously submitted for any other degree.

Xin Xu

Publications

Publications:

Xu X, Williams TC, Divne C, Pretorius IS and Paulsen IT. Evolutionary engineering in *Saccharomyces cerevisiae* reveals a *TRK1*-dependent potassium influx mechanism for propionic acid tolerance. Biotechnology for biofuels. 2019 Dec;12(1):97.

Williams TC, **Xu X,** Ostrowski M, Pretorius IS, Paulsen IT. Positive-feedback, ratiometric biosensor expression improves high-throughput metabolite-producer screening efficiency in yeast. Synthetic Biology. 2017 Jan 1;2(1):ysw002.

Chen W, Deng W, **Xu X,** Zhao X, Vo JN, Anwer AG, Williams TC, Cui H, Goldys EM. Photoresponsive endosomal escape enhances gene delivery using liposome–polycation–DNA (LPD) nanovectors. Journal of Materials Chemistry B. 2018;6(32):5269-81.

Publications in preparation:

Williams TC, Kroukamp H, **Xu X**, Wightman E, Goold HD, Carpenter A, Wyk1 NA, Llorente B, Espinosa M, Daniel E, Nevalainen H, Curach N, Borneman AR, Bader J, Mitchell LA, Boeke JD, Johnson D, Pretorius IS and Paulsen IT. Synthesis and adaptive laboratory evolution of *Saccharomyces cerevisiae* synXIV. To be submitted to Cell.

Conference proceedings

Xu X, Williams TC, Divne C, Pretorius IS, and Paulsen IT. Evolution identifies increased potassium influx through the Trk1 transporter as the mediator of propionic acid tolerance in *Saccharomyces cerevisiae*. Poster presentation at the 7th Yeast 2.0 and Synthetic Genomes Conference, Sydney, Australia, 2018.

Xu X, Williams TC, Pretorius IS and Paulsen IT. Evolutionary engineering of propionic acid tolerance in *Saccharomyces cerevisiae*. Poster presentation at Metabolic Engineering 12, Munich, Germany, 2018.

Xu X, Williams TC, Pretorius IS and Paulsen IT. Improving propionic acid tolerance in yeast via adaptive laboratory evolution. Poster presentation at Synthetic Biology Australasia, Sydney, Australia, 2017.

Contributions

Chapter 2: Synthesis and laboratory evolution of *Saccharomyces cerevisiae* synthetic chromosome XIV.

This work was conceived by Paulsen, Pretorius, Williams and Kroukamp. The project is led by Williams, and the experiments were conducted collaboratively within Paulsen's 'Yeast 2.0 group'. Xu contributed to the construction of megachunk P and the megachunk I-M region, debugging of the G-L regions (including re-integration of megachunk G in the wild-type strain, replacement of synthetic seuquences with wild-typle sequences with CRISPR-Cas9, integration of an intron deleted chunk, Pooled PCRtag analysis of the G-K strain, and overexpression of *MPRL19* and fitness test), and data analyses. The manuscript was written by Williams with contributions from Kroukamp, Xu, Paulsen, and Pretorius.

Chapter 3: Positive-feedback, ratiometric biosensor expression improves high-throughput metabolite-producer screening efficiency in yeast.

This work was conceived by Williams and Paulsen. The biosensor was designed and constructed by Williams. Flow cytometry determination and cell sorting were performed by Williams and Xu assisted by Ostrowski. All data analyses were performed by Williams and Xu. The manuscript was written by Williams with contributions from Xu, Ostrowski, Paulsen, and Pretorius.

Chapter 4: SCRaMbLE and biosensor-mediated selection for improved PA production in yeast

This work was conceived by Williams, Xu and Paulsen. All experiments and data analyses were performed by Xu. The manuscript was written by Xu, with contributions from Williams, Paulsen and Pretorius.

Chapter 5: Evolutionary engineering in *Saccharomyces cerevisiae* reveals a *TRK1*-dependent potassium influx mechanism for propionic acid tolerance.

This work was conceived by Xu, Williams and Paulsen. All experiments were conducted by Xu. Data analyses were performed by Xu, Williams and Divne. The manuscript was written by Xu, with contributions from Williams, Divne, Paulsen and Pretorius.

Abstract

Metabolic engineering of the cell factory *Saccharomyces cerevisiae* (yeast) has enabled bio-production of a wide variety of chemicals, fuels, and pharmaceuticals. However, strain engineering is time-consuming, expensive, and limited by biological knowledge. With the rapid development of synthetic biology, novel approaches can be developed to revolutionise the problematic strain engineering processes. The global consortium '*Saccharomyces cerevisiae version* 2.0 (Sc2.0)' is constructing the world's first synthetic eukaryotic genome with whole genome design and synthesis. One of the design features of Sc2.0 is the inclusion of a synthetic evolution system termed 'SCRaMbLE' (Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution). This can generate high rates of genome rearrangements, which has great potential to be applied in yeast metabolic engineering by generating diverse and novel genomes with superior phenotypes.

Propionic acid (PA), a top platform chemical, has been widely used as a food preservative and a chemical intermediate in many industries. PA production via engineering yeast is potentially a sustainable alternative to current petroleum derived production. In this thesis, novel approaches have been established, combined with synthetic biology approaches and high-throughput screening, for the generation and selection of genetic diversity to improve PA production and PA tolerance.

Firstly, to build the ultimate Sc2.0 platform, I contributed to the construction and de-bugging of synthetic Chromosome XIV. The full synXIV was functional with comparable growth to the wild-type. Secondly, to facilitate the high-throughput screening of PA producers, I contributed to the development of an organic acid biosensor in yeast. The biosensor consists of the native War1p transcriptional regulator and *PDR12* promoter, which transduce the concentrations of

para-hydroxybenzoic acid (PHBA) and PA to a GFP signal. The dynamic range has been further increased by the positive feed-back expression of *WAR1*, and the GFP signal has been normalised to the constitutive expression of mCherry within each cell to control for intrinsic noise.

Thirdly, to improve PA production in yeast, SCRaMbLE of the PA-producing Wood-Werkman cycle was performed in a haploid strain and a synthetic-wild-type hybrid diploid yeast containing synthetic chromosomes III, VI, and IXR. SCRaMbLE of a diploid strain was found to be superior to SCRaMbLE of a haploid strain due to reduced SCRaMbLE-mediated viability loss and a subsequent higher frequency of cells with improved phenotypes. A novel strategy has been developed for SCRaMbLE and biosensor facilitated FACS screening. With this strategy, two strains with 1.7-fold to 2.4-fold improved PA titres were isolated and characterised using whole-genome sequencing. However, their parental diploid strain had all but one of the genes in the synthetic Wood-Werkman cycle deleted before the induction of SCRaMbLE. The two isolates also had lost the synthetic chromosomes. Improved production of PA was attributed to degradation of exogenous amino acids to propionyl-CoA via native enzymes, and the conversion of propionyl-CoA to propionate by the only remaining Wood-Werkman cycle gene, *scpC*.

Lastly, to overcome the toxicity of PA and pave the way for titre improvement in the future, adaptive laboratory evolution (ALE) was conducted with increasing concentrations of PA ranging from 15 mM to 45 mM. This approach successfully improved yeast tolerance to PA by more than 3-fold. Through genome sequencing and reverse engineering, three different single nucleotide substitutions in *TRK1* (encoding a high-affinity potassium transporter) were revealed to be the determinant of elevated PA tolerance. Potassium supplementation assays showed that mutated *TRK1* alleles, as well as addition of potassium in the medium, greatly elevated yeast

tolerance to PA. These experiments have demonstrated that PA tolerance in yeast is mediated by the uptake of potassium ions, which likely promotes the export of protons from PA molecules to maintain pH homeostasis and stabilise membrane potential as protons from PA molecules are exported. The mutations in *TRK1* mediate increased tolerance through increased potassium uptake and affinity. The mechanism not only functions in PA stress but also confers tolerance to multiple organic acid stress conditions.

Overall, this doctoral study has contributed to the construction and understanding of the synthetic Yeast 2.0 genome, and the utilisation of its genome evolution system SCRaMbLE. In particular, it has provided insights in the establishment of novel approaches for strain engineering, and has revealed an organic acid tolerance mechanism that can contribute to future organic acid and lignocellulosic ethanol production processes in yeast.

Abbreviations

3HP	3-hydroxypriopionic acid
3-PG	3-phospho-D-glycerate
ACC1	Acetyl-CoA carboxylase gene
acs_{SE}	Acetyl-CoA synthetase gene from *Salmonella enterica*
ADH2	Alcohol dehydrogenase gene
ALD6	Aldehyde dehydrogenase gene
ALE	Adaptive laboratory evolution
bgl	Beta-glucosidase gene
BsKtrB	*Bacillus subtilis* KtrB
CDS	Coding sequence
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
DiCRISPR	Delta-CRISPR
E. coli	*Escherichia coli*
EBD	Estrogen binding domain
EMS	Ethyl methane sulfonate
FACS	Fluorescence activated cell sorting
FDA	Food and Drug Administration

FSC	Forward-angle light scatter
G-3-P	Glyceraldehyde 3-phosphate
GAPN	Non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase
GC	Gas chromatography
GFP	Green fluorescent protein
GRAS	Generally recognised as safe
gRNA	Guide RNA
HI-CRISPR	Homology-integrated CRISPR−Cas
HPLC	High-performance liquid chromatography
IML	The intramembrane loop
INDELS	Insertions and deletions
IS	Insertion sequence
ITS	Internal transcribed spacer
LB	Luria-Bertani
mce	methyl malonyl-CoA epimerase gene
mcr_{Ca}	malonyl-CoA reductase gene from *Chloroflexus aurantiacus*
mtc	malonyl-CoA carboxytransferase gene
mut	methyl-malonyl-CoA mutase gene
NGS	Next-generation sequencing
OD_{600}	Optical density at 600 nm

ORF	Open reading frame
PA	Propionic acid
PAM	Protospacer adjacent motif
P. acidipropionici	*Propionibacterium acidipropionici*
PCB	Phycocyanobilin
pcs	Propionyl-coa synthase gene
PHBA	Para-hydroxybenzoic acid
PoPM	Pooled PCRTag mapping
pst	propionyl-CoA: succinyl-CoA transferase gene from *P. acidipropionici*
RBS	Ribosome-binding site
rDNA	Ribosomal DNA
RE	Restriction enzyme
S. cerevisiae	*Saccharomyces cerevisiae*
SC	Synthetic complete
Sc2.0	*Saccharomyces cerevisiae* version 2.0
scpC	propionyl-CoA: succinate-CoA transferase gene from *E. coli*
SCRaMbLE	Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution
ScTrk1p	*Saccharomyces cerevisiae* Trk1p
SD	Standard deviation

SNP	Single-nucleotide polymorphism
SSC	Side-angle light scatter
SSM	Secondary-structure matching
SYN	Synthetic
TCA	Tricarboxylic acid cycle
tracrRNA	Trans-encoded RNA
tRNA	Transfer RNA
UTR	Untranslated region
VpTrkH	*Vibrio parahaemolyticus* TrkH
WGS	Whole genome sequencing
WT	Wild type
YNB	Yeast nitrogen base
YOGE	Yeast oligo-mediated genome engineering
YPD	Yeast extract peptone dextrose
YPG	Yeast extract peptone glycerol

Chapter 1 : Introduction

1.1 *Saccharomyces cerevisiae* as a model eukaryote

The yeast *Saccharomyces cerevisiae*, known as baker's yeast, has a long history of use in food and beverage production. Yeast has several superior properties that make it the most deeply characterised and widely applied eukaryotic microorganism. For example: rapid growth, an easily manipulated genome, extensive knowledge of genetics, physiology and biochemistry, and a high-efficiency transformation system, support basic and applied yeast research. It is also a non-pathogenic microorganism, which has been 'Generally Recognised as Safe' (GRAS) by the American Food and Drug Administration (FDA).

The *S. cerevisiae* genome was completely sequenced in 1996. The total genome size is 12,068 kilobases with 5885 protein encoding genes identified (1). Following the release of the sequences, genome-wide mutants and genome-wide deletions of haploid and diploid yeast strain have been constructed to understand the function of the yeast genes. Other advances in techniques including DNA microarrays for studying the transcriptome (2, 3), synthetic lethality to understand genetic interactions (4), protein localization (5), enzymatic activities (6, 7), and protein-protein interactions by two-hybrid analysis (8, 9). Yeast genome and proteome databases including Saccharomyces Genome Database, Comprehensive Yeast Genome Database, Kyoto Encyclopedia of Genes and Genomes are also generally accessible. These greatly facilitate the development of yeast biotechnology.

Yeast is robust, stress tolerant and can grow at relatively low pH making it an ideal microbe for industrial fermentation. With rapid development of metabolic engineering techniques and synthetic biology tools, it has been engineered for producing a wide range of industrial products (Figure 1-1). Yeast is the preferred cell factory for producing biofuels such as ethanol (10), biobutanol (5, 11), and biodiesels (12) and building block chemicals such as lactic acid (13), succinic acid (14), and 1,2-propanediol (15). In addition, it has also been engineering to produce high-value pharmaceuticals such as human insulin (16, 17), the antimalarial precursor artemisinic acid (18), and hepatitis vaccines (19), and nutraceuticals such as resveratrol (20).

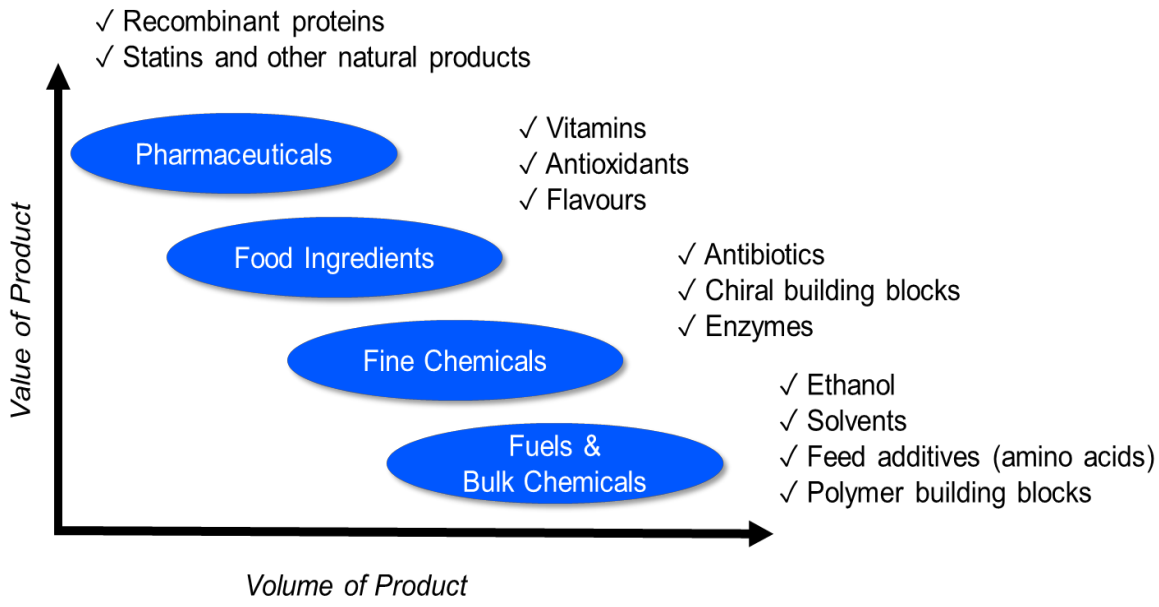


Figure 1-1. Various kinds of biotech products produced by *S. cerevisiae*.The products range from high-volume/low-value to low-volume/high value. Reproduced from Hone et al., 2012, Cellular and Molecular Life Sciences.

1.2 Metabolic engineering

In the last three decades, yeast strain improvement has predominantly been facilitated by the development of recombinant DNA techniques. Recombinant DNA technology, when used to enable the directed improvement of product formation or cellular properties through the modification of specific biochemical reactions or the introduction of new ones, has been defined as metabolic engineering (21).

Classic metabolic engineering strategies are mainly focused on direct modification of target genes, enzymes, and regulatory networks. Metabolic engineering can also involve the expression of heterologous enzymes and the elimination of the competing pathways. However, due to the complexity of biological systems and their tightly regulated networks, classical strain engineering does not always result in improved phenotypes. One additional issue is that in some cases, the genes or genetic networks that could improve a specific phenotype haven't been discovered yet. In these cases, inverse metabolic engineering strategies could be used to circumvent the limitation through the selection of the mutant with desired phenotype, followed by identification of the causal gene and implementation of the principle back into the industrial

strain to achieve a better yield (22, 23). Among the inverse engineering strategies, adaptive laboratory evolution (ALE, also known as evolutionary engineering) might be advantageous since it is effective, easy to set up and can generate a desired phenotype derived from multiple genetic changes (23).

1.2.1 Adaptive laboratory evolution (ALE)

Adaptive laboratory evolution (ALE) is a common approach to obtain improved phenotypes and to study the mechanism of evolution during prolonged growth under defined conditions. The principle of ALE is based on simple Darwinian evolution, except by artificial rather than natural selection.

Microorganisms have become a favourite choice for the application of ALE since they can be easily cultured in the laboratory conditions, replicate rapidly in a large population, and can evolve for several hundred generations in a few weeks. Microbial cells have an estimated mutation rate of 10^{-10} per base pair per replication (24). During prolonged culture, beneficial mutations can occur within a large population and accumulate when gradually exposed to the selective pressure (25). Thus, over time the evolved strains with beneficial mutations take over the population and can be isolated at the end of the experimental evolution period.

One common method of ALE is repeated batch fermentation in shake flasks. In this scenario, an aliquot of culture grown to a certain stage is transferred into fresh medium at intervals for another round of growth (Figure 1-2A). During the evolution, the nutrition is not limited, and the cells are mostly in exponential phase. In terms of adaption to stress conditions, the selective pressure could be increased gradually during ALE when an improvement in growth-rate or yield has been observed (26). This method has the advantages of low cost and replicability, where independent lineages can be evolved in parallel. However, the environmental conditions, such as pH, temperature, oxygen and nutrients, may not be consistent. An alternative method is using chemostats. In the chemostat, the environmental conditions can be controlled by a feed-back system and the growth rate is usually held constant by the supply of limited carbon and

nitrogen source, and the removal of waste and cells at defined rates (Figure 1-2B). This approach is costlier, and performed at a larger scale, and therefore and can't be performed with large numbers of replicates in parallel (27).

With the development of systems biology techniques, the analysis of ALE outcomes has been greatly enhanced by next-generation sequencing (NGS) and other omics techniques. To uncover the genotype-phenotype correlations, whole genome sequencing (WGS) is performed on evolved strains with the improved phenotypes. Various types of mutations are detected (Figure 1-2C): single-nucleotide polymorphisms (SNPs), insertions and deletions (indels), transposable element (insertion sequence, IS) movements, and deletions as well as duplications of large genomic regions (27). The causal mutations can then be reserve engineered into the parental strain to test their function and contribution to the improved phenotype. The mechanisms underlying the genetic change could be further analysed by transcriptomics, proteomics, and metabolomics.

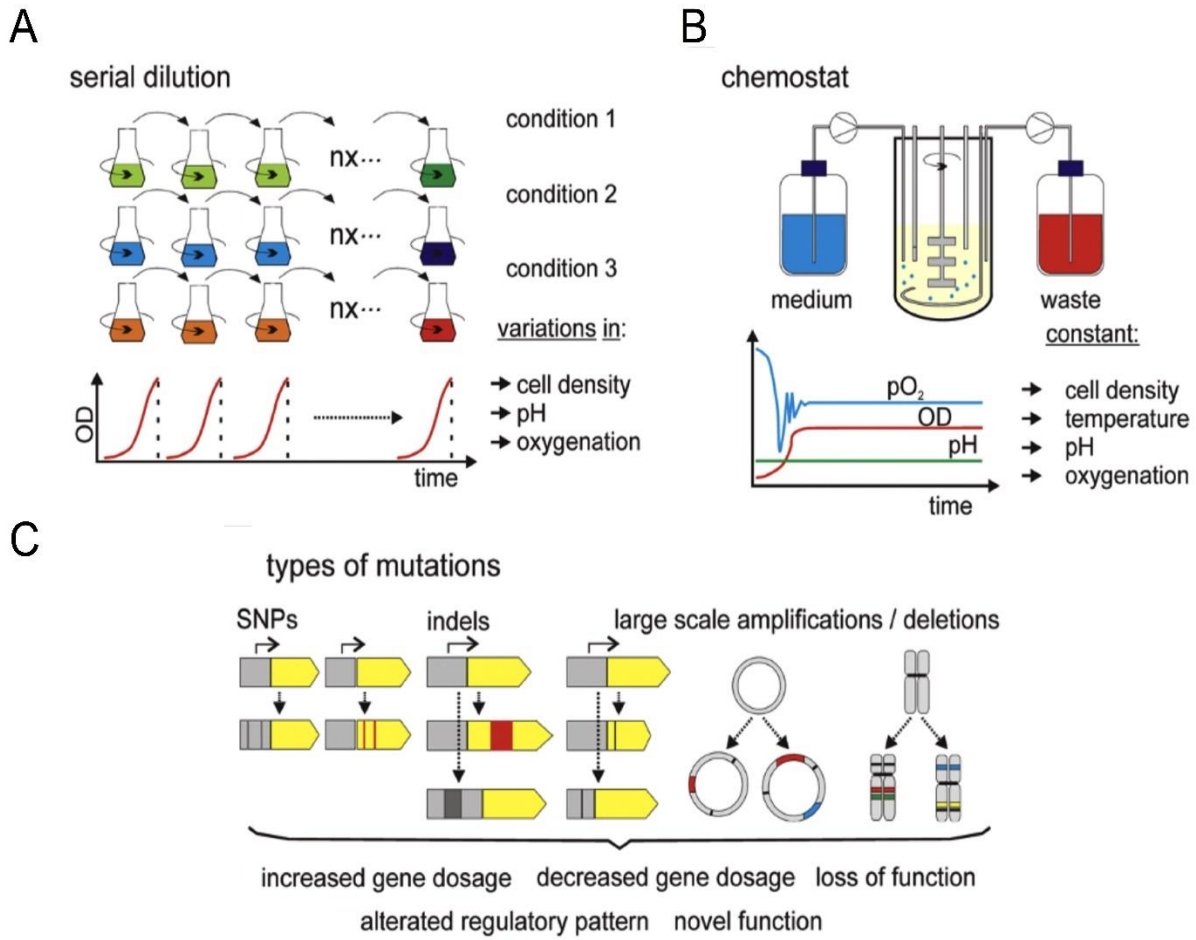


Figure 1-2. Adaptive laboratory evolution (ALE). Two common methods to perform ALE are (A) repeated batch fermentation in shake flasks and (B) chemostat cultures. (C) Various types of mutations that can be identified in ALE studies. Reprinted from Dragosits et al., 2013, Microbial Cell Factories.

ALE has been widely used to improve industrial strain and to inform rational engineering principles, especially for model organisms. Different engineering purposes can be achieved through ALE including increasing utilization rates of native substrate (28), activating latent metabolic pathways to use non-native substrates (29), increasing the growth rate of recombinant strains (30), increasing product titers and yields (31), and adaption to environmental stress conditions (32). The studies of ALE in *S. cerevisiae* mainly focus on two categories: evolution for substrate utilization and for tolerance to stress conditions (Table 1-1). Recently, facilitated by the development of synthetic biology tools, ALE can also be used to improve metabolite production in yeast. For instance, combined with heterologous pathway expression and the high-throughput biosensor screening, ALE has been demonstrated a powerful tool to increase the yield of muconic acid (33).

Table 1-1. Examples of ALE experiments conducted with *S. cerevisiae*

Strain	Environment	Phenotype	Mechanism	Ref.
CP1AB (an isogenic diploid derived from the haploid XC500A)	glucose-limited chemostat, 250 generations	3-fold greater biomass with improved efficiency of glucose utilization	DNA microarrays showed altered gene expression in the evolved strains was involved in glycolysis, the tricarboxylic acid cycle, oxidative phosphorylation, and metabolite transport	(1)
TMB3001 (expressing the xylose metabolic pathway, mutagenised with EMS before ALE)	chemostat, anaerobic growth on xylose, 460 generations	growth on xylose as the sole carbon source under strictly anaerobic condition	N/A	(2)
IMS0003 (engineered for both xylose and arabinose utilization)	chemostat, sequential batch cultivation on glucose, xylose, and arabinose, 40 days and 20 cycles	40% reduction in the time required to completely ferment a mixture containing 30 g/1 glucose, 15 g/1 xylose, and 15 g/1 arabinose	N/A	(3)
H131 (expressing the xylose metabolic pathway)	aerobic, anaerobic, sequential batch cultivation and in a xylose-limited chemostat, 190 generations	increased anaerobic growth rate (0.203±0.006 h^{-1}) and xylose consumption rate (1.866 g g^{-1} h^{-1}) along with high ethanol yield (0.41 g/g)	functional complementation of genomic DNA fragments of the evolved strain revealed that elevated expression level of xylose isomerase improved xylose assimilation.	(4)
BY4741	salt (NaCl) stress, 300 generations	improved fitness in salt stress condition	DNA microarrays and whole genome population sequencing showed that salt stress adaption is attributed to genome size increase and expression changes in several genes including *CTT1*, *MSN4*, *HLR1* and *MOT2*	(5)
CEN-PK 113-7D	batch culture in HCl and lactic acid stress, 10-12 weeks (277, 281, 317 generations)	31.3% and 200% increase in maximum specific growth rate treated with HCl and lactic acid, respectively	whole genome sequencing and RNA-seq analysis revealed an altered sterol composition and impaired iron uptake contributed to HCl tolerance whereas a multicellular morphology and rapid lactate degradation improved tolerance to lactic acid	(6)
W303-1A (mutagenised with EMS and non-mutagenised)	chemostat with increasing concentrations of ethanol, 192 days for non-mutagenised strain and 14-28 days for mutagenised strain.	survivability in lethal ethanol concentrations (at 20% (v/v) ethanol) was considerably improved.	the tolerance might be associated with higher glycerol production rates, which are important for maintaining redox balance by oxidizing NADH to NAD^+	(7)
CEN.PK (mutagenised with EMS before ALE)	batch selection for freezing–thawing stress, up to 68 generations	102-, 89-, 62-, and 1429-fold increased resistance to freezing–thawing, temperature, ethanol, and oxidative stress, respectively	N/A	(8)
YLH2 (expressing carotenoid production pathway, Δ*CTT1*)	ALE using a periodic hydrogen peroxide shocking strategy, 34 and 40 transfers	3-fold increase in carotenoids production	transcriptome analysis showed upregulation of genes were related with lipid biosynthesis and mevalonate biosynthesis pathways	(9)

N/A represents no relevant data being found; EMS is short for ethyl methane sulfonate.

1.3 Synthetic biology

Yeast has been engineered for producing a wide range of industrial products. In recent decades, the engineering process has been significantly advanced by the development of synthetic biology. Synthetic biology is a young interdisciplinary research field that brings together biology with engineering. The emerging of synthetic biology has been greatly facilitated by the ever-improving techniques and ever-decreasing cost of DNA sequencing and synthesis.

Traditionally, genetic engineering focuses on the manipulation of one gene or a pathway. However, synthetic biology, focuses on the design and construct of the biological systems in a bottom up way with wholescale changes being considered (42). It not only addresses the existing problems but also creates new functions.

The first wave of synthetic biology focused on construction of small genetic devices and modules (42). The devices and modules can be used to control gene expression and regulate protein function. The elements used for the construction of synthetic networks include promoters, terminators, and regulatory regions for transcriptional regulations; RNAi and ribosomal binding sites for translational regulations; phosphorylation cascades, protein degradation and localization signals for post-translational regulations. The second wave is capable of engineering large complex biological systems (42). The fast development of DNA assembly techniques has greatly facilitated this progress. Methods such as Golden Gate assembly (43), Gibson assembly (44), ligase cycling reaction (45), circular polymerase extension cloning (46), and yeast assembly (47), allow fast and scarless assembly of large DNA fragments. At this level, the synthetic genetic modules are integrated into the host strains to create a novel system. However, owing to the complex biological processes in cellular systems, engineered synthetic constructs may not function as predicted. The performance of the synthetic system is then analysed by omics approaches to give guidance to the future engineering.

The rapid development of synthetic biology has the potential to revolutionise the field of agriculture, environment, biology, biotechnology, medicine and human health. Currently,

synthetic biology tools have been applied in the field of metabolic engineering to improve bioproduction of fuels (48), chemicals (49), and pharmaceuticals (50). The following sections will introduce three kinds of synthetic biology tools that have been widely used in yeast metabolic engineering: basic elements for controlling gene expression such as promoters, terminators and transcription factors; biosensors as reporter systems to speed up the screening of cells of interest; chromosomal editing tools for genetic manipulation at a chromosomal level.

1.3.1 Gene expression toolbox

Promoters, terminators and transcription factors are basic elements for controlling gene expression. For the engineering of the genetic modules, the expression of each gene unit in the module needs to be controlled at a different level or induced at a different time-point. The native elements sometimes can't meet the requirement for the specific requirement. To tackle the problem, synthetic genetic circuits can be designed for precise control of gene expression.

Many studies have been performed to characterise native promoters and create synthetic promoters of *S. cerevisiae*. Constitutive promoters are the most commonly used in yeast metabolic engineering, e.g. p*TEF1*, p*TDH3*, p*PGK1* and p*ADH2*, with their expression strength well characterised under different carbon sources (51). Alternatively, inducible promoters, such as p*GAL1* and p*CUP1*, can be used where gene expression can be turned on in response to a chemical trigger. Their usage is extremely helpful when there is a need to express genes that severely impact growth. To further enrich promoter libraries, synthetic promoters have been created either by rational engineering or by random mutagenesis of native promoters, followed by screening (52). For the latter, error-prone PCR is a common approach to create various mutants of native promoters. Promoter libraries of *TEF1* and *DAN1* have been created by error-prone PCR and the parameters of the promoters were also quantified so that the suitable promoter can be easily chosen for a specific design requirement (53). Compared with promoters, terminator studies have drawn less attention. However, terminators are also critical for gene expression as it can influence stability and abundance of mRNA (54). Over 30 terminators

were characterised for protein expression in *S. cerevisiae* and up to an 11-fold difference in expression was identified between an expression-enhancing terminator and the parent terminator (55).

The toolbox is further expanded by the engineering of transcription factors to create synthetic versions. Typically, transcription factors contain DNA-binding domains and transcriptional activation domains. By changing the sequences of target promoters or engineering the DNA-binding domains, synthetic transcription factors can bind various promoters to activate or repress their transcription (52). One of the commonly used synthetic transcription factors are zinc finger synthetic transcription factors (52). Their DNA binding domains have the advantages of small size, functional independence, modularity, and easy combination with activation and repression domains to generate synthetic zinc finger transcription factors (56, 57). The native zinc finger recognises a 3–4 base pair DNA sequence and can be engineered to bind to other non-native recognition sites (58, 59). Several zinc finger domains can be linked together to bind specific DNA sequences within a genome (59). A library of more than 100,000 synthetic zinc-finger transcription factors were created, and various phenotypes were observed in yeast cells expressing these synthetic transcription factors, such as drug resistance, thermotolerance or osmotolerance (60).

1.3.2 Biosensors

Transcription-based synthetic devices are also applied as biosensors for detecting a wide range of metabolites. Biosensors are genetically encoded devices that can sense environmental changes or a particular metabolite and trigger an output signal (Figure 1-3A). They are mainly divided into two classes: transcriptionally regulated biosensors and RNA biosensors (25). Transcriptional regulators are the dominant class of biosensors with the advantage of response to a variety of metabolites. Transcriptionally regulated biosensors are constructed by fusing the metabolite-responsive promoters to the open reading frame of the output gene. RNA biosensors regulate gene expression via transcription termination, translation initiation, and mRNA

stability. Riboswitches are a major group of RNA sensors that are widespread in nature. Riboswitches can bind to small molecules and modulate mRNA secondary structure in the 5' untranslated region (UTR) through altering the accessibility of a ribosome-binding site (RBS) to the ribosome. The binding of the riboswitch to the target metabolite therefore influences the translation process (25).

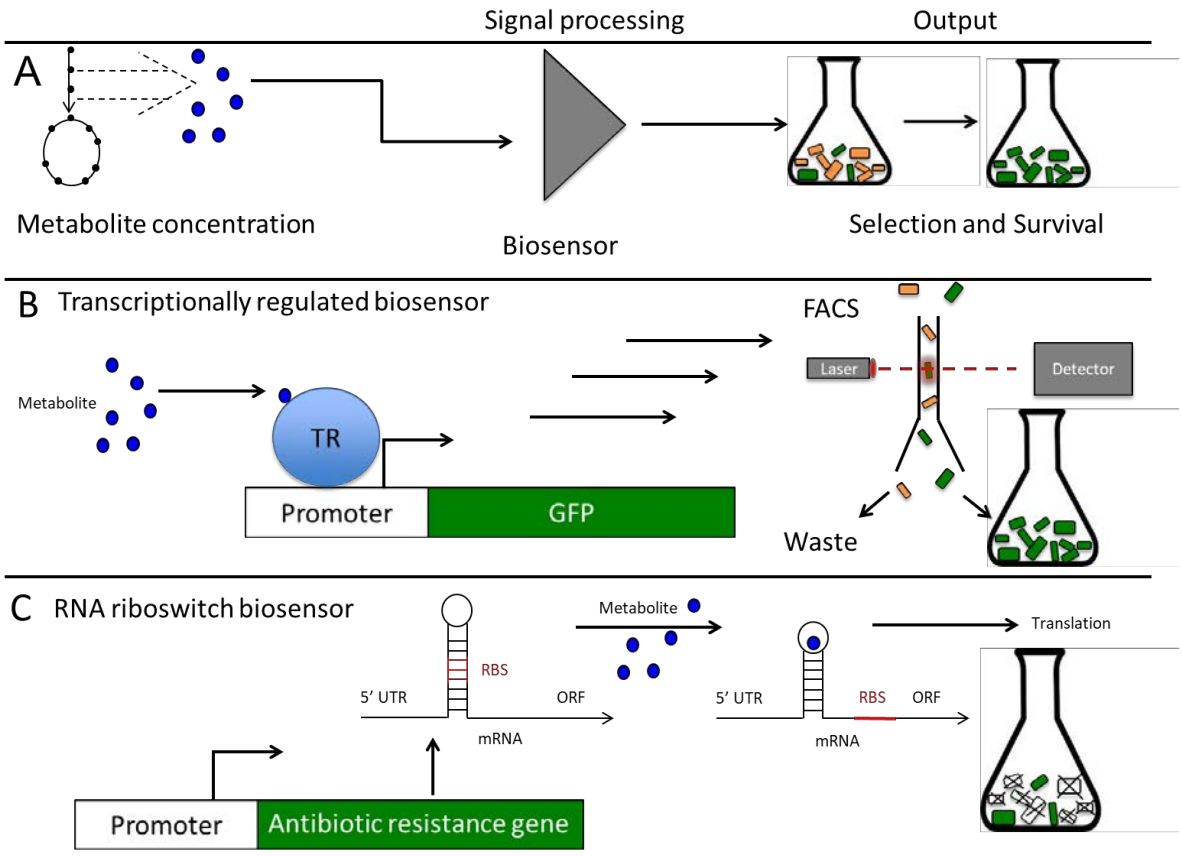


Figure 1-3. Biosensor based screening. (A) A general concept of in vivo biosensor. (B) A transcriptionally regulated biosensor. An allosterically controlled transcription regulator is activated by a specific metabolite produced in the cell and binds to the target promoter, generating green fluorescent protein (GFP) fluorescence signal. The producer cells generate high fluorescence (coloured green) can be screened by fluorescence activated cell sorting (FACS). (C) A riboswitch biosensor. A riboswitch binds specifically to a target metabolite and results in a conformational change of mRNA that exposes the RBS. Therefore, the translation is initiated with antibiotic resistance gene expressed. Figure reproduced from Williams et al., 2016, Trends in biotechnology.

There is an increasing amount of research focusing on developing metabolite biosensors. Biosensors can be coupled to an output signal easy for detection such as GFP and antibiotic resistance (Figure 1-3B&C). Combined with FACS or multi-well plate screening, biosensors can enable high-throughput screening. This is extremely attractive when comes to the screening

of a library of microbial strains for a desired phenotype. Compared with traditional analytical methods, such as gas chromatography (GC) and high-performance liquid chromatography (HPLC), biosensor-mediated screening has several advantages: first, it enables a higher-throughput of screening (up to millions of cells in a few seconds); second, the sample preparation is much easier, and the assay is less costly; third, coupled with FACS, a biosensor can monitor the dynamics of the metabolite in real-time within the living cells (61).

1.3.3 Chromosomal editing tools

Instead of manipulating one gene or one regulatory element, synthetic biology provides tools that facilitate engineering at the scale of whole chromosomes. This means multiple genes can be edited at the same time, which is highly desirable in metabolic pathway engineering.

DiCarlo et al. have established a high-frequency recombination engineering system named as yeast oligo-mediated genome engineering (YOGE) (62). To increase the oligonucleotide integration efficiency, genes responsible for mismatch repair were knocked out and single-stranded DNA recombinases were overexpressed. Taking advantage of the high endogenous homologous recombination machinery in yeast, YOGE can generate up to 10^5 recombinant mutants at each round and can be performed iteratively to increase the efficiency of the multi-locus modifications.

In bacteria and archaea, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) system function as an immune system to disrupt foreign invasive DNA (63). The CRISPR-Cas system has been demonstrated to be a high-efficient genome editing tool in a variety of organisms including yeast, plants, animals and human cell-lines (64-68). CRISPR-Cas9 has the advantage of no markers or plasmids needing to be maintained in a modified strain, and the guide RNA (gRNA) is easily synthesised for targeting to different genomic loci, and for multiplexing. In *S. cerevisiae*, the CRISPR-Cas system (Figure 1-4A) with double-stranded oligonucleotide donors increased homologous recombination rates by 130-fold (68). Another system termed ‘Homology-integrated CRISPR−Cas’ (HI-CRISPR)

achieved multigene disruption in *S. cerevisiae* in one-step with an ultra-high copy number plasmid carrying iCas9 (a Cas9 variant), trans-encoded RNA (tracrRNA), and a homology-integrated CRISPR RNA (crRNA) cassette (69). With this approach, *CAN1*, *ADE2*, and *LYP1*, were simultaneously knocked out with an efficiency ranging from 27 to 87%, and three genes *ATF2*, *GCY1*, and *YPR1*, involved in an artificial hydrocortisone biosynthetic pathway, were simultaneously disrupted with 100% efficiency (69). Delta-CRISPR (DiCRISPR) was established using the HI-CRISPR system with gRNA targeting the multiple delta sites in the yeast genome (Figure 1-4B) (70). With this platform, large biochemical pathways (up to 24 kb) were integrated at 18-copies in a single step (70).

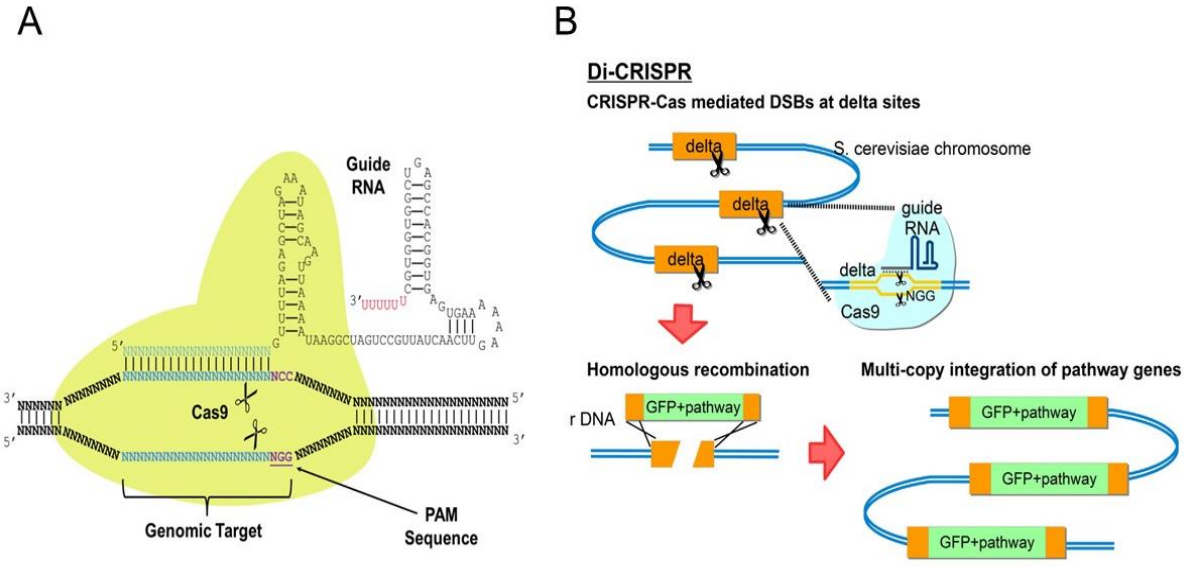


Figure 1-4. Schematic representations of the CRISPR/Cas9 system and the Di-CRISPR systems. (A) The CRISPR/Cas9 system. Cas9 protein complexes with gRNA to bind the target DNA and direct endonuclease activity proximal to the protospacer adjacent motif (PAM) sequence. Reproduced from DiCarlo et al., 2013, Nucleic Acids Research. (B) The Di-CRISPR systems. The Di-CRISPR system generates multiple double strand breaks at the delta sites by CRISPR-Cas to facilitate homologous recombination of pathways at these sites. Reproduced from Shi et al., 2016, Metabolic Engineering.

Chromosomal editing tools such as YOGE and CRISPR enable the evolution of the genome, editing at multiple gene locus and integration of large homologous pathways. However, the editing is still site-specific and mostly relies on the introduction of the oligonucleotide or donor DNA fragments, which becomes limiting for multiplex applications. As the synthesis of large

DNA sequences becomes cheaper, faster and more accurate, synthetic biologists have started to design and build entirely artificial chromosomes and even synthetic genomes.

1.4 *Saccharomyces cerevisiae* version 2.0 (Sc2.0)

The difficulty of engineering a native organism results from the complexity of the biological system. One way to tackle the problem is the construction of a simpler artificial genome. Nowadays the development in DNA synthesis and the computational tools have paved the way for the design and the synthesis at a genome level. Eventually, by the construction of an artificial genome, only the essential genes and advantageous functions will be maintained, and the evolutionary baggage can be cut down, making it simpler to understand and easier to customise (71).

To start with, the genome of poliovirus was synthesised in 2002. The synthetic poliovirus induced the same neurovirulent phenotype as wild-type poliovirus (72). In 2010, the first chemically synthesised bacterial genome of *Mycoplasma mycoides* (JCV-syn1.0) was created by researchers at the J. Craig Venter Institute (73). Based on JCV-syn1.0, a new version of a synthetic minimal genome was created (JCV-syn3.0) through design, build, and test cycle. (74). Following the success in viruses and bacterial, a global team of synthetic biologists led by Jef Boeke has started an ambitious project '*Saccharomyces cerevisiae version* 2.0 (Sc2.0)' to build the first synthetic eukaryotic genome. As part of the Sc2.0 consortium, our group is leading the synthesis of chromosomes XIV and XVI.

In the future, the synthetic yeast genome can contribute to a greater fundamental understanding of genetics and cell biology, and also serve as a platform that is easier to be engineered for producing wide range of industrial products. In addition, *S. cerevisiae* is also a model organism of choice in medicine related research. 30 % of known genes which result to human disease have orthologs with yeast (75). Thus, Sc2.0 could be explored to understand genes related to human disease and generate therapeutics (76).

1.4.1 Design

The native haploid yeast genome contains 16 well characterised chromosomes with 5885 protein encoding genes identified, within which about 5000 genes are individually non-essential (1). Genetic redundancy and unstable elements have also been observed in yeast chromosomes (1). The main aim of building a synthetic yeast genome is to gain more understanding of yeast genomics and generate a genome that is simpler, more stable, and easier for engineering. Three design principles were applied in Sc2.0: the phenotype and the fitness of the synthetic strain should be as close as possible to the wild type; genetic elements that cause instability need to be removed; it should have genetic flexibility to facilitate future studies (77). According to these principles, a range of systematic changes were made *in silico* with the genome editing platform Biostudio (77, 78), summarised in Table 1-2. The changes are as follows: unstable or redundant elements including retrotransposons, subtelomeric repeats and introns were removed; the repetitive transfer RNA (tRNA) genes were relocated to a 'neo-chromosome' to test their function and stability in the future; in addition, the following elements were replaced: first, TAG stop codons were swapped with TAA to release one stop codon to enable the introduction of new artificial amino acids in the future; second, short regions were synonymously recoded to introduce or delete restriction sites for the integration of the synthetic gene fragments; lastly, strings of codons were also recoded synonymously as 'PCRtags', which functions as watermarks to distinguish the synthetic sequences from the wild-type sequences (77). Another important feature embedded in the synthetic chromosome is the symmetrical loxPsym sites. They were inserted 3 bp after the stop codon of all non-essential genes and other major landmarks and can be applied to generate genomic diversity upon induction of a heterologous Cre-recombinase enzyme (77).

Table 1-2. Summary statistics for design of Sc2.0.

	WT size	SYN size	No. of stop codon swaps	No. of loxPsym sites added	bp of PCRTag recoded	bp of RE sites recoded	No. of tRNA deleted	bp of tRNA deleted	bp of repeats deleted
chr01	230208	181030	19	62	3535	210	4	372	3987
chr02	813184	770035	93	271	13651	1215	13	993	7030
chr03	316617	272195	44	100	5272	250	10	794	7358
chr04	1531933	1454671	183	479	25398	2298	28	2261	11674
chr05	576874	536024	61	174	8760	813	20	1471	11181
chr06	270148	242745	30	69	4553	369	10	835	9297
chr07	1090940	1028952	126	380	17910	1572	36	2887	13284
chr08	562643	506705	61	186	9980	714	11	878	19019
chr09	439885	405513	54	142	7943	436	10	736	11632
chr10	745751	707459	85	249	12582	1102	24	1853	7523
chr11	666816	659617	68	199	11769	1017	15	1243	4214
chr12	1078177	999406	122	291	15129	1539	19	1646	10843
chr13	924431	883749	100	337	15911	0	21	1691	7673
chr14	784333	753096	96	260	13329	1113	14	1152	5115
chr15	1091291	1048343	147	399	18015	2058	20	1612	9542
chr16	948066	902994	127	334	15493	1374	17	1338	10048
Total	12071297	11352534	1416	3932	199230	16080	272	21762	149420

WT, wild type; SYN, synthetic. RE, restriction enzyme.

Reproduced from Richardson et al., 2017, Science.

1.4.2 Construction and de-bugging

Sc2.0 is a global consortium that involves leading research groups from six countries working together on chromosome construction. Sixteen chromosomes were assigned to each or several groups separately, and each chromosome was divided into many synthesis units and sub-units (Figure 1-5A). From the smallest unit-oligos, the synthetic sequences gradually build up by homologous recombination to make the whole synthetic chromosome (79). At the beginning, the synthetic chromosomes were built from oligonucleotides. To speed up the construction, DNA chunks are now synthesised from companies instead. Restriction enzyme sites were

designed at each end of the chunks allow them to be ligated into megachunks, and the megachunks are swapped into the chromosome by the iterative replacement of the auxotrophic markers *URA3* or *LEU2* (Figure 1-5B). By meiotic recombination, several synthetic arms combine into a whole synthetic chromosome. PCRTag analysis (80) is performed to ensure the replacement of wild-type sequences into synthetic sequences after integration of each megachunk. In the end, the complete genome of the synthetic strain is checked by whole genome resequencing (80).

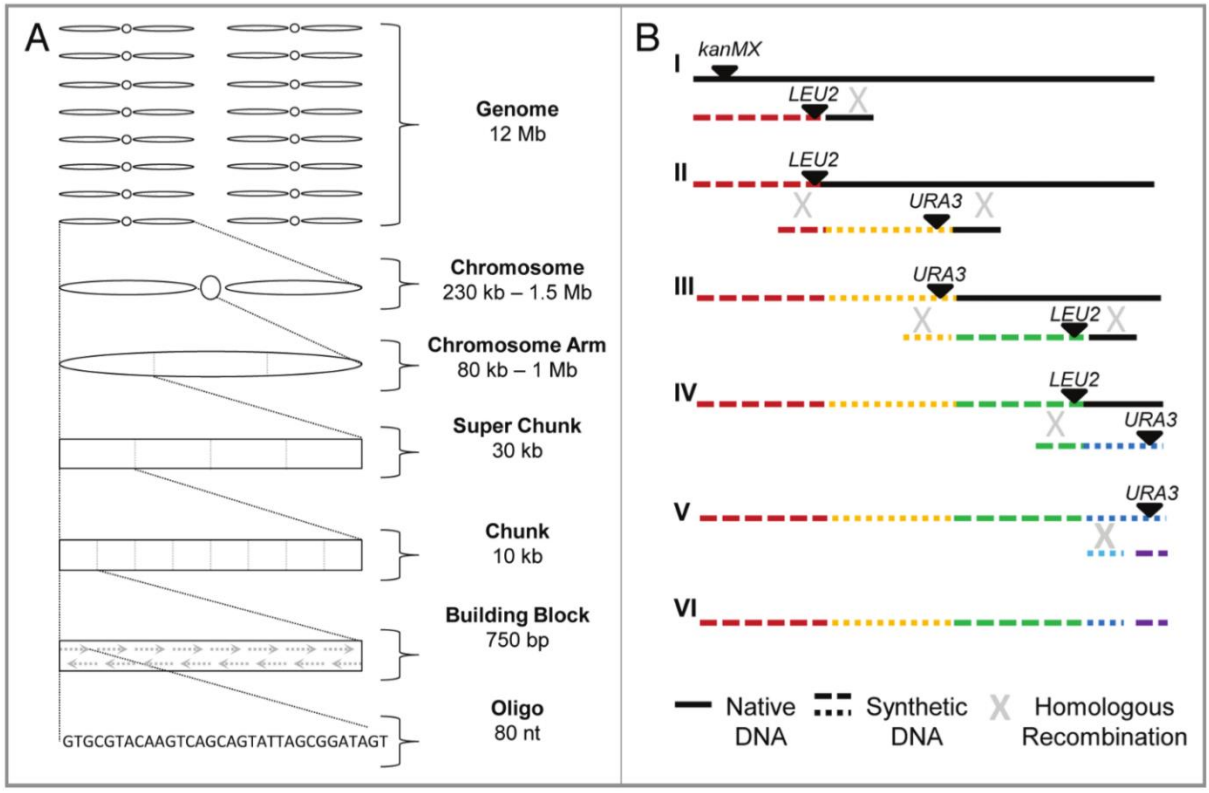


Figure 1-5. Construction of Sc2.0. (A) The yeast genome is subdivided into increasingly smaller units to facilitate assembly of the synthetic genome. The construction can start from different units. (B) The native DNA is replaced with synthetic DNA gradually by iterative rounds of homologous recombination. The native chromosome with *kanMX* marker (I) is targeted for integration of synthetic DNA with *LEU2* marker. Synthetic fragments gradually integrate and replace the native DNA by iterative rounds of homologous recombination, with alternated markers (*LEU2* or *URA3*) (II-IV). The final *URA3* marker is replaced by synthetic sequence without a marker and selected on 5-FOA to generate a complete synthetic chromosome (V). Reproduced from Dymond et al., 2012, Bioengineered Bugs.

In 2013, the first synthetic eukaryotic chromosome Sc synIII was completed (80). Followed by this success, the synthesis of another five synthetic chromosomes have been accomplished so far (81-85). The synthetic chromosomes are functional and don't cause visible growth defects.

During the construction, a 'design-build-test-learn' cycle was applied. A number of genetic changes resulted in the decreased fitness ('bugs') in the yeast strain. Some of these changes came from undesired mutations that arose during construction, and some were derived from the genome modifications designed *in silico*. A big challenge in the process of synthesis is de-bugging. Pooled PCRTag mapping (PoPM) (84) has been developed to map the region that decreases the fitness. In their study, through backcrossing (forming a diploid that then undergoes meiosis) a wild-type strain with the synthetic strain with a defect, the resulting progeny that carry mixtures of synthetic and wild-type DNA were divided into a normal growth pool and growth defect pool under a selective condition. The causal region was narrowed down after PCRTag analysis of two growth pools. Other approaches have been implemented for de-bugging including sequencing and 'omics' techniques (81). The function of corresponding genes towards cell fitness was then characterized and the genetic modifications were reverted to restore the growth of the synthetic strain.

In addition, deep knowledge about genome structure have been gained along the reconstruction of the synthetic genome. For example, the native ribosomal DNA (rDNA) on synthetic chromosomal XII was removed (85). From this study, the engineered rDNA with the ITS region swapped-in from the sequences of *Saccharomyces bayanus*, either expressed on a shuffling plasmid or on a relocated chromosomal region, generating a fully functional rDNA. It can be concluded that the ITS region is changeable and the rDNA region in the genome is flexible, supporting current theories of neutral drift in this genomic region. The circularised synthetic chromosome V was constructed in synV strain with no decreased fitness or stability (83). With all the modifications embedded in the synthetic chromosome V, the ring synV provides a model for studies of yeast genome rearrangement, ring chromosome evolution and human ring chromosome disorders (83).

The strains containing single synthetic chromosomes have been consolidated into a single strain (81). Mitchell's study indicated the triple synthetic strain (with synIII, synVI and synIXR) exhibited comparable fitness on yeast extract peptone dextrose (YPD) and yeast extract peptone

glycerol (YPG) but had decreased fitness on synthetic complete medium probably due to the deletion of twenty tRNAs. Through transcriptional and proteomic analysis, the triple synthetic strain didn't show major changes, which demonstrates it is likely feasible to build the ultimate strain with all the sixteen synthetic chromosomes (81).

1.4.3 SCRaMbLE - A whole genome rearrangement system

To facilitate the genome flexibility, a whole genome rearrangement system termed Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution (SCRaMbLE) has been implemented in Sc2.0 (77)**.** All non-essential genes and major landmarks have been designed to be flanked by the symmetrical loxPsym sites in the synthetic genome (Figure 1-6 A). Upon the activation of the site-specific Cre recombinase, recombination events between loxPsym sites can happen in either direction (86, 87). Consequently, genome rearrangement events occur including gene deletions, duplications, inversions and translocations (79, 88). The system is powerful to generate cells with diverse genome and various phenotypes (Figure 1-6 B).

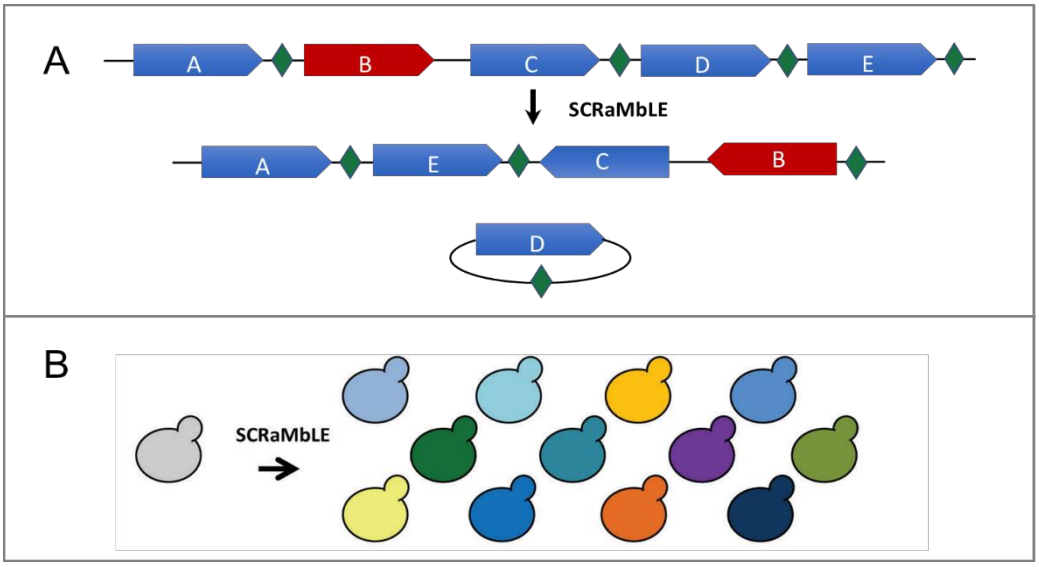


Figure 1-6. SCRaMbLE generates genome diversity. (A) LoxPsym sites (green diamonds) are inserted in the 3'UTR of all non-essential genes (blue arrows), but not in essential genes (red arrow). Upon induction of SCRaMbLE, various genome rearrangement events can happen: 'B' has been inverted and translocated, 'C' has been inverted, 'E' has been translocated, and 'D' has been deleted. (B) Induction of SCRaMbLE in a synthetic strain (gray) results in cells with diverse genome and phenotypes (colours). Reproduced from Dymond et al., 2012, Bioengineered Bugs.

The system can be used to gain more knowledge about genome structure, genome rearrangement and genome evolution. In addition, by iterative rounds of SCRaMbLE, redundant genes and unstable elements can be looped-out, and yeast minimal chromosomes and minimal genomes can in theory be obtained. Most importantly, the SCRaMbLE platform has great potential for industrial applications.

1.4.3.1 Control systems for SCRaMbLE

The synthetic chromosomes carrying dozens to hundreds of loxPsym sites (78) are supposed to be stable until a certain time point, when rearrangement events are triggered upon induction. Thus, a tightly regulated system is needed to control SCRaMbLE. A Cre-EBD variant is created by fusing the engineered Cre with the murine estrogen binding domain (EBD), which facilitates translocation of the protein from the cytosol to the nucleus upon binding with estrogen (86). Their study shows this Cre-EBD variant has low basal activity and is induced upon the addition of the estrogen. Combined with a daughter cell specific promoter pSCW1, the system should be tightly regulated via expression from the pSCW11-Cre-EBD plasmid, which theoretically only produces a pulse of recombinase activity once in each cell's lifetime (86). However, in previous studies, growth defects have been observed without the addition of estrogen (77, 88). This suggests the possibility of leaky expression in pSCW11Cre-EBD controlled SCRaMbLE. In more recent studies, an AND gate switch was developed for the precise control of *Cre* (89). In this model, the Cre-EBD is regulated by a galactose-inducible *pGAL1*. Thus, the activation of SCRaMbLE requires simultaneous presence of estrogen and galactose in the growth medium. Compared with the pSCW11Cre-EBD, this system gave very low basal expression level without induction and exhibited a stronger and more constant signal after the induction of SCRaMbLE (89). A light inducible system also has been applied to control SCRaMbLE upon exposure to red light (90). The extent of recombination is adjustable by the induction time and the concentration of the chromophore phycocyanobilin (PCB).

1.4.3.2 Application

SCRaMbLE has demonstrated to be powerful to generate genome diversity. SCRaMbLE of synthetic chromosome arm synIXR has shown that, from deep sequencing of 64 SCRaMbLEd strains, the recombinations only happened among the loxPsym sites, and no ectopic recombinations or other rearrangements were detected in the nonsynthetic genome (88). Shen's study showed every SCRaMbLEd genome was unique.

As a tool for creating diverse rearranged genomes, SCRaMbLE can serve as a method for generating novel genetic diversity as input for screening of strains with favourable industrial phenotypes. For example, SCRaMbLE can be used to improve strain tolerance under different stress conditions. With a reporter of SCRaMbLEd cells, strains with increased tolerance to ethanol, acetic acid and heat have been isolated after SCRaMbLE of synXII (91). SCRaMbLE was also performed on a *S. cerevisiae* haploid and a heterozygous diploid mated with *Saccharomyces paradoxus* to evolve heat and caffeine tolerant strains (92). Moreover, SCRaMbLE is a powerful tool to optimise metabolic pathways to improve substrate utilisation or increase the yield of different products. As an attractive cell factory, *S. cerevisiae* has been engineered with numerous heterologous pathways. However, heterologous pathway engineering is always a problematic issue due to the limited understanding of complex biological systems. Traditionally, a metabolic pathway optimised by overexpressing one critical gene or a few genes in the pathway or manipulating the genes that catalyse the reactions of the precursors, and deleting genes that encode enzymes which compete for carbon flux or cofactors. With the induction of SCRaMbLE, diverse genome rearranged strains will be generated with global changes of gene content, gene direction and gene locations within a short time. SCRaMbLE of the xylose oxidoreductase pathway in a synV strain dramatically increased its growth on xylose, which is an attractive alternative carbon source, abundant in lignocellulosic hydrolysates (93). Heterologous synthesis pathways of carotene and carotenoids have been expressed in semi-synthetic yeast strains. Yields of carotene and carotenoids have been improved dramatically through SCRaMbLE in several studies (89, 90, 94). The pigment

producing violacein biosynthesis pathway from *Chromobacterium* was expressed in a synV strain (93). After SCRaMbLE, dark colonies were selected since the dark colour correlates with more violacein pigment production. A SCRaMbLEd strain producing 2.3-times more violacein was obtained (93).

The SCRaMbLEd mutants were subjected to further analysis to identify the genetic changes that are responsible for the phenotype. PCRTag analysis, backcrossing, long read sequencing and other omics analysis have been used to reveal the genotype-phenotype correlations (89, 91, 92, 95). The causal genetic mutations have revealed novel principles for the engineering of metabolic pathways. Currently, most of the pathways that have been chosen for SCRaMbLE to improve the biosynthesis are the ones indicating colour changes, such as carotene and carotenoids, since the change of the yield can be directly estimated visually, which expedites and simplifies the screening process. The potential of SCRaMbLE to improve the yield of many industrially relevant bio-products is yet to be fully explored.

1.5 Microbial PA production

Organic acids are small molecular compounds that have one or more carboxylic acid groups (96). They are building block chemicals, widely used in many industries. Currently, organic acids are derived from petroleum, which is not a sustainable resource and also pollutes the environment. As many organic acids naturally exist as metabolites in microorganisms, microbial organic acid production is an attractive alternative to petroleum production. For instance, lactic acid, as an important chemical intermediate and the precursor for renewable polymers, has been produced in large quantities by lactic acid bacteria (97, 98). However, in the process of bacterial fermentation for organic acid production, a large amount of neutralizing agents, such as $CaCO_3$ or base, need to be added to maintain neutral pH, due to the pH sensitivity of these bacterial producers (99, 100). This also leads to the production of organic acid in salt form. Thus, additional recovery steps are required for the purification of protonated organic acid, which increases the cost of downstream processing, and hinders further

development of these bacterial production processes (99-101). Alternatively, yeast is an attractive cell factory to produce organic acids as it is robust, can grow at relatively high fermentation temperature and is easily genetically engineered. Specifically, yeast is tolerant to low pH and can therefore produce organic acids in their protonated forms, saving downstream purification costs. It has been engineered for the heterologous production of multiple organic acids, such as lactic acid (102), succinic acid (103), muconic acid (104), para-hydroxybenzoic acid (PHBA) (105) and 3-hydroxypriopionic acid (3HP) (106).

Propionic acid (PA) is a three-carbon organic acid. It has been widely used as a preservative for human and animal food (107). PA also serves as a chemical intermediate in plastic, pharmaceutical, cosmetic, paint, and herbicide industries (107-110). The U.S. Department of Energy recognises PA as one of the top 30 'platform' chemicals (111). Global production of PA in 2014 was 400 kilotons with USD 1.07 billion in revenue and was predicted to reach USD 1.55 billion by 2020 (112). However, PA is currently still produced from finite oil reserves. Microbial fermentation of PA is gaining strong attention as an alternative to petroleum refining.

1.5.1 Microbial PA production

Propionibacteria, including *Propionibacterium thoenii*, *Propionibacterium freudenreichii*, *Propionibacterium shermanii*, *Propionibacterium acidipropionici*, and *Propionibacterium beijingense*, are natural PA producers (113). The reactions involved in PA biosynthesis was revealed in 1960 and the pathway is known as the Wood-Werkman cycle (107, 114). As shown in Figure 1-7, PA biosynthesis can be divided into two steps: first, pyruvate is converted to succinate through (S)-methylmalonyl-CoA-pyruvate transcarboxylase, malate dehydrogenase, fumarase and fumarate reductase; second, succinate is dissimilated into PA through CoA transferase, (R)-methylmalonyl-CoA mutase, methylmalonyl-CoA racemase and (S)-methylmalonyl-CoA-pyruvate transcarboxylase (107, 113). Microbial PA production through propionibacteria has drawn much attention. Different carbon sources (115, 116), culture conditions (117, 118), new culture methods such as fed-batch culture (119), extractive

fermentation (120), and cell-immobilised fermentation (121) have already been applied in many studies to improve PA production. Despite these efforts, PA production through propionibacteria still can't compete with the petroleum-derived method. Metabolic engineering and synthetic biology approaches have the potential to further improve PA production. However, little progress has been made through metabolic engineering of propionibateria owing to their high GC content, limited knowledge of their metabolism (122, 123) and few tools for genetic manipulation (124). In addition, propionibacteria commonly develop antibiotic resistance spontaneously and have a low transformation efficiency (125, 126). All these disadvantages have forced researchers to consider using model microorganisms for PA production.

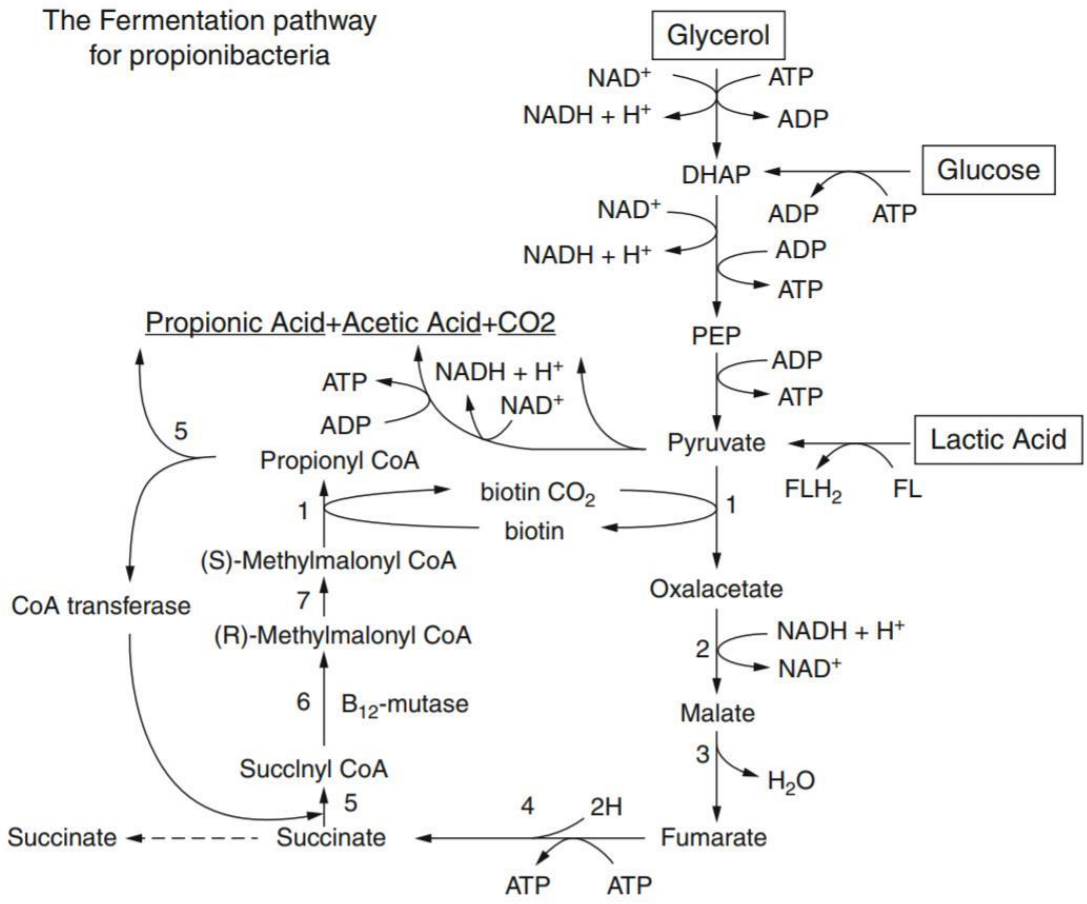


Figure 1-7. The succinate-propionate pathway in propionibacteria. 1 (S)-methylmalonyl-CoA-pyruvate transcarboxylase, 2 malate dehydrogenase, 3 fumarase, 4 fumarate reductase, 5 CoA transferase, 6 (R)-methylmalonyl-CoA mutase, and 7 methylmalonyl-CoA racemase. Reproduced from Zidwick et al., 2013, The Prokaryotes.

As a favourable cell factory for producing organic acids, yeast is a promising alternative for heterologous PA production. A synthetic Wood-Werkman cycle has been redesigned for expression in yeast (Williams et al., unpublished data) by our group. Based on native yeast metabolism, only four enzymes in the original Wood-Werkman cycle were required for construction of a functional PA pathway. The following enzymes were designed in the synthetic pathway, the Vitamin B_{12}-dependent methyl-malonyl-CoA mutase (encoded by *mutA* and *mutB*) from *Saccharopolyspora erythraea*, the methyl malonyl-CoA epimerase (*mce*), the methyl malonyl-CoA carboxytransferase (*mtcA, mtcB, mtcC, mtcD*), and the propionyl-CoA succinyl-CoA transferase (*pst*) from *P. acidipropionici* (Figure 1-8). An *Escherichia coli* (*E. coli*) propionyl-CoA: succinate-CoA transferase (*scpC*) was also co-expressed to enhance propionate production. All the codons were optimised for expression in yeast. A strong constitutive promoter was used to express each gene constitutively at a high level. In total, five strong promoters and six terminators were used in order to keep the chance of homologous recombination between repeated elements as low as possible. In addition, all the genes in the synthetic pathway were flanked by the symmetry loxPsym sites, which enable SCRaMbLing the pathway in the synthetic genome upon induction.

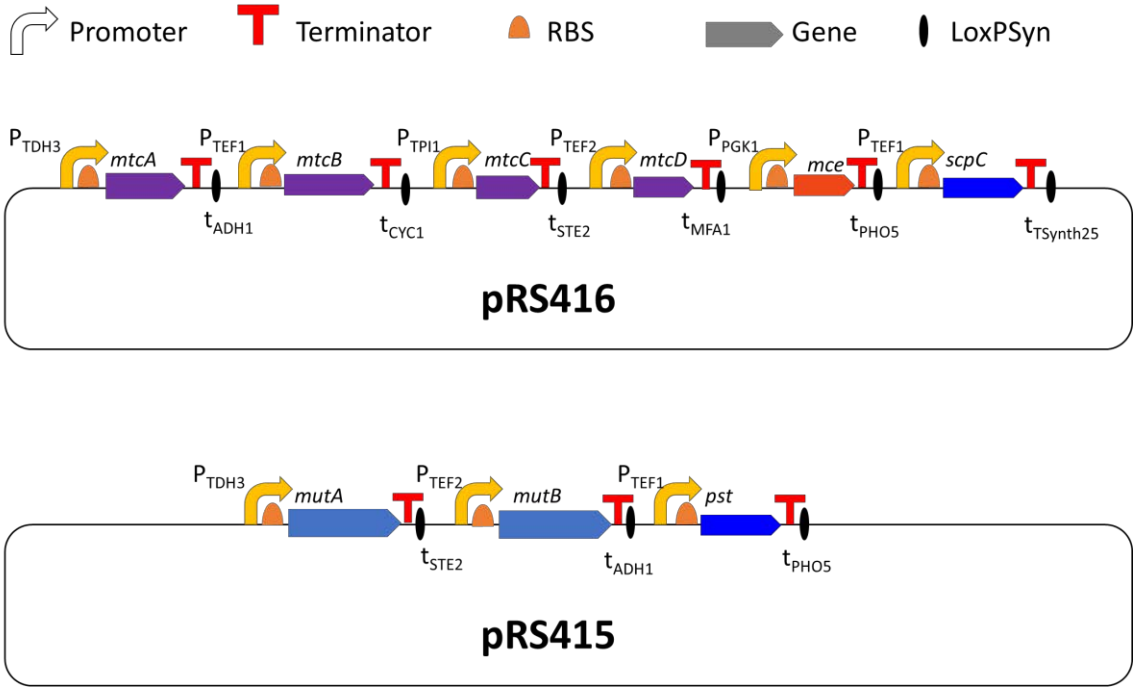


Figure 1-8. Synthetic Wood-Werkman cycle genes and regulatory elements. The genes of the Wood-Werkman cycle are depicted with their promoters, terminators, and flanking loxPsym sites. The genes were distributed over two yeast centromeric vectors with auxotrophic *URA3* (pRS416) and *LEU2* (pRS415) selective markers respectively. Reprinted from Williams et al., unpublished data.

One problem for expressing this pathway in yeast is that methyl-malonyl-CoA mutase requires vitamin-B_{12} as a co-factor. As yeast is not able to produce Vitamin-B_{12} itself, it needs to be supplemented with Vitamin-B_{12} during fermentation (Williams et al., unpublished data). From their study, when fermentation was performed anaerobically in serum bottles, the BY4741 strain expressing the synthetic Wood-Werkman cycle accumulated up to 0.048 mM PA. In aerobic bioreactor fermentation at pH 3.5 and pH 6, the PA yield reached 1.1 mM and 0.85 mM, respectively.

Although the PA yield from yeast fermentation is far from reaching industrial scale, it could possibly be enhanced by induction of SCRaMbLE system. Through SCRaMbLing the synthetic Wood-Werkman cycle with the synthetic chromosomes from Sc2.0, diverse genome rearrangements would happen, possibly generating the ones with improved PA production.

1.5.2 Microbial organic acid tolerance

One big obstacle of microbial organic acid production is that organic acids are generally toxic to many of the microorganisms. When the extracellular pH is below the dissociation constant (pKa), organic acids are in their protonated (non-dissociated) form and can therefore diffuse freely across the cell membrane. After entering the near neutral cytosol, they dissociate into the protons and the anions. The protons can build up and acidify the cytosol. This influences many cellular processes, including decreasing DNA and RNA synthesis rates, disrupting internal pH homeostasis, and affecting lipid organization and the function of cellular membranes (127-130). In addition, the accumulation of the anions is also toxic to the cells.

Extensive studies have been performed on *S. cerevisiae* to investigate the weak acid toxicity/tolerance mechanism. The acetic acid that is released during hydrolysis of lignocellulosic feedstocks and the subsequent toxicity of acetic acid is a big challenge for bioethanol production using lignocellulosic feedstocks (131). Transcriptome analysis and genome-wide screening have been performed to study the response of yeast to acetic acid (132, 133). Formic acid, benzoic acid and PA are widely used as food preservatives. Yeast genes that confer tolerance or sensitivity to these acids were also explored by screening of the yeast gene deletion library. However, the genome-wide screening of the deletion library and the transcriptome identified numerous genes and biological processes, making it hard to identify the critical genes and the underlying mechanisms. In addition, they were both performed using only a short exposure to weak acids, which does not replicate the organic acid toxicity that would be encountered in an industrial setting. Furthermore, the genome-wide deletion library was screened on solid medium, which is also significantly different to an industrial setting, which involves serial transfers in liquid medium.

ALE is a powerful approach to improve a strain's tolerance under a specific stress condition. It has already been applied to generate yeast strains tolerant to acetic acid (134), 3-HP (26), and lactic acid (39). In these studies, starting with a relatively moderate concentration, ALE

gradually evolved strains with stable and long-term acid tolerance. The phenotype-genotype relationships were then probed using omics techniques and functional analysis.

Yeast's base capacity for PA production has already been demonstrated (Williams et al., unpublished data). However, PA is highly toxic to yeast, which would hinder the yield improvement in future, as very high product-titers are typically required for any industrial fermentation to be economically viable. ALE is an attractive approach to generate yeast strains with long-term tolerance to PA and reveal novel engineering principles for microbial PA production combined with sequencing and reverse-engineering techniques.

1.6 Scope of thesis

The overall aim of this project is to establish novel tools for the generation and selection of genetic diversity to improve PA production and PA tolerance. This thesis is therefore focused on the integration of synthetic biology and high-throughput screening principles to explore alternatives to traditional metabolic engineering for the production of metabolites such as PA. This study will provide insights into the utilisation of the synthetic yeast genome and SCRaMbLE system, and provide novel design and engineering principles related to the use of these cutting-edge and relatively unexplored techniques.

Specific aims:

Aim 1: To contribute to the development of the world's first synthetic eukaryotic genome (Sc2.0 project) through the construction and de-bugging of synthetic chromosomes XIV. The synthetic genome evolution system (SCRaMbLE) implemented in Sc2.0 has the potential to be applied to improve yeast heterologous production, such as PA production (Chapter 2).

Aim 2: To establish a biosensor that responds to PA and can be used for high-throughput screening of PA producers (Chapter 3).

Aim 3: To utilise the synthetic genome evolution system (SCRaMbLE) in conjunction with biosensor-mediated high-throughput screening to isolate yeast with improved PA production, and provide refined methods for this approach (Chapter 4).

Aim 4: To improve the PA tolerance in yeast using ALE, and decipher the underlying mechanism of PA tolerance, which will reduce the PA toxicity problem in terms of PA production, and pave the way for further improvement of PA yield (Chapter 5).

1.7 Reference

1. Goffeau A, Barrell BG, Bussey H, Davis R, Dujon B, Feldmann H, et al. Life with 6000 genes. Science. 1996;274(5287):546-67.

2. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nature genetics. 1999;21(1s):33.

3. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science. 1997;278(5338):680-6.

4. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. science. 2004;303(5659):808-13.

5. Ross-Macdonald P, Sheehan A, Friddle C, Roeder GS, Snyder M. [29] Transposon mutagenesis for the analysis of protein production, function, and localization. Methods in enzymology. 303: Elsevier; 1999. p. 512-32.

6. Martzen MR, McCraith SM, Spinelli SL, Torres FM, Fields S, Grayhack EJ, et al. A biochemical genomics approach for identifying genes by the activity of their products. Science. 1999;286(5442):1153-5.

7. Hizicky EM, Martzen MR, McCraith SM, Spinelli SL, Xing F, Hull NP, et al. Biochemical genomics approach to map activities to genes activities to gene. Methods in Enzymology. 350: Elsevier; 2002. p. 546-59.

8. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, et al. A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. Nature. 2000;403(6770):623.

9. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences. 2001;98(8):4569-74.

10. Medina VG, Almering MJ, van Maris AJ, Pronk JT. Elimination of glycerol production in anaerobic cultures of a Saccharomyces cerevisiae strain engineered to use acetic acid as an electron acceptor. Applied and environmental microbiology. 2010;76(1):190-5.

11. Hong K-K, Nielsen J. Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. Cellular and Molecular Life Sciences. 2012;69(16):2671-90.

12. Yu KO, Jung J, Kim SW, Park CH, Han SO. Synthesis of FAEEs from glycerol in engineered Saccharomyces cerevisiae using endogenously produced ethanol by heterologous expression of an unspecific bacterial acyltransferase. Biotechnology and bioengineering. 2012;109(1):110-5.

13. Zhao L, Wang J, Zhou J, Liu L, Du G, Chen J. Modification of carbon flux in Sacchromyces cerevisiae to improve L-lactic acid production. Wei sheng wu xue bao= Acta microbiologica Sinica. 2011;51(1):50-8.

14. Raab AM, Gebhardt G, Bolotina N, Weuster-Botz D, Lang C. Metabolic engineering of Saccharomyces cerevisiae for the biotechnological production of succinic acid. Metabolic engineering. 2010;12(6):518-25.

15. Lee W, DaSilva NA. Application of sequential integration for metabolic engineering of 1, 2-propanediol production in yeast. Metabolic engineering. 2006;8(1):58-65.

16. Vai M, Brambilla L, Orlandi I, Rota N, Ranzi BM, Alberghina L, et al. Improved secretion of native human insulin-like growth factor 1 from gas1 mutant Saccharomyces cerevisiae cells. Applied and environmental microbiology. 2000;66(12):5477-9.

17. Gellerfors P, Axelsson K, Helander A, Johansson S, Kenne L, Lindqvist S, et al. Isolation and characterization of a glycosylated form of human insulin-like growth factor I produced in Saccharomyces cerevisiae. Journal of Biological Chemistry. 1989;264(19):11444-9.

18. Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature. 2006;440(7086):940.

19. Kim E-J, Park Y-K, Lim H-K, Park Y-C, Seo J-H. Expression of hepatitis B surface antigen S domain in recombinant Saccharomyces cerevisiae using GAL1 promoter. Journal of biotechnology. 2009;141(3-4):155-9.

20. Becker JV, Armstrong GO, van der Merwe MJ, Lambrechts MG, Vivier MA, Pretorius IS. Metabolic engineering of Saccharomyces cerevisiae for the synthesis of the wine-related antioxidant resveratrol. FEMS Yeast Research. 2003;4(1):79-85.

21. Bailey JE. Toward a science of metabolic engineering. Science. 1991;252(5013):1668-75.

22. Bailey JE, Sburlati A, Hatzimanikatis V, Lee K, Renner WA, Tsai PS. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. Biotechnology and Bioengineering. 1996;52(1):109-21.

23. Çakar ZP, Turanlı-Yıldız B, Alkım C, Yılmaz Ü. Evolutionary engineering of Saccharomyces cerevisiae for improved industrially important properties. FEMS yeast research. 2012;12(2):171-82.

24. Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. Proceedings of the National Academy of Sciences. 1991;88(16):7160-4.

25. Williams TC, Pretorius IS, Paulsen IT. Synthetic evolution of metabolic productivity using biosensors. Trends in biotechnology. 2016;34(5):371-81.

26. Kildegaard KR, Hallström BM, Blicher TH, Sonnenschein N, Jensen NB, Sherstyk S, et al. Evolution reveals a glutathione-dependent mechanism of 3-hydroxypropionic acid tolerance. Metabolic engineering. 2014;26:57-66.

27. Dragosits M, Mattanovich D. Adaptive laboratory evolution–principles and applications for biotechnology. Microbial cell factories. 2013;12(1):64.

28. LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, et al. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. Applied and environmental microbiology. 2015;81(1):17-30.

29. Lee D-H, Palsson BØ. Adaptive evolution of Escherichia coli K-12 MG1655 during growth on a Nonnative carbon source, L-1, 2-propanediol. Applied and environmental microbiology. 2010;76(13):4158-68.

30. Sonderegger M, Jeppsson M, Larsson C, Gorwa-Grauslund MF, Boles E, Olsson L, et al. Fermentation performance of engineered and evolved xylose-fermenting Saccharomyces cerevisiae strains. Biotechnology and bioengineering. 2004;87(1):90-8.

31. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, et al. In silico design and adaptive evolution of Escherichia coli for production of lactic acid. Biotechnology and bioengineering. 2005;91(5):643-8.

32. Çakar ZP, Seker UO, Tamerler C, Sonderegger M, Sauer U. Evolutionary engineering of multiple-stress resistant Saccharomyces cerevisiae. FEMS yeast research. 2005;5(6-7):569-78.

33. Leavitt JM, Wagner JM, Tu CC, Tong A, Liu Y, Alper HS. Biosensor-Enabled Directed Evolution to Improve Muconic Acid Production in Saccharomyces cerevisiae. Biotechnology journal. 2017;12(10):1600687.

34. Ferea TL, Botstein D, Brown PO, Rosenzweig RF. Systematic changes in gene expression patterns following adaptive evolution in yeast. Proceedings of the National Academy of Sciences. 1999;96(17):9721-6.

35. Sonderegger M, Sauer U. Evolutionary engineering of Saccharomyces cerevisiae for anaerobic growth on xylose. Applied and environmental microbiology. 2003;69(4):1990-8.

36. Wisselink HW, Toirkens MJ, Wu Q, Pronk JT, van Maris AJ. Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered Saccharomyces cerevisiae strains. Applied and environmental microbiology. 2009;75(4):907-14.

37. Zhou H, Cheng J-s, Wang BL, Fink GR, Stephanopoulos G. Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by Saccharomyces cerevisiae. Metabolic engineering. 2012;14(6):611-22.

38. Dhar R, Sägesser R, Weikert C, Yuan J, Wagner A. Adaptation of Saccharomyces cerevisiae to saline stress through laboratory evolution. Journal of evolutionary biology. 2011;24(5):1135-53.

39. Fletcher E, Feizi A, Bisschops MM, Hallström BM, Khoomrung S, Siewers V, et al. Evolutionary engineering reveals divergent paths when yeast is adapted to different acidic environments. Metabolic engineering. 2017;39:19-28.

40. Stanley D, Fraser S, Chambers PJ, Rogers P, Stanley GA. Generation and characterisation of stable ethanol-tolerant mutants of Saccharomyces cerevisiae. Journal of industrial microbiology & biotechnology. 2010;37(2):139-49.

41. Reyes LH, Gomez JM, Kao KC. Improving carotenoids production in yeast via adaptive laboratory evolution. Metabolic engineering. 2014;21:26-33.

42. Purnick PE, Weiss R. The second wave of synthetic biology: from modules to systems. Nature reviews Molecular cell biology. 2009;10(6):410.

43. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. A modular cloning system for standardized assembly of multigene constructs. PloS one. 2011;6(2):e16765.

44. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison III CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods. 2009;6(5):343.

45. Kok Sd, Stanton LH, Slaby T, Durot M, Holmes VF, Patel KG, et al. Rapid and reliable DNA assembly via ligase cycling reaction. ACS synthetic biology. 2014;3(2):97-106.

46. Quan J, Tian J. Circular polymerase extension cloning of complex gene libraries and pathways. PloS one. 2009;4(7):e6441.

47. Muller H, Annaluru N, Schwerzmann JW, Richardson SM, Dymond JS, Cooper EM, et al. Assembling large DNA segments in yeast. Gene Synthesis: Springer; 2012. p. 133-50.

48. Rabinovitch-Deere CA, Oliver JW, Rodriguez GM, Atsumi S. Synthetic biology and metabolic engineering approaches to produce biofuels. Chemical reviews. 2013;113(7):4611-32.

49. Jarboe LR, Zhang X, Wang X, Moore JC, Shanmugam K, Ingram LO. Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. BioMed Research International. 2010;2010.

50. Khalil AS, Collins JJ. Synthetic biology: applications come of age. Nature Reviews Genetics. 2010;11(5):367.

51. Peng B, Williams TC, Henry M, Nielsen LK, Vickers CE. Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. Microbial cell factories. 2015;14(1):91.

52. Redden H, Morse N, Alper HS. The synthetic biology toolbox for tuning gene expression in yeast. FEMS Yeast Res. 2015;15(1):1-10.

53. Tyo KE, Nevoigt E, Stephanopoulos G. Directed evolution of promoters and tandem gene arrays for customizing RNA synthesis rates and regulation. Methods in enzymology. 2011;497:135-55.

54. Yamanishi M, Ito Y, Kintaka R, Imamura C, Katahira S, Ikeuchi A, et al. A genome-wide activity assessment of terminator regions in Saccharomyces cerevisiae provides a " terminatome "toolbox. ACS synthetic biology. 2013;2(6):337-47.

55. Curran KA, Karim AS, Gupta A, Alper HS. Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metabolic engineering. 2013;19:88-97.

56. Beerli RR, Barbas III CF. Engineering polydactyl zinc-finger transcription factors. Nature biotechnology. 2002;20(2):135.

57. Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO. Zif268 protein–DNA complex refined at 1.6 Å: a model system for understanding zinc finger–DNA interactions. Structure. 1996;4(10):1171-80.

58. Jamieson AC, Wang H, Kim S-H. A zinc finger directory for high-affinity DNA recognition. Proceedings of the National Academy of Sciences. 1996;93(23):12834-9.

59. Beerli RR, Segal DJ, Dreier B, Barbas CF. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proceedings of the National Academy of Sciences. 1998;95(25):14628-33.

60. Park K-S, Lee D-k, Lee H, Lee Y, Jang Y-S, Kim YH, et al. Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. Nature biotechnology. 2003;21(10):1208.

61. Liu D, Evans T, Zhang F. Applications and advances of metabolite biosensors for metabolic engineering. Metabolic engineering. 2015;31:35-43.

62. DiCarlo JE, Conley AJ, Penttilä M, Jäntti J, Wang HH, Church GM. Yeast oligo-mediated genome engineering (YOGE). ACS synthetic biology. 2013;2(12):741-9.

63. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010;327(5962):167-70.

64. Feng Z, Zhang B, Ding W, Liu X, Yang D-L, Wei P, et al. Efficient genome editing in plants using a CRISPR/Cas system. Cell research. 2013;23(10):1229.

65. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology. 2013;31(3):227.

66. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823-6.

67. Bortesi L, Fischer R. The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnology advances. 2015;33(1):41-52.

68. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic acids research. 2013;41(7):4336-43.

69. Bao Z, Xiao H, Liang J, Zhang L, Xiong X, Sun N, et al. Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae. ACS synthetic biology. 2014;4(5):585-94.

70. Shi S, Liang Y, Zhang MM, Ang EL, Zhao H. A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in Saccharomyces cerevisiae. Metabolic engineering. 2016;33:19-27.

71. Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new engineering rules for an emerging discipline. Molecular systems biology. 2006;2(1).

72. Cello J, Paul AV, Wimmer E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. science. 2002;297(5583):1016-8.

73. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. science. 2010:1190719.

74. Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, et al. Design and synthesis of a minimal bacterial genome. Science. 2016;351(6280):aad6253.

75. Foury F. Human genetic diseases: a cross-talk between man and yeast. Gene. 1997;195(1):1-10.

76. Kannan K, Gibson DG. Yeast genome, by design. Science. 2017;355(6329):1024-5.

77. Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature. 2011;477(7365):471.

78. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, et al. Design of a synthetic yeast genome. Science. 2017;355(6329):1040-4.

79. Dymond J, Boeke J. The Saccharomyces cerevisiae SCRaMbLE system and genome minimization. Bioengineered. 2012;3(3):170-3.

80. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, et al. Total synthesis of a functional designer eukaryotic chromosome. science. 2014;344(6179):55-8.

81. Mitchell LA, Wang A, Stracquadanio G, Kuang Z, Wang X, Yang K, et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. Science. 2017;355(6329):eaaf4831.

82. Shen Y, Wang Y, Chen T, Gao F, Gong J, Abramczyk D, et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. Science. 2017;355(6329):eaaf4791.

83. Xie Z-X, Li B-Z, Mitchell LA, Wu Y, Qi X, Jin Z, et al. "Perfect" designer chromosome V and behavior of a ring derivative. Science. 2017;355(6329):eaaf4704.

84. Wu Y, Li B-Z, Zhao M, Mitchell LA, Xie Z-X, Lin Q-H, et al. Bug mapping and fitness testing of chemically synthesized chromosome X. Science. 2017;355(6329):eaaf4706.

85. Zhang W, Zhao G, Luo Z, Lin Y, Wang L, Guo Y, et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. Science. 2017;355(6329):eaaf3981.

86. Lindstrom DL, Gottschling DE. The mother enrichment program: a genetic system for facile replicative life span analysis in Saccharomyces cerevisiae. Genetics. 2009.

87. Hoess RH, Wierzbicki A, Abremski K. The role of the loxP spacer region in PI site-specific recombination. Nucleic acids research. 1986;14(5):2287-300.

88. Shen Y, Stracquadanio G, Wang Y, Yang K, Mitchell LA, Xue Y, et al. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. Genome research. 2016;26(1):36-49.

89. Jia B, Wu Y, Li B-Z, Mitchell LA, Liu H, Pan S, et al. Precise control of SCRaMbLE in synthetic haploid and diploid yeast. Nature communications. 2018;9(1):1933.

90. Hochrein L, Mitchell LA, Schulz K, Messerschmidt K, Mueller-Roeber B. L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. Nature communications. 2018;9(1):1931.

91. Luo Z, Wang L, Wang Y, Zhang W, Guo Y, Shen Y, et al. Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. Nature communications. 2018;9(1):1930.

92. Shen M, Wu Y, Yang K, Li Y, Xu H. Heterozygous diploid and interspecies SCRaMbLEing Nat. Nature communications; 2018.

93. Blount B, Gowers GF, Ho J, Ledesma-Amaro R, Jovicevic D, McKiernan R, et al. Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome. Nature communications. 2018;9(1):1932.

94. Wu Y, Zhu R-Y, Mitchell LA, Ma L, Liu R, Zhao M, et al. In vitro DNA SCRaMbLE. Nature communications. 2018;9(1):1935.

95. Liu W, Luo Z, Wang Y, Pham NT, Tuck L, Pérez-Pi I, et al. Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods. Nature communications. 2018;9(1):1936.

96. Yin X, Li J, Shin H-d, Du G, Liu L, Chen J. Metabolic engineering in the biotechnological production of organic acids in the tricarboxylic acid cycle of microorganisms: advances and prospects. Biotechnology advances. 2015;33(6):830-41.

97. Abdel-Rahman MA, Tashiro Y, Sonomoto K. Lactic acid production from lignocellulose-derived sugars using lactic acid bacteria: overview and limits. Journal of biotechnology. 2011;156(4):286-301.

98. Ghaffar T, Irshad M, Anwar Z, Aqil T, Zulifqar Z, Tariq A, et al. Recent trends in lactic acid biotechnology: a brief review on production to purification. Journal of radiation research and applied Sciences. 2014;7(2):222-9.

99. Joglekar H, Rahman I, Babu S, Kulkarni B, Joshi A. Comparative assessment of downstream processing options for lactic acid. Separation and purification technology. 2006;52(1):1-17.

100. Cheng K-K, Zhao X-B, Zeng J, Wu R-C, Xu Y-Z, Liu D-H, et al. Downstream processing of biotechnological produced succinic acid. Applied microbiology and biotechnology. 2012;95(4):841-50.

101. Werpy T, Petersen G. Top value added chemicals from biomass: volume I--results of screening for potential candidates from sugars and synthesis gas. National Renewable Energy Lab., Golden, CO (US); 2004.

102. Ishida N, Saitoh S, Tokuhiro K, Nagamori E, Matsuyama T, Kitamoto K, et al. Efficient production of L-lactic acid by metabolically engineered Saccharomyces cerevisiae with a genome-integrated L-lactate dehydrogenase gene. Applied and environmental microbiology. 2005;71(4):1964-70.

103. Otero JM, Cimini D, Patil KR, Poulsen SG, Olsson L, Nielsen J. Industrial systems biology of Saccharomyces cerevisiae enables novel succinic acid cell factory. PloS one. 2013;8(1):e54144.

104. Curran KA, Leavitt JM, Karim AS, Alper HS. Metabolic engineering of muconic acid production in Saccharomyces cerevisiae. Metabolic engineering. 2013;15:55-66.

105. Williams T, Averesch N, Winter G, Plan M, Vickers C, Nielsen L, et al. Quorum-sensing linked RNA interference for dynamic metabolic pathway control in Saccharomyces cerevisiae. Metabolic engineering. 2015;29:124-34.

106. Borodina I, Kildegaard KR, Jensen NB, Blicher TH, Maury J, Sherstyk S, et al. Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via β-alanine. Metabolic engineering. 2015;27:57-64.

107. Zidwick MJ, Chen J-S, Rogers P. Organic acid and solvent production: propionic and butyric acids and ethanol. The Prokaryotes: Springer; 2013. p. 135-67.

108. Wemmenhove E, van Valenberg HJ, Zwietering MH, van Hooijdonk TC, Wells-Bennik MH. Minimal inhibitory concentrations of undissociated lactic, acetic, citric and propionic acid for Listeria monocytogenes under conditions relevant to cheese. Food microbiology. 2016;58:63-7.

109. Álvarez-Chávez CR, Edwards S, Moure-Eraso R, Geiser K. Sustainability of bio-based plastics: general comparative analysis and recommendations for improvement. Journal of Cleaner Production. 2012;23(1):47-56.

110. Hebert RF. Stable indole-3-propionate salts of S-adenosyl-L-methionine. Google Patents; 2017.

111. Vörös A, Horváth B, Hunyadkürti J, McDowell A, Barnard E, Patrick S, et al. Complete genome sequences of three Propionibacterium acnes isolates from the type IA2 cluster. Journal of bacteriology. 2012;194(6):1621-2.

112. Propionic Acid Market for Animal Feed & Grain Preservatives, Calcium & Sodium Propionates, Cellulose Acetate Propionate and Other Applications: Global Industry Perspective, Comprehensive Analysis, Size, Share, Growth, Segment, Trends and Forecast, 2014 - 2020 [Available from: https://www.marketresearchstore.com/report/propionic-acid-market-for-animal-feed-grain-z39993.

113. Liu L, Zhu Y, Li J, Wang M, Lee P, Du G, et al. Microbial production of propionic acid from propionibacteria: current state, challenges and perspectives. Critical reviews in biotechnology. 2012;32(4):374-81.

114. Swick RW, Wood HG. The role of transcarboxylation in propionic acid fermentation. Proceedings of the National Academy of Sciences. 1960;46(1):28-41.

115. Himmi E, Bories A, Boussaid A, Hassani L. Propionic acid fermentation of glycerol and glucose by Propionibacterium acidipropionici and Propionibacterium freudenreichii ssp. shermanii. Applied Microbiology and Biotechnology. 2000;53(4):435-40.

116. Quesada-Chanto A, Wagner F. Microbial production of propionic acid and vitamin B 12 using molasses or sugar. Applied microbiology and biotechnology. 1994;41(4):378-83.

117. Quesada-Chanto A, Schmid-Meyer A, Schroeder A, Carvalho-Jonas M, Blanco I, Jonas R. Effect of oxygen supply on biomass, organic acids and vitamin B12 production by

Propionibacterium shermanii. World Journal of Microbiology and Biotechnology. 1998;14(6):843-6.

118. Koussémon M, Combet-Blanc Y, Ollivier B. Glucose fermentation by Propionibacterium microaerophilum: effect of pH on metabolism and bioenergetic. Current microbiology. 2003;46(2):0141-5.

119. Zhu Y, Li J, Tan M, Liu L, Jiang L, Sun J, et al. Optimization and scale-up of propionic acid production by propionic acid-tolerant Propionibacterium acidipropionici with glycerol as the carbon source. Bioresource technology. 2010;101(22):8902-6.

120. Jin Z, Yang ST. Extractive fermentation for enhanced propionic acid production from lactose by Propionibacterium acidipropionici. Biotechnology progress. 1998;14(3):457-65.

121. Zhu L, Wei P, Cai J, Zhu X, Wang Z, Huang L, et al. Improving the productivity of propionic acid with FBB-immobilized cells of an adapted acid-tolerant Propionibacterium acidipropionici. Bioresource technology. 2012;112:248-53.

122. Falentin H, Deutsch S-M, Jan G, Loux V, Thierry A, Parayre S, et al. The complete genome of Propionibacterium freudenreichii CIRM-BIA1T, a hardy actinobacterium with food and probiotic applications. PloS one. 2010;5(7):e11748.

123. Parizzi LP, Grassi MCB, Llerena LA, Carazzolle MF, Queiroz VL, Lunardi I, et al. The genome sequence of Propionibacterium acidipropionici provides insights into its biotechnological and industrial potential. BMC genomics. 2012;13(1):562.

124. Zhuge X, Liu L, Shin H-d, Chen RR, Li J, Du G, et al. Development of a Propionibacterium-Escherichia coli shuttle vector as a useful tool for metabolic engineering of Propionibacterium jensenii, an efficient producer of propionic acid. Applied and environmental microbiology. 2013:AEM. 00737-13.

125. Kiatpapan P, Murooka Y. Genetic manipulation system in propionibacteria. Journal of bioscience and bioengineering. 2002;93(1):1-8.

126. Gonzalez-Garcia RA, McCubbin T, Navone L, Stowers C, Nielsen LK, Marcellin E. Microbial Propionic Acid Production. Fermentation. 2017;3(2):21.

127. Fernandes A, Mira N, Vargas R, Canelhas I, Sá-Correia I. Saccharomyces cerevisiae adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. Biochemical and biophysical research communications. 2005;337(1):95-103.

128. Piper P, Mahe Y, Thompson S, Pandjaitan R, Holyoak C, Egner R, et al. The Pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast. The EMBO journal. 1998;17(15):4257-65.

129. Stratford M, Anslow P. Comparison of the inhibitory action on Saccharomyces cerevisiae of weak-acid preservatives, uncouplers, and medium-chain fatty acids. FEMS microbiology letters. 1996;142(1):53-8.

130. Raja N, Goodson M, Smith D, Rowbury R. Decreased DNA damage by acid and increased repair of acid-damaged DNA in acid-habituated Escherichia coli. Journal of applied bacteriology. 1991;70(6):507-11.

131. Jönsson LJ, Alriksson B, Nilvebrant N-O. Bioconversion of lignocellulose: inhibitors and detoxification. Biotechnology for biofuels. 2013;6(1):16.

132. Mira NP, Palma M, Guerreiro JF, Sá-Correia I. Genome-wide identification of Saccharomyces cerevisiae genes required for tolerance to acetic acid. Microbial cell factories. 2010;9(1):79.

133. Li B-Z, Yuan Y-J. Transcriptome shifts in response to furfural and acetic acid in Saccharomyces cerevisiae. Applied microbiology and biotechnology. 2010;86(6):1915-24.

134. González-Ramos D, de Vries ARG, Grijseels SS, Berkum MC, Swinnen S, Broek M, et al. A new laboratory evolution approach to select for constitutive acetic acid tolerance in

Saccharomyces cerevisiae and identification of causal mutations. Biotechnology for biofuels. 2016;9(1):173.

Chapter 2 : Synthesis and laboratory evolution of *Saccharomyces cerevisiae* synthetic chromosome XIV

Synthesis and laboratory evolution of *Saccharomyces cerevisiae* synthetic chromosome XIV

Thomas C. Williams [1,2], Heinrich Kroukamp [1], Xin Xu [1], Elizabeth Wightman [1], Hugh D. Goold [1,3], Alexander Carpenter [1,2], Briardo Llorente [1,2], Niel Van Wyk[1], Monica Espinosa [1,2], Elizabeth Daniel [1], Helena Nevalainen [1], Natalie Curach [1], Anthony R. Borneman [4], Joel Bader [5], Leslie A. Mitchell [6], Jef D. Boeke [6], Daniel Johnson [4], Isak S. Pretorius [1], Ian T. Paulsen [1 *]

[1] Department of Molecular Sciences, Macquarie University, NSW, Australia

[2] CSIRO Synthetic Biology Future Science Platform, Canberra, ACT 2601, Australia

[3] New South Wales Department of Primary Industries, Locked Bag 21, Orange, NSW 2800, Australia

[4] The Australian Wine Research Institute, PO Box 197, Adelaide, SA 5064, Australia

[5] High Throughput Biology Center and Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205, USA.

[6] Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, New York University (NYU) Langone Medical Center, New York City, NY 10016, USA.

*Corresponding author: Ian T. Paulsen

E-mail address: ian.paulsen@mq.edu.au

Abstract

The Yeast 2.0 Synthetic Biology Project uses the yeast *Saccharomyces cerevisiae* as the base for a synthetic genome. The major aim of this international study is to gain knowledge about the fundamental properties of chromosomes by synthetically constructing them. This emerging field allows for researchers to explore and eventually unlock the potential for understanding the biological basis of eukaryotic genes. This project marks a milestone in the advancement of synthetic genomics and metabolic engineering. As part of the Yeast 2.0 consortium, Macquarie University is building chromosome XIV (synXIV). Unlike previous synthetic genome projects, the Yeast 2.0 genome contains a number of substantial alterations such as deletion of repetitive DNA elements, relocation of tRNA genes, use of synthetic telomeres, and recoding of a stop codon. The synthetic yeast genome also includes an inducible evolution system termed SCRaMbLE (Synthetic Chromosome Recombination and Modification by LoxP Meditated Evolution) that will enable the formation of distinct *S. cerevisiae* genomes that vary in gene content and genomic architecture. PCR-tag screening, fitness tests and whole-genome sequencing revealed several errors in synXIV, which led to growth defects when compared to the native strain. Redesign of the chromosome construction, back crossing and adaptive laboratory evolution (ALE) were utilised to amend perturbations in the strain's growth.

Introduction

The field of synthetic genomics encompasses the design, construction, and characterisation of whole genomes. This new approach to genomics provides several unique opportunities. For example, the ability to make global genetic changes that are too numerous to implement in a step-wise manner, the capacity to discover new biological knowledge through the classic 'design-build-test-learn' cycle of synthetic biology, and the potential to design genomes that encode superior industrial phenotypes [1,2] are all enabled by synthetic genomics. The synthetic yeast genome project 'Sc2.0' exemplifies these new possibilities via genome streamlining (removal of transposons and non-essential introns), genome 'defragmentation' via the

relocation of all transfer RNA genes to a separate neo-chromosome, telomere standardisation, and through modularisation of each non-essential promoter-gene cassette with flanking loxPsym recombination sites. The 12 Mbp *Saccharomyces cerevisiae* genome is comprised of sixteen chromosomes, which are being built by an international consortium adhering to central design principles [3]. The Sc2.0 consortium has already completed six synthetic yeast chromosomes, resulting in new fundamental biological knowledge and genome construction technology. For example, novel growth-defect 'debugging' [4] and chromosome consolidation [5] techniques have been developed, knowledge of in-depth phenotypic characterisation of designed chromosomes [6,7] and the degree of genome plasticity with regard to ribosomal gene clusters have been gained [8], and the effects of chromosome re-design on the three-dimensional genomic architecture have been explored [9].

The most significant design feature incorporated in Sc2.0 is an inducible evolution system termed SCRaMbLE (synthetic chromosome recombination and modification by LoxP mediated evolution). Induction of a heterologous Cre-recombinase enzyme results in inversions, duplications, translocations, and deletions of gene(s) between LoxP sites [9-14]. The induction of SCRaMbLE can in theory generate an incalculable number of genomes with unique gene content and genomic architecture [14], making it an extremely powerful tool for generating genetic diversity prior to laboratory evolution, and for understanding the genomic basis of selected phenotypes [15,16].

Results and discussion

SynXIV design and construction

S. cerevisiae synXIV was redesigned according to Sc2.0 principles using the BioStudio software package [3]. Briefly, 256 loxPsym sites were inserted 3 bp after the stop codons of non-essential genes, 14 introns were removed, ORFs were synonymously recoded to contain a total of 1040 PCR-tags, 90 stop codons were swapped to the 'TAA' variant to free-up the TGA codon, native telomeres were replaced with a standardised synthetic version, and all transposon

and tRNA sequences were removed. These changes resulted in a 753,097 bp synXIV, representing a 4 % size reduction of the native 784,333 bp version. SynXIV was segregated into twenty-two 'megachunks' (labelled A to X) ranging in length between 27 and 39 kb, each comprising four or five 5-10 kb 'chunks' (labelled A1, A2 etc.). Each chunk was designed with 40 bp of homology to each adjacent chunk to facilitate yeast homologous recombination based assembly [17]. These homologous regions also contained restriction enzyme sites so that chunks could be ligated together *in vitro* prior to transformation. The last chunk of each megachunk contained either a *LEU2* or *URA3* selectable marker gene for leucine or uracil biosynthesis respectively. The successful integration of a megachunk would over-write the marker left by the previous megachunk, enabling both positive and negative selection of megachunk transformants as previously described with the 'SWAP-In' method [3]. Colonies able to grow on medium selective for the incoming megachunk and not on medium selective for the previous megachunk were then screened for correct homologous recombination of synthetic megachunk DNA using high-throughput real-time PCR with primers specific to the synonymously recoded PCR tags in synthetic ORFs. Colonies with amplification of all synthetic primer pairs, and no amplification with wild-type equivalents were deemed to have successfully integrated a megachunk.

To facilitate rapid construction, the megachunks comprising synXIV were integrated in two separate strains in parallel. Strain 1 contained megachunks A to G while Strain 2 contained megachunks G to X (Figure 2-1). As a eukaryote, *S. cerevisiae* can undergo meiosis, meaning that a diploid cross of strains 1 and 2 would create haploid progeny with random recombination events between the two versions of partially synthetic synXIV. To enable the identification of haploid progeny that were more likely to have undergone meiotic recombination within the shared megachunk G region, selective markers were placed within the wild-type regions of partially synthetic chromosomes (Figure 2-1A). This meant that sporulated haploid colonies could be selected for further analysis based on an inability to grow on media selective for the markers in the wild-type chromosome segments. Haploid colonies that were unable to grow on

all selective media after random spore isolation were screened for the presence of each synthetic megachunk using PCR-tag analysis. After the first attempt, no isolates were found to contain an entirely synthetic version of chromosome XIV, as meiotic crossover was on average too frequent throughout the tested colonies. Two isolates that had a greater total amount of synthetic DNA, a larger 'overlap' for crossover, and were of opposite mating type (Figure 2-1B) were therefore used in a second cross to form a strain containing a complete, contiguous version of synXIV.

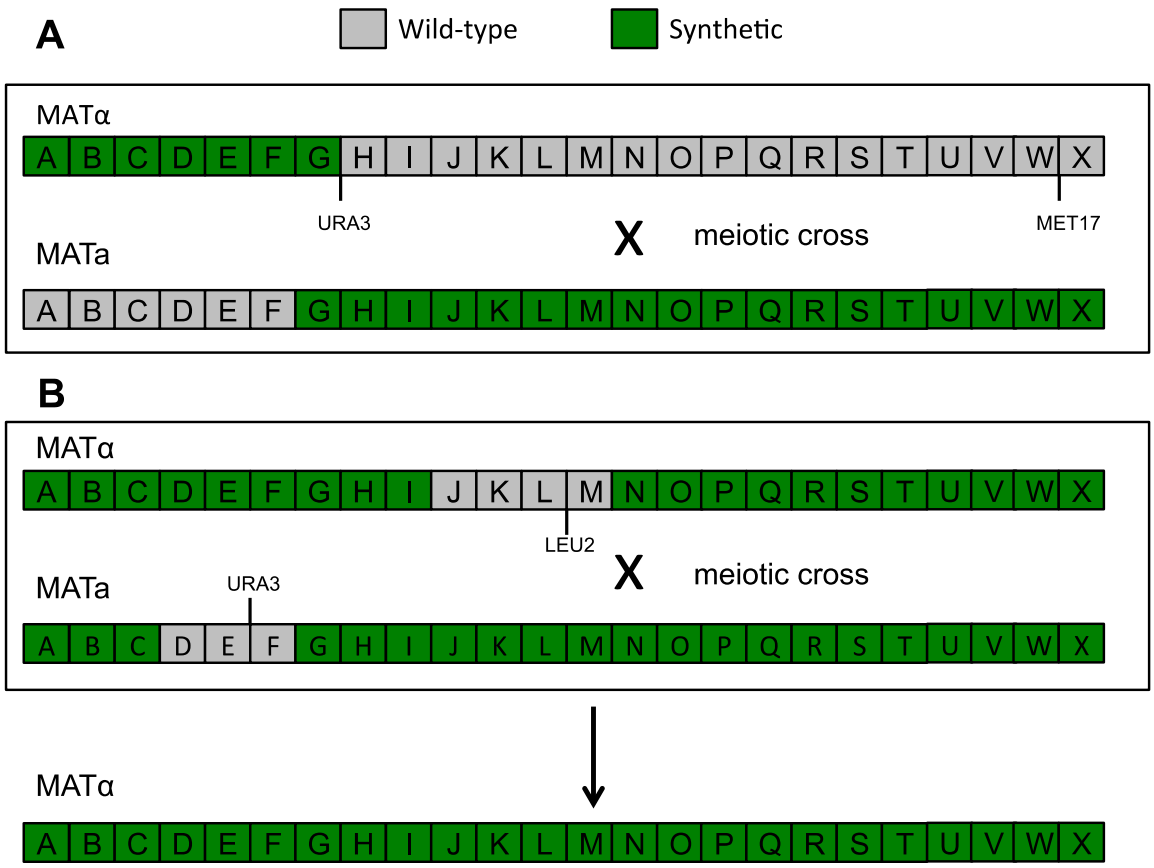


Figure 2-1. Parallel construction and strain crossing strategy for synXIV. (A) Synthetic megachunks A to X were integrated across two separate strains in parallel, with *URA3* and *MET17* markers integrated at the *YNL223W* and *YNR063W* wild-type loci of the synthetic A-G strain respectively, prior to mating with a synthetic G-X strain. (B) Haploid strains resulting from the first cross were crossed together with screening for *LEU2* (*YNL125C* locus) and *URA3* marker (*YNL269W* locus) loss from wild-type regions to give a fully synthetic version of chromosome XIV.

SynXIV was initially designed for construction in four different strains overlapping by one megachunk (A-G, G-M, M-T, and T-X). However, problems encountered during this process led to the abandonment of two intermediate strains and the subsequent construction of synXIV in two separate strains overlapping at megachunk G (Figure 2-1A). The introduction of megachunk P (strain synXIV M-P) took over ten attempts. This, combined with the fact that the one positive colony appeared after seven days of incubation, suggested that this strain had a growth-defect. The majority of megachunk P integrant colonies that were screened had positive PCR-tag amplification for both synthetic and wild-type primers. Whole genome re-sequencing and read-depth analysis of this strain revealed aneuploidy of several chromosomes. This aneuploidy may explain the difficulty in integrating megachunk P, as genome instability and multiple chromosome copies would reduce the likelihood of correct megachunk integration and wild-type DNA replacement. This strain was discontinued and megachunks N-P were integrated in the existing G-M strain. In another example of unexpected chromosomal structural rearrangement, whole-genome sequencing of the T-X strain revealed that a region spanning chunks U4 to V3 was inserted approximately ten times. This strain was therefore abandoned and megachunks Q-X were integrated in the existing G-P strain. Only due to frequent whole-genome sequencing were these chromosomal rearrangements detected and corrected. This work demonstrates that as large-scale synthetic biology projects with many sequential genetic modifications become more common, it will become pertinent to view whole-genome sequencing as an essential quality control. During this QC process we also identified several sequence discrepancies such as missing LoxP sites, missing PCR-tags, and non-swapped stop codons. These discrepancies and their repair are detailed Table S2-1 and Table S2-2.

The synXIV strain comprising of megachunks G-X incorporated a growth-defect after the integration of the first megachunk, which was retained in all subsequent strains. However, when megachunk G was re-integrated in a wild-type BY4741 strain, the growth defect was not repeatable. Whole-genome re-sequencing of the original megachunk G integrant strain revealed

that chunks G1 and G2 had been integrated approximately four times, as indicated by read depth relative to surrounding chromosomal loci. When a descendant of this strain that had megachunks G-O integrated was also sequenced, the multiple insertions were no longer present, and had presumably been looped out of the genome via homologous recombination. However, this strain and all subsequent megachunk integration strains retained a severe growth defect. To ascertain the cause of this problem, backcrossing and pooled fast/slow strain sequencing was carried out as previously described [4]. A synthetic chromosome region spanning megachunks J to L was found to have low-coverage in 'fast-grower' reads. Conversely, a wild-type chromosome region from megachunks H to L had low-coverage in the pooled 'slow-grower' reads, suggesting that a synXIV modification within megachunks G to L was the cause of the defect.

An independent line of enquiry also indicated that the megachunk G to L region was the cause of a major growth-defect, and further narrowed the location down to chunk J1. During the final meiotic crossing of partially synthetic strains to produce a fully synthetic version of chromosome XIV (Figure 2-1B), two near-complete strains were identified that had wild-type I-J and J regions respectively. These strains had improved fitness relative to two strains with fully-synthetic versions of chromosome XIV, suggesting that the cause of a growth-defect lay within the megachunk I-J region, supporting the back-crossing and pooled sequencing analysis. Integration of megachunk I in one of these strains (SynXIV.29) did not cause any growth-defect, indicating that the defect lay outside of megachunk I. When synthetic chunk J1 was then introduced, the severe growth defect on YP-glycerol was re-established (Figure 2-2A). Subsequent integration of the wild-type J1 region did not restore normal growth, initially leading us to dismiss this region as the cause of the growth-defect. Fortuitously, during the integration of synthetic chunk J1, several strains were identified as having correct integration according to PCR-tag analysis, and one of these strains was found to not have the growth-defect (Figure 2-2A). Whole genome sequencing of slow and fast-growing versions of the J1 integrants revealed the fast-growing isolate (J1.8) was missing a single loxPsym site 3′ of the

MRPL19 gene, whereas in the slow-growing isolate (J1.4), this loxPsym was present. The *MRPL19* gene encodes a mitochondrial ribosomal protein, and deletion of this gene causes a respiratory growth defect [18]. Further analysis of the re-sequenced genomes showed that the slow-growing isolate had no reads mapping to the yeast mitochondrial reference genome, while the fast-growing isolate did. Loss of mitochondrial DNA is consistent with the fact that re-integration of wild-type chunk J1 did not restore growth, as yeast cannot de-novo regenerate the mitochondrial genome once it has been lost [19]. It is also consistent with the complete lack of respiratory growth on YP-glycerol seen from this defect (Figure 2-2).

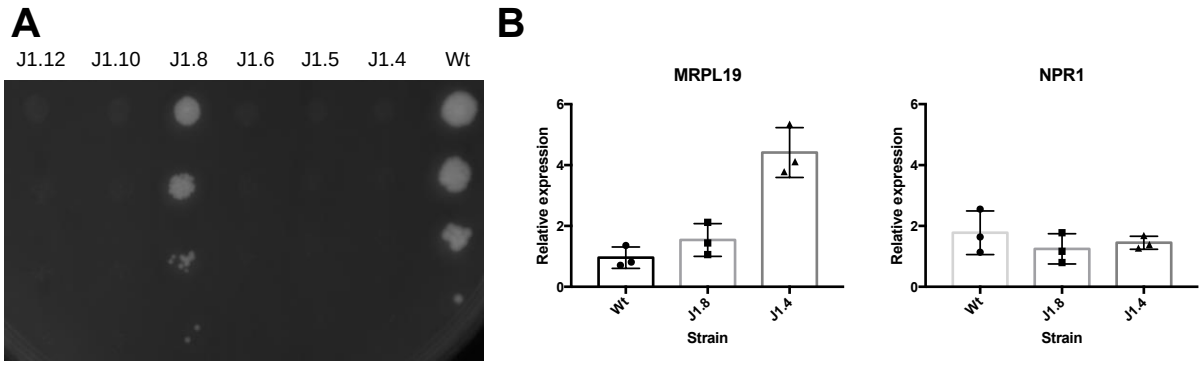


Figure 2-2. Chunk J1 growth-defect analysis. (A) YP-glycerol fitness test of chunk J1 integrants and the Wt (BY4741) strain. 3 µL aliquots of 10-fold serial dilutions of exponentially growing liquid cultures were spotted onto agar plates and incubated at 30 °C for 5-days. (B) RT-qPCR of the *MRPL19* and *NPR1* genes was carried out on cDNA from BY4741 (Wt), repaired synXIV (J1.8), and growth-defect synXIV (J1.4) strains. Expression was normalised to the *ALG9* gene using the modified-Livak method as previously described [20]. Bars and error bars represent mean and standard deviation from three biological replicates. Individual expression values of replicates are also shown.

We hypothesised that the presence of a loxPsym site in the 3′ UTR of *MRPL19* could modulate transcriptional termination efficiency, and performed RT-qPCR on RNA extracted from exponentially growing wild-type, J1.4, and J1.8 strains. Interestingly, *MRPL19* transcript levels were significantly up-regulated by approximately 5-fold in the slow-growing J1.4 isolate, but were not significantly different between the wild-type and fast-growing J1.8 isolate (Figure 2-2B). The mRNA levels of the nearby *YNL184C* and *NPR1* ORFs were also measured to determine if the *MRPL19* 3′ UTR loxPsym had any effects on their transcription. No mRNA was detected from the *YNL184C* ORF in any of the strains, while *NPR1* expression levels were

not significantly different between the strains. To test whether the observed up-regulation of *MRPL19* mRNA could be the cause of the growth defect in the slow-growing J1.4 strain, the native *MRPL19* gene and terminator were over-expressed from the strong-constitutive *TDH3* promoter [21] in the wild-type strain from the pRS413 vector. *MRPL19* over-expression did not cause a growth defect, suggesting the mechanism of the growth defect is not related to *MRPL19* over-expression.

Like all nuclear-encoded mitochondrial proteins, the *MRPL19* protein is targeted to the mitochondria. Consistent with this, the *MRPL19* protein sequence includes a predicted mitochondrial targeting peptide signal in the N-terminus [22]. In addition to protein targeting signals, many nuclear encoded mitochondrial genes have motifs in the 3′ UTR of their mRNA that facilitate localisation to the outside of mitochondria via recognition by the *PUF3* protein for co-translation and import [23]. The *MRPL19* mRNA has a PUF3p recognition motif, and when *PUF3* is deleted there is no *MRPL19* mRNA localisation to the mitochondria [23]. The addition of a loxPsym site in the 3′ UTR of *MRPL19* may therefore interfere with mitochondrial mRNA targeting, leading to the observed growth defect. To test this hypothesis, we designed a GFP fusion protein that retained the entire *MRPL19* coding sequence in order to account for the possibility that other RNA or protein signals are important for mitochondrial protein import (Figure 2-3). Versions with the native 3′ UTR and with the loxPsym containing 3′ UTR were synthesised and tested for the import of GFP into yeast mitochondria (Figure 2-3). Confocal microscopy of yeast cells with stained mitochondria (mitotracker) showed that the cytosol-localised control (‘Free GFP’) had not GFP signal correlation with mitotracker signal, while the native *MRPL19-GFP* construct (‘Native’) resulted in successful co-localisation of GFP with the mitochondria (Figure 2-3). In contrast, the insertion of a loxPsym motif in the 3′ UTR of *MRPL19* appeared to interfere with mitochondrial localisation and import, as GFP signal was commonly clustered around mitochondria (Figure 2-3). We therefore conclude the loxPsym motif in the 3′ UTR of *MRPL19* interferes with correct mitochondrial import of this protein, leading to the observed growth defect under respiratory conditions (Figure 2-2).

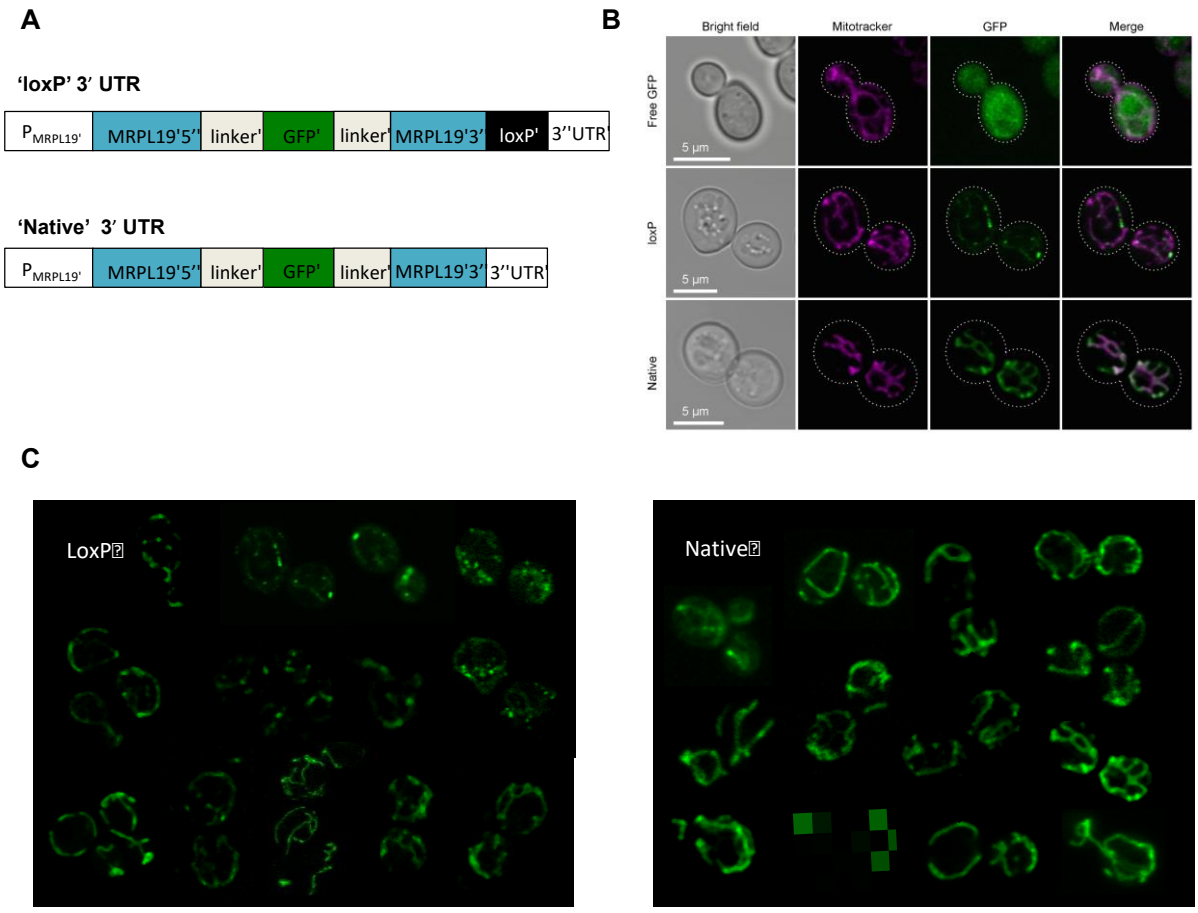


Figure 2-3. Mitochondrial import of MRPL19-GFP fusion proteins with native and loxPsym containing 3′ UTR. (A) Two synthetic *MRPL19* promoter-gene-3′UTR constructs were designed with a super folder GFP expressed from the middle of the native ORF, separated by peptide linkers. One version contained a loxPsym motif 3 bp after the stop codon (termed 'loxP') while the second version contained no loxP within the native 3′UTR (termed 'Native'). (B) BY4741 strains expressing either of these two constructs, or a cytosol localised GFP (termed 'Free GFP' [24]) were grown in the presence of 100 nM Mitotracker Red (ThermoFisher) to stain active mitochondria. (C) Multiple cells from different triplicate populations of both 'loxP' and 'Native' populations with mitochondrial localised GFP signals. An Olympus FV 1000 confocal microscope was used to visualise yeast cells with bright field, mitotracker and GFP signals.

Although the removal of the *MRPL19* 3′ UTR loxPsym dramatically improved growth on both YPD and YP-glycerol, there was still a growth-defect on YP-glycerol medium at 30 °C (Figure 2-2A). To repair and understand the cause of this defect, the growth-defect containing J1.8 strain was subjected to adaptive laboratory evolution (ALE) [25] on liquid YP-glycerol medium in triplicate for approximately 90 generations by passaging into fresh medium every 24 hours (Figure 2-4). The wild-type BY4741 strain was also evolved in parallel to enable identification of mutations that enhance glycerol utilisation, mutations related to general adaptation to YP media, and for any 'hitch-hiker' mutations occurring due to genetic drift. Mutations that

enhance glycerol utilisation were important to control for, given that the laboratory *S. cerevisiae* strains grow poorly on glycerol and ALE of superior glycerol utilisation phenotypes has recently been reported [26]. Initially, each passage was inoculated at an OD_{600} of 0.1, but inoculation was decreased to 0.05 after 24 generations (vertical dashed line, Figure 2-4) to enable faster accrual of generations. OD_{600} was measured after each 24 h period to serve as a proxy for fitness (Figure 2-4). The J1.8 strain showed a ~50 % improvement in final OD_{600} on YP-glycerol medium after 90 generations, while the wild-type control strain showed a 38 % improvement after 120 generations. Fitness testing of the wild-type strain (wt), parental synXIV strain (J1.8), and a mixed population from one of the J1.8 evolutionary lineages (J1.8e) revealed that growth on YP-glycerol was restored to wild-type levels, but that the adaptive mutations had not been acquired by all cells in the J1.8e population, as determined by colony size heterogeneity (Figure 2-4). One of the final mixed populations was therefore plated onto YPG medium to identify fast-growing isolates. One of these isolates (J1.8e1) was tested for fitness (Figure 2-4), whole-genome sequenced, and used for further analysis.

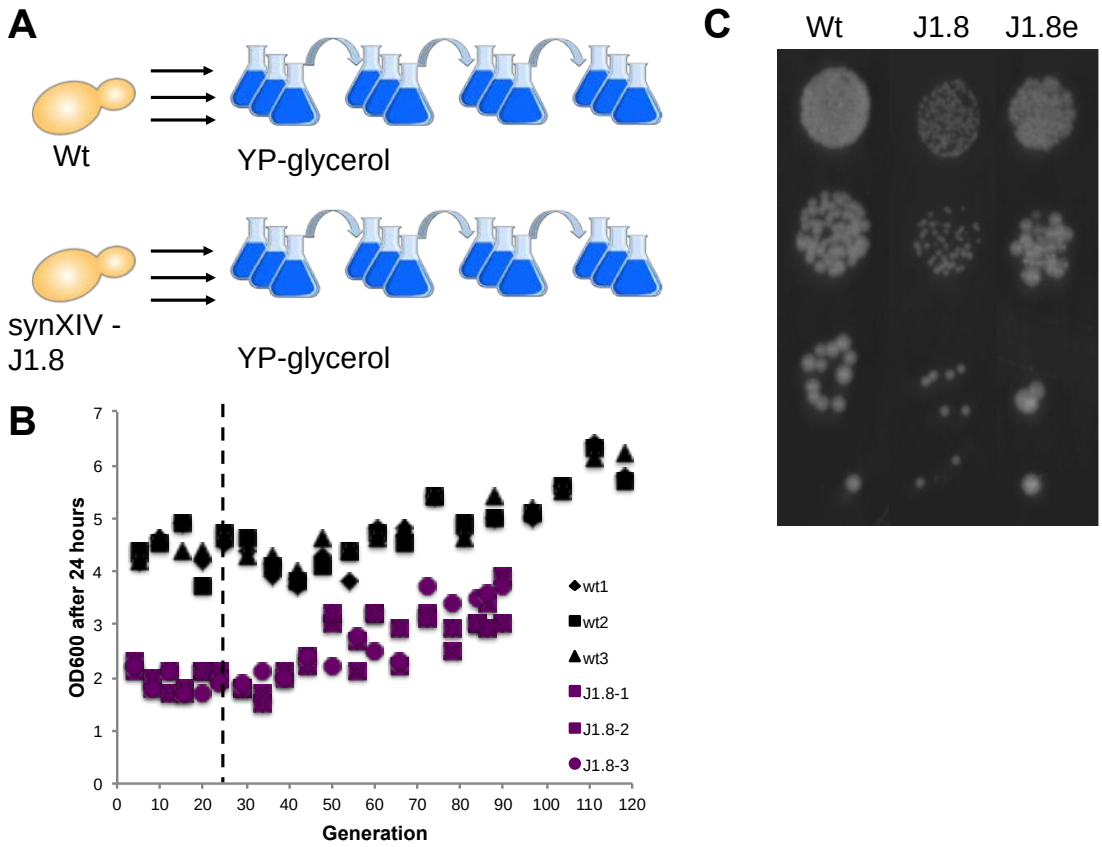


Figure 2-4. Adaptive laboratory evolution of synXIV and Wt strains on YP-glycerol medium. (A) BY4741 (Wt) and synXIV strains (J1.8) were grown in YP-glycerol medium with passaging to fresh medium every 24-hours. (B) OD_{600nm} was measured after each 24 h passage to track any improvements in final biomass and therefore fitness. After approximately 24 generations the inoculation OD_{600nm} was changed from 0.1 to 0.05 to increase the rate of generation accrual per passage. (C) At the end of the evolution experiment the fitness of the parental wild-type and J1.8 strains were compared with one of the evolved lineages of J1.8 (J1.8e) on YP-glycerol at 30 °C.

Whole-genome sequencing of the isolates and final evolved mixed-populations was carried out to compare mutations that might have caused the initial growth-defect during the construction of synXIV. While no point mutations were detected anywhere in the genomes of the evolved lineages, they did have a higher relative copy number of the ribosomal DNA repeat located on chromosome XII. Evolved lineages had approximately eight more rDNA copies compared to the parental J1.8 strain, and the total estimated rDNA copy number of the evolved lineages was more similar to the original strain that was used for megachunk integration (*lap3::KanMX* from the knockout collection). This suggested that higher rDNA copy-number was more conducive with fitness, and more similar to the wild-type state. The copy-number of the ribosomal DNA

locus is known to be highly plastic and easily modulated by cells [27,28]. It is therefore possible that the rDNA copy number is important for normal growth in the context of synthetic chromosome XIV and its construction history. Interestingly, it has been found that exposure of yeast to lithium acetate during DNA transformation can lead to alteration in rDNA copy-number [29]. Given that the synXIV strain underwent approximately 30 transformations in its life-history, it is possible that these repeated lithium acetate exposures promoted rDNA copy-number change. To test if increased rDNA copy-number was the cause of the evolved phenotype, the rDNA locus was cloned and expressed from the low copy centromeric pRS413 vector in the parental J1.8 strain. This construct had no effect on growth, suggesting that rDNA copy number on its own was not the causative mutation arising from the ALE experiment. It is also possible that the rDNA encoded mitochondrial *TAR1* protein [30] plays a role in the quality control of defective mitochondria, particularly when mixed mitochondrial populations are inherited during mating [31].

The Tar1p response to defective mitochondria is proposed to be mediated via the amplification of the rDNA locus to increase *TAR1* expression [31], suggesting that the increased rDNA copy-number we observed in evolved populations could have occurred in response to defective mitochondria that were inherited from the crossing of partially synthetic strains to make synXIV (Figure 2-1). Defective mitochondria would have been present in the synthetic GX strain for many generations after the introduction of megachunk J and the growth defect associated with the *MRPL19* loxPsym motif. Expression of the *TAR1* gene from its native promoter on a pRS413 vector restored growth on YP-glycerol at 30 °C and 37 °C (Figure 2-4), suggesting that increased rDNA copy-number enabled higher *TAR1* expression and normal respiratory growth in the evolved lineages. Given that the evolved isolate J1.8e1 had restored rDNA locus copy number to that of the parent strain (Table 2-1) and retained an unaltered version of synXIV, this isolate was used as the final synXIV strain.

Table 2-1. Read depth of the ribosomal DNA locus (rDNA), surrounding chromosome XII sequence, and the ration of rDNA to surrounding chromosome XII sequence.

Strain	rDNA locus	ChrXII coverage	ratio
lap3::KanMX	11564	172	67
G	11141	168	66
G-O	7986	128	62
G-X	11988	204	59
J1.4	10510	186	56
J1.8	11512	188	61
J1.8e1	12688	188	67
J1.8e2	14561	214	68
J1.8e3	13893	197	70

Summary

Presented here is the construction and debugging of the fully synthetic chromosome XIV of *Saccharomyces cerevisiae*. Whole genome sequencing was a crucial approach to identify the genetic basis of fitness defects during chromosome construction, and subsequent modifications were made to create a synthetic chromosome that did not affect the fitness of the *S. cerevisiae* strain. This is the first instance of a completely synthetic chromosome being constructed in the southern hemisphere.

Materials and methods

Growth Media

S. cerevisiae strains were grown in synthetic dropout (SD) media containing Yeast Nitrogen Base Without Amino Acids mix (Sigma-Aldrich Y0626) supplemented with 10 g/L glucose, and amino acids at 100 mg/L to complement auxotrophies as appropriate. Alternatively, yeast was grown in 10 g/L yeast extract, 20 g/L peptone, 20 g/L dextrose (YPD), or YP-glycerol (20 g/L glycerol in place of dextrose). *E. coli* DH5α strains were grown in LB medium with ampicillin.

Growth conditions

All strains were cultivated in either 50 mL falcon tubes or 250 mL baffled shake-flasks with medium comprising 10 % of the total vessel volume. All growth experiments occurred at 30 °C with 200 rpm, 50 mm orbital shaking (InFors MultiTron HT).

Chunk preparation

DNA chunks comprising ~5-10 Kb of each megachunk were synthesised and sequence verified by Genscript (megachunks A-K and N-X), GeneArt (megachunks L, M), and GeneWiz (chunks E1, E2, E3, S4). Chunks were then either restriction digested using terminal, complementary sites incorporated in the design changes, or PCR amplified using primers that anneal to the very 5´ or 3´ ends of each chunk with Phusion polymerase (New England Biolabs). Plasmid digested or PCR amplified chunks were excised from agarose gels and quantified. Chunks were pooled together such that the relative amounts of each chunk were approximately halved such that chunk 1 > chunk 2 > chunk 3 > chunk 4, with the amount of chunk 4 being 400-800 ng. Restriction digested chunks were ligated over night at 16 °C using T4 DNA ligase (New England Biolabs).

Yeast transformation and marker-loss screening

Cells were transformed using the lithium acetate/polyethylene glycol/ssDNA transformation method [32]. After 2-5 days of incubation on selective media at 30 °C, colonies were replica-plated onto media selective for the marker gene used to integrate the previous megachunk, with those not able to grow used for further analysis.

DNA extraction and PCR-tag analysis

Genomic DNA was extracted using a lithium acetate-SDS solution for cell disruption followed by ethanol mediated DNA precipitation as previously described [33]. Crude DNA extracts were transferred to a 384-well plate compatible with the Echo 550 acoustic liquid transfer system (Labcyte), as were primer pairs for each PCR-tag in SynXIV (15 μM). 4.75 μL aliquots of 1x

SYBR green mastermix were added to each well of a 384-well qPCR plate using an epMotion liquid handling robot. 200 nL of crude gDNA and 50 nL of each primer-pair was transferred to each 384-well qPCR plate well using an Echo550 (Labcyte Inc.). The plate was centrifuged briefly to ensure transferred droplets were suspended in the SYBR-green mix. qPCR was carried out using a Lightcycler 480 with an initial 95 °C denaturation of 3 minutes followed by 15 cycle of 30 s at 95 °C, 30 s at 70 °C with a decrease of 1 °C each cycle, and extension at 72 °C for 30 s. The same denaturation and extension condition were then used for a further 20 cycles, except with constant annealing at 55 °C. SYBR-green fluorescence was measured at the end of each extension step. After cycling, a melt curve was generated by heating from 50 °C to 95 °C with fluorescence measurements every 5 s. For each megachunk, positive and negative controls were used that comprised of mixed synthetic chunk DNA or BY4741 DNA respectively. Any PCR-tags resulting in aberrant amplification were excluded from analysis of transformant DNA. Megachunk integration was accepted when all synthetic PCR-tags and no wild-type PCR-tags resulted in amplification.

CRISPR-Cas9 mediated genome modification

All CRISPR-Cas9 mediated homologous recombination was carried out by using a previously reported strategy that utilises a single episomal plasmid (pRS423) that contains both guide RNA and Cas9 expression cassettes [24]. New 20 bp guide RNA sequences were encoded in 5´ extensions of primers that target the 3´ end of the *SNR52* promoter (reverse primers) and the 5´ end of the structural crispr RNA (forward primer). ~ 100 ng of the linear PCR product resulting from this reaction was used to co-transform yeast with 1-5 µg of donor DNA with homology to the target guide-RNA locus. Colonies growing on SD –histidine media were screened for desired mutations using PCR-tag analysis and/or loci specific primers.

SynXIV discrepancy repair

As a default option, sequence discrepancies were repaired using our previously developed crispr-Cas9 system [24], whereby synthetic chunk DNA served as donor for homologous

recombination. Discrepancies 1-10 (Table S2-1) were repaired by targeting the *EGT2* ORF with CRISPR-Cas9 and synthetic chunks A3-A4 as donor DNA. The *EGT2* ORF was fully synonymously recoded as part of an error in the synXIV design phase, and we had subsequently reverted this sequence to wild-type. However, this made no difference to fitness, so the wild-type sequence was used as a guide-RNA target during the repair of discrepancies 1-10, leaving the sequence in its original fully recoded state. The two missing TAA stop codons on chunk M3 (*YNL113W* and *YNL114C* ORFs) had no available guide-RNA sites that were absent in the donor DNA, and were therefore repaired using a previously developed SNP reconstruction method [34]. A *URA3* expression cassette flanked by correct synthetic versions of the two genes, which had 200 bp of homology to one-another were used to replace the target locus. After verifying correct integration and stop-codon replacement, the strain was grown for several generations in YPD medium to enable 'loop-out' of the *URA3* cassette via the 200 bp homologous regions on the re-integrated *YNL113W* and *YNL114C* ORFs. The population was then plated on minimal medium containing uracil and 5-FOA to select for those cells without functional *URA3* expression. Loss of the marker was verified using PCR. Discrepancies that were located on terminal, marker-containing chunks (numbers 12-13 and 18-20, Table S2-1) were repaired by integrating the relevant chunk using selection for its marker (*LEU2* in both cases). The marker was then removed by targeting it using a *LEU2* specific CRISPR-Cas9 cassette and 3′ chunk donor DNA.

MRPL19-GFP fusion protein design

The *MRPL19* protein was internally tagged with super-folder Green Fluorescent Protein (sfGFP) [35] in its coding sequence with and without the loxPsym site in the 3′ UTR, and a cytosol-localised GFP control [24], respectively. We inserted an in-frame *sfGFP* sequence inside the coding sequence of *MRPL19* (between position 282 and 283) because this gene encodes a predicted mitochondrial N-terminal peptide targeting signal [22] and a 3′ UTR mRNA signal that mediates mRNA localization to mitochondria-bound polysomes involved in mitochondria

protein import [23]. The mitochondrial N-terminal peptide targeting signal was identified using MitoFates [22] and by generating and analysing a 3D protein model of Mrpl19p using SWISS-MODEL [36], which also revealed an N-terminal β-hairpin motif predicted to target proteins to mitochondria [37]. These mitochondrial targeting signals would have been disrupted by placing the sfGFP at either the N- or C-terminus of Mrpl19p. To promote proper folding of this fusion protein, we flanked the sfGFP with flexible linkers (L) [38] halfway through the *MRPL19* ORF. The resulting fusion protein had the following design: *MRPL1994-L-sfGFP-L-95MRPL19*. The native promoter and terminator regions were maintained, except for the version containing a loxPsym site 3 bp after the stop codon. These two cassettes were synthesised by Genscript Inc. and cloned onto pRS416 vectors using XhoI and NotI (Table S2-3).

Confocal microscopy

BY4741 strains transformed with *MRPL19-sfGFP-loxP-pRS416*, *MRPL19-sfGFP-Native 3′ UTR-pRS416* or with cytosol localised GFP expression (*pPDR12-GFP-pRS416* [24]) were pre-cultured twice in minimal medium without uracil before being inoculated at an OD_{600nm} of 0.4 in fresh medium. Cells were treated with 100 nM Mitotracker Red FM (ThermoFisher M22425) for 3-4 hours with shaking at 30 °C. Cells were kept on ice prior to microscopic examination. Visualisation of GFP and Mitotracker Red FM signals was performed using an Olympus FV 1000 confocal laser-scanning microscope. Microscopy images were analysed using ImageJ (https://imagej.nih.gov/ij/index.html). Images shown are representative of cells in independent biological triplicate populations.

Diploid formation

Strains to be crossed had selective markers integrated in wild-type chromosome XIV regions to provide a means of screening randomly isolated spores for those with a greater chance having crossed over in the desired megachunk. Strains of opposite mating type and with complementary auxotrophies were grown overnight separately in 5 mL of selective SD media. 500 µL of each culture was used to inoculate the same non-selective 5 mL of SD medium,

which was incubated overnight at 30 °C without shaking. The overnight culture was washed twice in sterile MilliQ water before being plated on solid medium selective for the respective auxotrophies in each strain, such that only diploids would form colonies. Putative diploid colonies were checked using 'mating type' primers to verify the presence of both 'a' and 'alpha' alleles at the *MAT* locus, indicating the formation of a diploid.

Sporulation, random spore isolation, and random spore screening

To initiate sporulation, diploid colonies were grown overnight in 5 mL of selective liquid SD medium, washed once with sterile MilliQ water, and plated on 10 g/L potassium acetate medium. Plates were incubated in the dark at room temperature for 4-7 days. Once asci were visible under light microscopy, as many cells as possible were scraped from the potassium acetate plate and resuspended in 500 µL of sterile MilliQ water with 10 units of zymolyase and 20 µL of beta-mercaptoethanol. This solution was incubated at 37 °C for 3-4 hours before being transferred to a 250 mL flask containing 20 mL of acid washed glass beads and 30 mL of sterile MilliQ water. Flasks were incubated at 30 °C overnight with 200 rpm shaking. The liquid fraction was recovered, washed once in sterile MilliQ, and a dilution series down to 10^{-3} plated on YPD with incubation at 30 °C for 1-2 days. Colonies were replica plated onto SD plates selective for each of the auxotrophic markers present in the haploid parent strains, and any colonies found not growing on each plate type were selected for PCR-tag analysis using the 2st tag of each of the 22 synXIV megachunks. Colonies with synthetic PCR-tag amplification and without wild-type PCR-tag amplification were deemed likely to contain the corresponding megachunk, and further screened using all PCR-tags for the relevant megachunks.

RNA extraction

1.5 mL samples of mid-exponential phase cultures (OD_{600} of 0.5 – 2.5) were pelleted by spinning at 12,000 x g for 2 minutes and removing the supernatant. Pellets were resuspended in 1 mL of RNAlater (ThermoFisher Scientific catalogue number AM7020) and stored at -20 °C. RNA was extracted after pelleting cells and removing RNAlater solution using the Zymo

Research YeaStar RNA extraction kit (catalogue number R1002) according to the user manual. Co-purified DNA was removed from RNA extracts using TURBO™ DNase (ThermoFisher Scientific catalogue number AM2238) according to the user manual.

RT-qPCR

100 – 1000 ng of purified RNA was used for reverse transcription using an 18 nucleotide poly-T primer and SuperScript™ III Reverse Transcriptase (ThermoFisher Scientific 18080093) according to the user manual. A no-enzyme control was included for each RNA sample and subsequently used for qPCR to verify that no genomic DNA was contributing to cDNA concentration estimates. Reverse transcribed samples were diluted 1:100 in MilliQ water prior to qPCR analysis. Relative expression was performed using the modified-Livak method (amplification efficiency measured for each primer-pair and not assumed to be log2) with *ALG9* as a housekeeping gene [39], as previously described[20], except that expression values of the control strain were not arbitrarily normalised to '1'.

Whole-genome sequencing

A yeast genomic DNA extraction kit (ThermoFisher catalog number 78870) was used to isolate DNA according to the manufacturer's instructions. Sequencing and library preparation were carried out by Macrogen Inc. using a True-Seq Nano kit with 470 bp inserts, and paired-end illumina HiSeq 2500 sequencing, or by the Ramaciotti Centre for Genomics using Nextera XT library preparation and 2x 150 bp paired end sequencing using a NextSeq500 (Sequencing of samples from the adaptive laboratory evolution experiment). Reads were analysed using Geneious Pro v10.2.2 software [40]. Paired-end reads were mapped to an edited version of the S288C reference genome where native chromosome 14 was replaced with synthetic chromosome XIV. The Geneious alignment algorithm was used to map reads to the reference genome using 'highest sensitivity' settings. Analysis of the resultant assembly was completed visually by assessing read coverage, and read disagreement with the reference sequence. The raw reads were of high-quality (Q30 > 91 %, Q20 > 95%), and were therefore not trimmed

prior to assembly. Average read depth of 190 was typically achieved from the Macrogen sequencing, while 50-fold coverage was used for the samples sequenced at the Ramaciotti Centre for Genomics. Single Nucleotide Polymorphisms (SNPs) and their effect on ORF reading frames and codons were detected using the Geneious 'Find Variations/SNPs' function with a variant p-value threshold of 10^{-6} and variant frequency of ≥ 50 %.

Acknowledgements

The Synthetic Biology initiative at Macquarie University is financially supported by an internal grant from the University, and external grants from Bioplatforms Australia, the New South Wales (NSW) Chief Scientist and Engineer, and the NSW Government's Department of Primary Industries. Ian Paulsen is supported by an Australian Research Council Laureate Fellowship. TCW and BL are supported by Fellowships from the CSIRO Synthetic Biology Future Science Platform and Macquarie University.

References

1 Pretorius, I. S. Synthetic genome engineering forging new frontiers for wine yeast. *Crit. Rev. Biotechnol.* **37**, 112-136, doi:10.1080/07388551.2016.1214945 (2017).

2 Pretorius, I. S. & Boeke, J. D. Yeast 2.0—connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Res.* **18**, foy032-foy032, doi:10.1093/femsyr/foy032 (2018).

3 Richardson, S. M. *et al.* Design of a synthetic yeast genome. *Science* **355**, 1040 (2017).

4 Wu, Y. *et al.* Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355** (2017).

5 Mitchell, L. A. *et al.* Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355** (2017).

6 Shen, Y. *et al.* Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355** (2017).

7 Xie, Z.-X. *et al.* "Perfect" designer chromosome V and behavior of a ring derivative. *Science* **355** (2017).

8 Zhang, W. *et al.* Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355** (2017).

9 Mercy, G. *et al.* 3D organization of synthetic and scrambled chromosomes. *Science* **355** (2017).

10 Annaluru, N. *et al.* Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science* **344**, 55-58, doi:10.1126/science.1249252 (2014).

11 Dymond, J. & Boeke, J. The Saccharomyces cerevisiae SCRaMbLE system and genome minimization. *Bioengineered Bugs* **3**, 168-171, doi:10.4161/bbug.19543 (2012).

12 Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476, doi:http://www.nature.com/nature/journal/v477/n7365/abs/nature10403.html - supplementary-information (2011).

13 Jovicevic, D., Blount, B. A. & Ellis, T. Total synthesis of a eukaryotic chromosome: Redesigning and SCRaMbLE-ing yeast. *Bioessays* **36**, 855-860, doi:10.1002/bies.201400086 (2014).

14 Shen, Y. *et al.* SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res.* **26**, 36-49, doi:10.1101/gr.193433.115 (2016).

15 Vickers, C. E., Williams, T. C., Peng, B. & Cherry, J. Recent advances in synthetic biology for engineering isoprenoid production in yeast. *Curr. Opin. Chem. Biol.* **40**, 47-56, doi:https://doi.org/10.1016/j.cbpa.2017.05.017 (2017).

16 Williams, T. C., Pretorius, I. S. & Paulsen, I. T. Synthetic Evolution of Metabolic Productivity Using Biosensors. *Trends Biotechnol.* **34**, 371-381, doi:10.1016/j.tibtech.2016.02.002 (2016).

17 Gibson, D. G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci.* **105**, 20404-20409, doi:10.1073/pnas.0811011106 (2008).

18 Merz, S. & Westermann, B. Genome-wide deletion mutant analysis reveals genes required for respiratory growth, mitochondrial genome maintenance and mitochondrial protein synthesis in Saccharomyces cerevisiae. *Genome Biology* **10**, R95, doi:10.1186/gb-2009-10-9-r95 (2009).

19 Parisi, M. A., Xu, B. & Clayton, D. A. A human mitochondrial transcriptional activator can functionally replace a yeast mitochondrial HMG-box protein both in vivo and in vitro. *Mol. Cell. Biol.* **13**, 1951, doi:10.1128/MCB.13.3.1951 (1993).

20 Williams, T. C. *et al.* Quorum-sensing linked RNA interference for dynamic metabolic pathway control in Saccharomyces cerevisiae. *Metab. Eng.* **29**, 124-134, doi:http://dx.doi.org/10.1016/j.ymben.2015.03.008 (2015).

21 Peng, B., Williams, T. C., Henry, M., Nielsen, L. K. & Vickers, C. E. Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microb. Cell. Fact.* **14**, 91, doi:10.1186/s12934-015-0278-5 (2015).

22 Fukasawa, Y. *et al.* MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Molecular & cellular proteomics : MCP* **14**, 1113-1126, doi:10.1074/mcp.M114.043083 (2015).

23 Saint-Georges, Y. *et al.* Yeast Mitochondrial Biogenesis: A Role for the PUF RNA-Binding Protein Puf3p in mRNA Localization. *PLOS ONE* **3**, e2293, doi:10.1371/journal.pone.0002293 (2008).

24 Williams, T. C., Xu, X., Ostrowski, M., Pretorius, I. S. & Paulsen, I. T. Positive-feedback, ratiometric biosensor expression improves high-throughput metabolite-producer screening efficiency in yeast. *Synthetic Biology* **2**, doi:10.1093/synbio/ysw002 (2017).

25 Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution - principles and applications for biotechnology. *Microb. Cell. Fact.* **12**, 64 (2013).

26 Strucko, T. *et al.* Laboratory evolution reveals regulatory and metabolic trade-offs of glycerol utilization in Saccharomyces cerevisiae. *Metab. Eng.* **47**, 73-82, doi:https://doi.org/10.1016/j.ymben.2018.03.006 (2018).

27 Kobayashi, T. Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cellular and molecular life sciences : CMLS* **68**, 1395-1403, doi:10.1007/s00018-010-0613-2 (2011).

28 Kobayashi, T., Heck, D. J., Nomura, M. & Horiuchi, T. Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes Dev.* **12**, 3821-3830, doi:10.1101/gad.12.24.3821 (1998).

29 Kwan, E. X., Wang, X. S., Amemiya, H. M., Brewer, B. J. & Raghuraman, M. K. rDNA Copy Number Variants Are Frequent Passenger Mutations in Saccharomyces cerevisiae Deletion Collections and de Novo Transformants. *G3 (Bethesda, Md.)* **6**, 2829-2838, doi:10.1534/g3.116.030296 (2016).

30 Bonawitz, N. D., Chatenay-Lapointe, M., Wearn, C. M. & Shadel, G. S. Expression of the rDNA-encoded mitochondrial protein Tar1p is stringently controlled and responds differentially to mitochondrial respiratory demand and dysfunction. *Curr. Genet.* **54**, 83-94, doi:10.1007/s00294-008-0203-0 (2008).

31 Poole, A. M., Kobayashi, T. & Ganley, A. R. D. A positive role for yeast extrachromosomal rDNA circles? *Bioessays* **34**, 725-729, doi:10.1002/bies.201200037 (2012).

32 Gietz, R. D. & Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protocols* **2**, 35-37 (2007).

33 Lõoke, M., Kristjuhan, K. & Kristjuhan, A. EXTRACTION OF GENOMIC DNA FROM YEASTS FOR PCR-BASED APPLICATIONS. *BioTechniques* **50**, 325-328, doi:10.2144/000113672 (2011).

34 Brennan, T. C., Kromer, J. O. & Nielsen, L. K. Physiological and transcriptional responses of Saccharomyces cerevisiae to d-limonene show changes to the cell wall but not to the plasma membrane. *Appl. Environ. Microbiol.* **79**, doi:10.1128/aem.00463-13 (2013).

35 Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79, doi:10.1038/nbt1172

https://www.nature.com/articles/nbt1172 - supplementary-information (2005).

36 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-w303, doi:10.1093/nar/gky427 (2018).

37 Jores, T. *et al.* Characterization of the targeting signal in mitochondrial β-barrel proteins. *Nature Communications* **7**, 12036, doi:10.1038/ncomms12036

https://www.nature.com/articles/ncomms12036 - supplementary-information (2016).

38 Edwards, W. R., Busse, K., Allemann, R. K. & Jones, D. D. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Res.* **36**, e78, doi:10.1093/nar/gkn363 (2008).

39 Teste, M.-A., Duquenne, M., Francois, J. & Parrou, J.-L. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. *BMC Mol. Biol.* **10**, 99 (2009).

40 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).

Chapter 3 : Positive-feedback, ratiometric biosensor expression improves high-throughput metabolite-producer screening efficiency in yeast

Synthetic Biology, 2017, 2(1): ysw002

doi: 10.1093/synbio/ysw002
Research article

Positive-feedback, ratiometric biosensor expression improves high-throughput metabolite-producer screening efficiency in yeast

Thomas C. Williams, Xin Xu, Martin Ostrowski, Isak S. Pretorius, and Ian T. Paulsen*

Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW, Australia

*Corresponding author: E-mail: ian.paulsen@mq.edu.au

Abstract

Biosensors are valuable and versatile tools in synthetic biology that are used to modulate gene expression in response to a wide range of stimuli. Ligand responsive transcription factors are a class of biosensor that can be used to couple intracellular metabolite concentration with gene expression to enable dynamic regulation and high-throughput metabolite producer screening. We have established the *Saccharomyces cerevisiae WAR1* transcriptional regulator and *PDR12* promoter as an organic acid biosensor that can be used to detect varying levels of para-hydroxybenzoic acid (PHBA) production from the shikimate pathway and output green fluorescent protein (GFP) expression in response. The dynamic range of GFP expression in response to PHBA was dramatically increased by engineering positive-feedback expression of the *WAR1* transcriptional regulator from its target *PDR12* promoter. In addition, the noise in GFP expression at the population-level was controlled by normalising GFP fluorescence to constitutively expressed mCherry fluorescence within each cell. These biosensor modifications increased the high-throughput screening efficiency of yeast cells engineered to produce PHBA by 5,000-fold, enabling accurate fluorescence activated cell sorting isolation of producer cells that were mixed at a ratio of 1 in 10,000 with non-producers. Positive-feedback, ratiometric transcriptional regulator expression is likely applicable to many other transcription-factor/promoter pairs used in synthetic biology and metabolic engineering for both dynamic regulation and high-throughput screening applications.

Key words: biosensor; adaptive laboratory evolution; synthetic biology; metabolic engineering; yeast.

1. Introduction

Advances in synthetic biology and metabolic engineering are beginning to enable alternative routes for chemical, fuel, and pharmaceutical manufacturing. Microorganisms can now be engineered to convert renewable agricultural resources into a wide array of metabolites that can serve as either replacements or alternatives to existing industrial products. However, there are significant challenges involved in engineering microbial cells to produce desired products at commercially viable titres, rates, and yields. Traditional approaches involve balancing the expression levels of metabolic pathway enzymes leading to the desired product, the elimination of enzymes that compete for carbon flux, and the balancing of cellular redox and energy states (Nielsen and Keasling 2016). These biological components often interact synergistically to control complex metabolic regulation, and their rearrangement/optimisation therefore entails a combinatorial explosion of complex traits that require building and testing for performance. Although *in silico* modelling

Submitted: 17 October 2016; **Received (in revised form):** 14 November 2016. **Accepted:** 29 November 2016

approaches have proven invaluable for understanding metabolism and improving metabolite production (Wiechert 2002; Patil et al. 2004), metabolic engineering 'design principles' are yet to be fully elucidated due to our incomplete knowledge of living systems. Subsequently, it can take many iterations of the classical design-build-test cycle to achieve engineering objectives, some of which may even be impossible to meet using available biological knowledge.

An elegant way to overcome the challenges associated with engineering in biology is to apply a selective pressure to a genetically diverse population so that cells with the desired phenotype can be isolated. Using selective pressure in this way means that biological knowledge and the limited capacity to design-build-test no longer limit the solutions available to biological engineering problems. While methods such as adaptive laboratory evolution can be used to generate populations with superior tolerance to growth inhibiting chemicals (Goodarzi et al. 2010; Kildegaard et al. 2014; Almario et al. 2013; Brennan et al. 2015), altered substrate specificity (Wisselink et al. 2009; Garcia Sanchez et al. 2010; Quan et al. 2012; Zhou et al. 2012), and increased temperature tolerance (Riehle et al. 2003; Caspeta et al. 2014), this approach cannot be readily applied to the selection of high metabolite yields because phenotypes such as metabolic productivity are usually not naturally coupled to cell survival. One of the most promising ways to make this connection and select for cells with higher metabolic productivity is to use a biosensor that detects the intracellular concentration of a metabolite of interest and outputs a survival function in response (Williams et al. 2016). The output is typically green fluorescent protein (GFP) expression so that cells that are more productive within an evolving or mutated population can be isolated using fluorescence activated cell sorting (FACS) (Williams et al. 2016). The advantage of using FACS is that thousands of cells can be screened per second, as opposed to per week or month using traditional analytical methods for metabolites.

Biosensors are gaining prominence both for the dynamic control of metabolic pathways in response to metabolites, and as conduits between metabolite productivity and cell survival for high-throughput screening (Liu et al. 2015; Zhang et al. 2015; Williams et al. 2016). Allosterically regulated transcription factors are abundant in nature, with a large number characterised in terms of their activating ligands and target promoters (Tropel et al. 2004; Ramos et al. 2005; Gallegos et al. 1997; Taylor et al. 2016). These proteins have therefore typically been the first option for those seeking metabolite biosensors for metabolic engineering applications and high-throughput screening (Taylor et al. 2016; Williams et al. 2016). However, many naturally occurring transcription factors have activation dynamics that are not ideal for high-throughput metabolite-producer cell screening. Features such as high basal expression levels, low dynamic range of expression, and stochastic variation in gene expression across a population (noise) can all reduce the effective screening throughput by making it difficult to distinguish 'producer' cells from non-producers within a genetically diverse population.

With the aim of establishing a generic synthetic biology approach to select for enhanced metabolite producers we have developed an amplified ligand-responsive transcription-factor biosensor with a built-in ratiometric noise suppressor to enhance the efficacy of high-throughput screening. As a proof of concept, we focused on transcriptional regulators that respond to organic acids, as there is great interest in producing these compounds biologically (Chen and Nielsen 2016). Organic acids are used to make a variety of products such as plastics, solvents, polymers and chemical 'building blocks', animal feed, nylons, flavours and fragrances, as well as food and beverage ingredients (Sauer et al. 2008). Most organic acids are currently produced from petrochemical feed-stocks, and there is therefore significant interest in implementing renewable and more environmentally friendly production processes using microbial hosts (Sauer et al. 2008). The yeast *Saccharomyces cerevisiae* is a well characterised eukaryotic model organism, industrial work-horse, and synthetic biology 'chassis' cell (Pretorius 2016). *S. cerevisiae* is also preferred industrial producer of organic acids due to its ability to grow at a low pH, enabling low purification costs and reducing microbial contamination from non-sterile substrates (Sauer et al 2008; Abbott et al. 2009). Consequently, there have been intensive efforts in metabolic engineering of yeast for the production of valuable organic acids such as lactic (Ishida et al. 2006; Baek et al. 2015), succinic (Raab et al. 2010; Agren et al. 2013; Ito et al. 2014), malic (Zelle et al. 2008), 3-hydroxypropionic (Borodina et al. 2015; Kildegaard et al. 2015), 4-hydroxybenzoic and para-aminobenzoic (Krömer et al. 2012; Averesch and Krömer 2014; Williams et al. 2015), itaconic (Blazeck et al. 2014), and muconic (Curran et al. 2013) acids.

2. Materials and methods

2.1 Growth media

All *S. cerevisiae* strains were grown in synthetic dropout (SD) media containing Yeast Nitrogen Base Without Amino Acids mix (Sigma-Aldrich Y0626) supplemented with 10 g/l glucose, and amino acids at 100 mg/l to complement auxotrophies as appropriate. pH was adjusted to either 6.5 or 3.5 as indicated in individual experiments. *Escherichia coli* DH5α strains were grown in LB medium with ampicillin.

2.2. Growth conditions

For all dose–response GFP measurement and cell-sorting experiments, culturing methods and pre-culture methods were as follows. Glycerol stocked strains were inoculated into 5 ml of SD medium in a 50-ml sterile Falcon tube and grown for 8–10 hours at 30 °C, 200 rpm in a shaking incubator. These cultures were then used to inoculate 10 ml of SD medium at an OD_{600nm} of 0.02–0.04 and grown overnight for approximately 18 hours. The next morning, exponentially growing cultures (OD_{600nm} 0.1–1.5) were used to inoculate experimental cultures in triplicate, at an OD_{600nm} of 0.1. Populations were grown in 24-well micro-plates in a total volume of 1.5 ml containing any organic acids involved in promoter screening or dose–response experiments at the indicated concentrations. Weak acid response experiments were carried out in SD medium at pH 3.5 such that acid molecules outside of cells would freely diffuse inside, potentially being available for promoter/GFP activation (Holyoak et al. 1999). Cell sorting and GFP measurement experiments with para-hydroxybenzoic acid (PHBA) producer populations were carried out in SD medium at pH 6.5 such that any organic acids produced and secreted by individual cells would not enter other cells (Holyoak et al. 1999) and interfere with the GFP readout from their own PHBA production capacity (or lack of). For cell sorting experiments, pre-cultured producer and non-producer populations were inoculated at exactly OD_{600nm} 0.05 in separate tubes prior to being mixed together at the indicated ratios within the same 50-ml Falcon tube. To reach the dilutions ranging from 1:10 to $1:10^5$ producers to non-producers, serial 10-fold dilutions were made starting from the 1:10 population. Both multi-well plate and Falcon-tube cultures were grown at 30 °C with 200 rpm

shaking in an InFors incubator for 3 hours prior to GFP measurement.

2.3 Flow cytometry

A BD Influx flow cytometer was used for all fluorescent protein measurements and cell sorting. For GFP measurement a 200 mW 488-nm laser was used for excitation with emission filters at 540 ± 30 nm. A GFP negative control strain (415.C, Table 3) was used to measure auto-fluorescence in parallel to GFP positive strains, and the 540 ± 30 nm PMT voltage was set such that the mean auto-fluorescence value was either 2 or 4, as indicated in the supplementary data (fluorescence of no-GFP control strain 415.C, Table 3). Mean GFP values from experimental populations were divided by auto-fluorescence, and all conditions were measured in triplicate cultures, with mean and SDs reported. Mixed producer/non-producer populations were sorted according to GFP or GFP:mCherry fluorescence whereby gates that excluded any cells below a GFP fluorescence level that occurred in a yeast population containing the *pPDR12-GFP* or *pPDR12-GFP-pPDA1-mCherry* biosensor (measured in parallel), but not containing a production pathway (strains GFP.415, +FB.GFP.415, +FB.GFP.mCherry.415, Table 3). Alternatively, in the case of the GFP-only sorting experiment (Fig. 2d), a gate encompassing the top 3.4% of the mixed population was used for sorting. mCherry fluorescence was measured simultaneously to GFP fluorescence in individual cells using the 488-nm laser. A 692/40-emission filter was used to measure mCherry fluorescence with PMT gain set to 75 and the 540/30 PMT gain set to 35. These gain settings were imposed with the goal of enhancing mCherry signal while reducing the affect of GFP emission leakage into the mCherry channel. Likewise the 692/40-emission filter was used in place of the 610/20 filter (closer to the mCherry emission optima) in order to reduce GFP leakage into the mCherry channel. mCherry expression was verified by comparing the mCherry.415 strain 692/40 values to those of the 415.C control strain. Overlay histograms of GFP and GFP:mCherry distributions were made using GraphPad Prism 7 software. 10,000 events were recorded and FCS files were converted to csv prior to being imported into GraphPad Prism, where GFP fluorescence values were grouped into 100 arbitrary units (au) bins, and GFP:mCherry values into 0.05 au bins. Fluorescence values/ratios were plotted on the x-axes with event numbers on the y-axes.

2.4 Strain and plasmid construction

Primers, plasmids, and strains used in this study are shown in Tables 1, 2, and 3, respectively. Strains were constructed by transforming plasmids into the relevant yeast strain using the lithium acetate method (Gietz et al. 2007) and selecting for growth on appropriate auxotrophic dropout agar plates. All *in silico* cloning, Gibson assembly, and primer design was carried out using Geneious Pro software (Kearse et al. 2012), version 9.1.5. *E. coli* DH5α was used for cloning using standard techniques (Sambrook and Russell 2001) unless otherwise mentioned. For the construction of pRS415 biosensor plasmids, *PDR12* (primers 1/2), *YGP1* (primers 3/4), and *TPO2* (primers 5/6) promoter regions were amplified from *S. cerevisiae* BY4741 DNA and Gibson Assembled (Gibson et al. 2009) 5′ of the *yEGFP-ADH1t* sequence amplified from the p413-TR-SSRE-GFP plasmid (Chen and Weiss 2005) (primers 7/8, 9/10, 11/12 for amplification of *yEGFP-ADH1t* and assembly with *pPDR12*, *pYGP1*, and *pTPO2* respectively). For *pPDR12-yEGFP* assembly the pRS415 plasmid was linearized via PCR amplification with primers pRS-F and pRS-R, while for *pYGP1-yEGFP* and *pTPO2-yEGFP*, pRS415 was linearized via digestion with SmaI and Eco53KI enzymes respectively. The high copy-number pPDR12-yEGFP-pRS425 plasmid was made using the same approach as for pPDR12-yEGFP-pRS415, except with pRS425.

The native *WAR1* promoter was replaced with the DNA binding target of the WAR1p transcriptional regulator (*PDR12* promoter) to create a positive feedback loop using CRISPR-Cas9-mediated (Clustered Regularly Interspersed Short Palindromic Repeat) targeting and homologous recombination. The CRISPR guide RNA expression cassette and *Streptococcus pyogenes* Cas9 gene from the work reported by DiCarlo et al. (2013) were first combined on the same high copy number plasmid with the *HIS3* marker (pRS423). The guide RNA expression cassette comprising the *SNR52* promoter, *CAN1.Y* targeting guide RNA, structural crRNA, and *CYC1* terminator from the p426-gRNA.CAN1.Y plasmid (DiCarlo et al. 2013) were PCR amplified (primers 13/14) and Gibson assembled into PCR linearised pRS423 (primers 15/16) to create the 'gRNA-423' plasmid. The *TEF1* promoter, *Cas9* gene, and *CYC1* terminator from p*TEF1-Cas9*-pRS414 (DiCarlo et al. 2013) were PCR amplified with primers 17/18 to create 40-bp homologous overlaps to the *SmaI* digested gRNA-423 plasmid, and the two were Gibson Assembled to make Cas9-gRNA-423. This plasmid still contained the guide RNA sequence designed to target the *CAN1.Y* locus from the original CRISPR study (DiCarlo et al. 2013). To generate a *WAR1* promoter-targeting guide RNA for CRISPR-mediated knock-in of the *PDR12* promoter, the entire Cas9-gRNA-423 plasmid was PCR amplified using primers that bind either side of the existing *CAN1.Y* guide RNA sequence. These primers (19/20) contained 20 nt 5′ extensions that encode a new CRISPR guide RNA specific for the *WAR1* promoter region. The new CRISPR guide RNA encoding extensions on the forward and reverse primers for the Cas9-gRNA-423 plasmid were also 100% homologous to one another so that the linearized PCR product could be circularised to create Cas9-pWAR1-crRNA-423 using Gibson Assembly (Gibson et al. 2009) or Yeast Assembly (Gibson et al. 2008; Shao et al. 2009). In this case, Yeast Assembly was used where PCR linearized and *Dpn*I-treated plasmid (to remove template DNA) was co-transformed with the *PDR12* promoter DNA. The *PDR12* promoter was PCR amplified from *S. cerevisiae* BY4741 genomic DNA using primers (21/22) with 5′ extensions that encode 60 bp of homology with the *WAR1* promoter region such that the 221 bp 5′ of the *WAR1* start codon would be replaced with the 780 bp *PDR12* promoter. CRISPR-mediated promoter replacement was achieved by co-transforming approximately 200 ng of Cas9-pWAR1-crRNA-423 and 5 μg of *PDR12* promoter PCR product. Cells were plated on SD –His medium to select for the CRISPR plasmid and screened using primers that anneal outside of the *WAR1* promoter region (23/24). Only 3 out of 28 colonies screened had the *WAR1* promoter replaced with the *PDR12* promoter using this method.

The PHBA production pathway comprised a feedback-resistant version (Q166K) (Hartmann et al. 2003; Fukuda et al. 1992) of the 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase enzyme (*ARO4* gene) and a chorismate pyruvate lyase enzyme encoded by a codon optimised version of the *E. coli UBiC* gene, which was previously shown to function in *S. cerevisiae* using a synthetic quorum sensing circuit (Williams et al. 2013, 2015). Here, *UBiC* and *ARO4* were expressed constitutively using the *TEF1* promoter, from the pRS416 plasmid. A previously described plasmid (*pTEF1-UBiC-CYC1t-pTEF1-ARO4-CYC1t*-pRS406) (Williams et al. 2016) was digested with *Sal*I and *Not*I to release the *UBiC-ARO4* expression cassette, which was gel purified prior

Table 1. Primers used in this study.

Primer number/name	5′ to 3′ sequence
1/pPDR12F	CGAGGTCGACGGTATCGATCTAAACCAAAGATGGATTGTTTACCA
2/pPDR12R	ACCTTTAGACATTTTTTTATTAATAAGAACAATAACA
3/pYGP1F	ATCGAATTCCTGCAGCCCAGCGTGCTATTTTTTAAAAAGGGC
4/pYGP1R	CCTTTAGACATTTTCTATTACTGTATTACTTAACTGACGA
5/pTPO2F	GCCGCCACCGCGGTGGAGCCTATGCAAAAACCCTTCCCC
6/pTPO2R	CCTTTAGACATATTTGTTTTGTGTATTATTTTTGTGA
7/pPDR12-yEGFPF	ATTAATAAAAAAATGTCTAAAGGTGAAGAATTATTCACTGG
8/pPDR12-yEGFPR	GGAATTCGATATCAAGCTTAATATTACCCTGTTATCCCTAGCGG
9/pYGP1-yEGFPF	AGTAATAGAAAATGTCTAAAGGTGAAGAATTATTCACTGG
10/pYGP1-yEGFPR	AGAACTAGTGGATCCCCCATATTACCCTGTTATCCCTAGCGG
11/pTPO2-yEGFPF	ACAAAACAAATATGTCTAAAGGTGAAGAATTATTCACTGG
12/pTPO2-yEGFPR	AGGGAACAAAAGCTGGAGATATTACCCTGTTATCCCTAGCGG
13/crRNAF	CGAGGTCGACGGTATCGAGCTTCTTTGAAAAGATAATGT
14/crRNAR	GGAATTCGATATCAAGCTTAGGCCGCAAATTAAAGCCTTC
15/pRSF	TAAGCTTGATATCGAATTCC
16/pRSR	TCGATACCGTCGACCTCG
17/Cas9F	ATCGAATTCCTGCAGCCCCATAGCTTCAAAATGTTTCTACTCCT
18/Cas9R	AGAACTAGTGGATCCCCCGGCCGCAAATTAAAGCCTTC
19/pWAR1-crRNAF	TAGTGTGTATTGACTGTGATGTTTTAGAGCTAGAAATAGCAAGTTA
20/pWAR1-crRNAR	ATCACAGTCAATACACACTAGATCATTTATCTTTCACTGCG
21/pPDR12F	TTACAAGTTCGTGCATATATAGAAAGAATTCTGTTGTTGTAATTGTCATAACTATTGAGCTCTAAACCAA AGATGGATTGTTTACCA
22/pPDR12R	ATTATCATTATTGATTTCTTTCCCGACGGCAACGCCAGTTATTGCAATCTGCGTGTCCATTTTTTTATTAA TAAGAACAATAACA
23/pWAR1checkF	AACCTGCTGAACCAACAAAACC
24/pWAR1checkR	AACTTTTTGGTCGGTCTTTG
25/4HB-KanMX F	GCCGCCACCGCGGTGGAGTAGGTCTAGAGATCTGTTTAGCTTGC
26/4HB-KanMX R	AGGGAACAAAAGCTGGAGATTAAGGGTTCTCGAGAGCTCG
27/pPDA1 F	GTATTCTGATAAATCTAAAGAGA
28/pPDA1 R	TGGCACAAATGTGGTTTCCT
29/mCherry F	ATGGTTTCTAAGGGTGAAGAAGACA
30/mCherry R	TTACTTGTACAATTCGTCCATACCAC
31/CYC1t F	AAGGCCCCTTTTCCTTTGTC
32/CYC1t R	CGACGATGAGAGTGTAAACTGC
33/UBiC-pTEF1 F	TATTATGTCGACCTCGAGGCACACACCATAG
34/UBiC-CYC1t R	TATTATGCGGCCGCACGATGAGAGTGTAAACTGC

to cloning into *SalI*/*NotI* digested pRS416 to create *pTEF1-UBiC-CYC1t-pTEF1-ARO4-CYC1t*-pRS416 (named UA-416). A pRS416 plasmid containing only the *pTEF1-UBiC-CYCt* expression cassette (U-416) was created by digesting this cassette from the UA-416 plasmid using *SalI*/*Eco*RI enzymes and cloning the gel purified product into pRS416. To facilitate easy identification of sorted-cells that contain the PHBA production genes, the G418 antibiotic resistance marker (*KanMX*) was PCR amplified from pUG6 DNA using primer pair 25/26 and Gibson assembled onto the *Eco53KI* linearised UA-416 plasmid to make UA-KanMX-416.

A ratiometric biosensor was made by expressing a red fluorescent protein (mCherry) from the constitutive *PDA1* promoter (Peng et al. 2015) on the same pRS415 plasmid that the *pPDR12-yEGFP* biosensor is encoded on. The *PDA1* promoter and *CYC1* terminator were PCR amplified from BY4741 genomic DNA using primer pairs 27/28 and 31/32, while the mCherry open reading frame was synthesised by IDT as a gBlock and PCR amplified using primer pair 29/30 for Gibson assembly with *pPDA1* and *CYC1t* into *SmaI* linearised *pPDR12-GFP*-415 plasmid to make *pPDR12-GFP-mCherry*-415. The same assembly reaction was carried out with empty, *Sma*I linearised pRS415 vector to make mCherry-415.

2.5 PCR-mediated identification of producers

Cells with GFP:mCherry fluorescence values deemed to indicate the presence of the *UBiC* and *ARO4*^{Q166K} genes were sorted directly onto SD minus leucine and uracil agar plates. After 3–4 days DNA was extracted from single colonies by resuspending them in 50 μl of MilliQ water with 16 units of zymolyase enzyme (Zymo Research catalogue number E1005), incubating at 37 °C for 30–60 minutes then at 95 °C for 10 minutes. Two microlitres of crude DNA extract was used as template for PCR using primers 33/34 that anneal to the *TEF1* promoter and *CYC1* terminator of the *UBiC* gene. Touch-down PCR (Korbie and Mattick 2008) was used with the annealing temperature decreasing from 65 to 50 °C at 1 °C per cycle, then staying constant at 50 °C for an additional 20 cycles. Denaturation at 95 °C for 30 seconds and extension at 72 °C for 1 minute 20 seconds was used for each cycle. GoTaq 2x master mix (Promega) was used with a total reaction volume of 20 μl in 96-well plates. Known non-producer and producer colonies were used as negative and positive controls, respectively. All PCR products were visualised on 1% agarose gels stained with SYBR safe (LifeTechnologies) and ran at 100 V for 30 minutes.

Table 2. Plasmids used in this study.

Name	Details	Origin
pRS415	Yeast centromeric plasmid, LEU2 marker	Euroscarf [52]
pRS416	Yeast centromeric plasmid, URA3 marker	Euroscarf (Sikorski and Hieter 1989)
pRS425	Yeast 2 micron plasmid, LEU2 marker	Euroscarf (Christianson et al. 1992)
pUG6	Contains G418 resistance marker KanMX	Euroscarf (Güldener et al. 1996)
TR-SSRE	yEGFP containing plasmid	Chen and Weiss (2005)
pPDR12-GFP-415	*pPDR12-yEGFP-ADH1t*-pRS415	This study
pYGP1-GFP-415	*pYGP1-yEGFP-ADH1t*-pRS415	This study
pTPO2-GFP-415	*pTPO2-yEGFP-ADH1t*-pRS415	This study
pPDR12-GFP-425	*pPDR12-yEGFP-ADH1t*-pRS425	This study
U-416	*pTEF1-UBiC-CYC1t*-pRS416	This study
UA-416	*pTEF1-ARO4-CYC1t-pTEF1-UBiC-CYC1t*-pRS416	This study
UA-KanMX-416	*pTEF1-ARO4-CYC1t-pTEF1-UBiC-CYC1t-KanMX*-pRS416	This study
pPDR12-GFP- mCherry-415	*pPDR12-yEGFP-ADH1t-pPDA1-mCherry-CYC1t*-pRS415	This study
mCherry-415	*pPDA1-mCherry-CYC1t*-pRS415	This study

Table 3. Yeast strains used in this study.

Name	Genotype, plasmids	Notes	Origin
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Haploid auxotrophic laboratory strain, mating type 'a'	Euroscarf
BY4742	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0	Haploid auxotrophic laboratory strain, mating type 'α'	Euroscarf
PDR12	BY4742, pPDR12-GFP-415	*PDR12* promoter regulated GFP expression	This study
YGP1	BY4742, pYGP1-GFP-415	*YGP1* promoter regulated GFP expression	This study
TPO2	BY4742, pTPO2-GFP-415	*TPO2* promoter regulated GFP expression	This study
415.C	BY4741, pRS415, pRS416	No GFP control strain	This study
425.C	BY4741, pRS425, pRS416	No GFP control strain	This study
GFP.415	BY4741, pPDR12-GFP-415, pRS416	Centromeric plasmid biosensor strain	This study
GFP.415.U	BY4741, pPDR12-GFP-415, U-416	Centromeric plasmid biosensor strain with *UBiC* expression	This study
GFP.415.UA	BY4741, pPDR12-GFP-415, UA-416	Centromeric plasmid biosensor strain with *UBiC* and *ARO4* expression	This study
GFP.415.UA.KanMX	BY4741, pPDR12-GFP-415, UA-KanMX-416	Centromeric plasmid biosensor strain with *UBiC*, *ARO4*, and *KanMX* expression	This study
GFP.425	BY4741, pPDR12-GFP-425, pRS416	Episomal plasmid biosensor strain	This study
+FB.GFP.415	BY4741, pWAR1::pPDR12, pPDR12-GFP-415, pRS416	Positive feedback, centromeric plasmid biosensor strain	This study
+FB.GFP.415.U	BY4741, pWAR1::pPDR12, pPDR12-GFP-415, U-416	Positive feedback, centromeric plasmid biosensor strain with *UBiC* expression	This study
+FB.GFP.415.UA	BY4741, pWAR1::pPDR12, pPDR12-GFP-415, UA-416	Positive feedback, centromeric plasmid biosensor strain with *UBiC* and *ARO4* expression	This study
+FB.GFP.mCherry.415	BY4741, pWAR1::pPDR12, pPDR12-GFP-pPDA1-mCherry-415, pRS416	Positive feedback, ratiometric centromeric plasmid biosensor strain	This study
+FB.GFP.mCherry.415.UA	BY4741, pWAR1::pPDR12, pPDR12-GFP-pPDA1-mCherry-415, UA-416	Positive feedback, ratiometric centromeric plasmid biosensor strain with *UBiC* and *ARO4* expression	This study

3. Results and discussion

3.1 Identification and testing of organic acid responsive promoters

A literature search identified two native yeast promoters that respond to organic acid bio-products. The first operates via the War1p transcriptional regulator protein, which up-regulates transcription of the membrane acid efflux protein encoding *PDR12* gene (Piper et al. 1998). Although the exact mechanism of War1p activation is incompletely understood, it appears to involve a combination of direct interaction between intracellular dissociated acid molecules and War1p, and phosphorylation of the protein via an unknown mechanism (Kren et al. 2003; Schüller et al. 2004; Gregori et al. 2008). The second weak acid response module results in the up-regulation of expression of the membrane protein encoding *TPO2* and *YGP1* genes from the

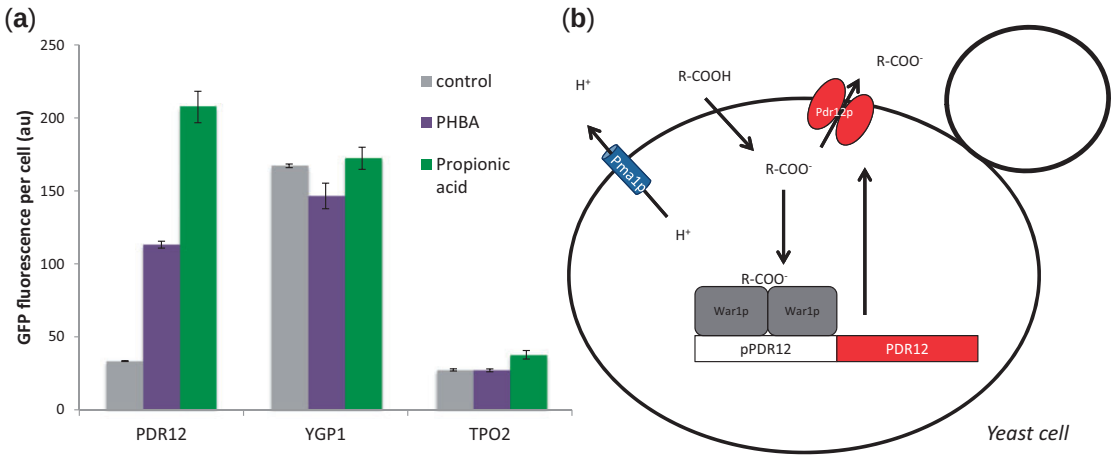


Figure 1. PHBA and PA responsive promoter screening. (a) Three different promoters that were previously reported to be up-regulated in the presence of organic acids (*PDR12*, *YGP1*, and *TPO1*) were used to control GFP expression in cells treated with water (control), 50 mM PHBA, or 50 mM PA. Mean GFP values from triplicate cultures are shown with error bars representing ± 1 SD. Flow cytometry was used to quantify GFP expression after 3 hours of cultivation. (b) The weak acid response system in yeast maintains neutral cytosolic pH via the exportation of organic acid anions and hydrogen ions form the cell. Organic acids can freely diffuse into the cell in their non-dissociated form (R-COOH) when the pH is below their pKa. Once inside the cell they encounter a neutral pH and exist in a dissociated anionic form ($R\text{-}COO^-$) (Holyoak et al. 1999). The hydrogen ions accumulated after acid dissociation are exported from the cell via the Pma1p membrane protein, while acid anions bind to the War1p transcription factor which up-regulates expression of the *PDR12* gene encoding a membrane transporter. Pdr12p then transports organic acid anions outside of the cell (Piper et al. 1998).

Haa1p transcriptional regulator as part of a rapid response to acid stress (Fernandes et al. 2005). We therefore tested the capacity of the *PDR12*, *YGP1*, and *TPO2* promoters to regulate the expression of GFP in response to two industrially relevant organic acids, PHBA and propionic acid (PA) (Fig. 1a). The *YGP1* promoter showed no difference in GFP expression levels when treated with either PHBA or PA, while the *TPO2* promoter displayed slightly up-regulated GFP expression in the presence of PA. In contrast, the *PDR12* promoter up-regulated GFP expression approximately 3.4- and 6.2-fold above the level of the control in the presence of PHBA and PA, respectively. This positive response indicated that the War1p-mediated *PDR12* promoter (Fig. 1b) has potential to act as a biosensor of exogenously added weak organic acids such as PHBA and PA.

3.2 Discrimination of intracellular production levels and genetic variants

Critical characteristics for a metabolite-reporter are to: detect metabolites originating from within the cell; distinguish between a wide range of different levels of cellular production from within a genetically diverse population of rapidly growing cells; and facilitate rapid selection by high-throughput screening methods. The *pPDR12-GFP* biosensor was tested against these performance criteria using PHBA production from the shikimate pathway in yeast (Fig. 2a). PHBA is an aromatic molecule used in liquid crystal polymers, with applications in the electronics and fibre industries and an estimated market of $150 million (USD) per annum (Krömer et al. 2012). It has previously been established that expression of an *E. coli* chorismate pyruvate lyase (*UBiC* gene), and feedback resistant *S. cerevisiae* DAHP synthase results in PHBA production in yeast (Williams et al. 2015, 2016). As a proof-of-concept, the *UBiC* and $ARO4^{Q166K}$ genes were expressed from strong-constitutive *TEF1* promoters in a *pPDR12-GFP* biosensor-containing strain (Fig. 2a). Expression of only the *UBiC* gene resulted in a slight but significant ($P = 0.009$) increase in average GFP expression (Fig. 2b) from 28 to 35 arbitrary units (au). This is highly consistent with previous results that demonstrated a small increase in PHBA production from yeast cultures expressing *UBiC* (26–46 μM) (Williams et al. 2015). The $ARO4^{Q166K}$ enzyme resists feedback inhibition by downstream metabolites in the shikimate pathway (Fukuda et al. 1992; Hartmann et al. 2003), and over-expression of feedback resistant versions of this enzyme is known to result in an approximately 4- to 5-fold increase in shikimate pathway flux (Luttik et al. 2008), and a 6-fold increase in PHBA production (from 46 to 297 μM)[35]. In concordance with these results, we observed a significant ($P = 1.7 \times 10^{-6}$) 3-fold increase in *pPDR12-GFP* expression from cells expressing both *UBiC* and $ARO4^{Q166K}$ genes (Fig. 2b). These results indicate that War1p-mediated *pPDR12-GFP* expression can be modulated by intracellular PHBA levels and is sensitive to variations in metabolic flux though a production pathway.

The second performance objective for a metabolite production biosensor is the capacity to enable selection of productive cells from a mixed population of producers and non-producers (Fig. 2c). The *pPDR12-GFP* biosensor was tested for this characteristic using a simple experiment where cells that contain the genes required for PHBA production (*UBiC* and $ARO4^{Q166K}$) also express the geneticin resistance gene *KanMX*. This strain (GFP.415.UA.KanMX, Table 3) was mixed in a 1:1 ratio with an equivalent biosensor containing strain without PHBA production genes (GFP.415, Table 3). After 3 hours of growth cells within the top 3.4% of GFP fluorescence values were sorted from the population directly onto YPD agar plates. This gate was drawn to visually encompass the top fraction of the population where more events from the producer-only population were likely to occur. To determine if sorted cells actually contain the PHBA production genes, these colonies were replica-plated onto YPD agar containing geneticin (G418), which only producer cells have the capacity to grow on. The majority of sorted colonies (23/35) tested using this method were found to have the G418 resistance and were therefore correctly isolated from the mixed producer/non-producer population based on GFP fluorescence. Conversely, when cells in the bottom 10% of GFP fluorescence values were sorted, as expected none were found to be producers (0/35 able to grow on YPD-G418 plates). This experiment was carried out in growth medium at a pH of 6.5 so that any PHBA molecules (pKa = 4.54) excreted by producer cells exist in their dissociated form outside of the cell, and are unable to

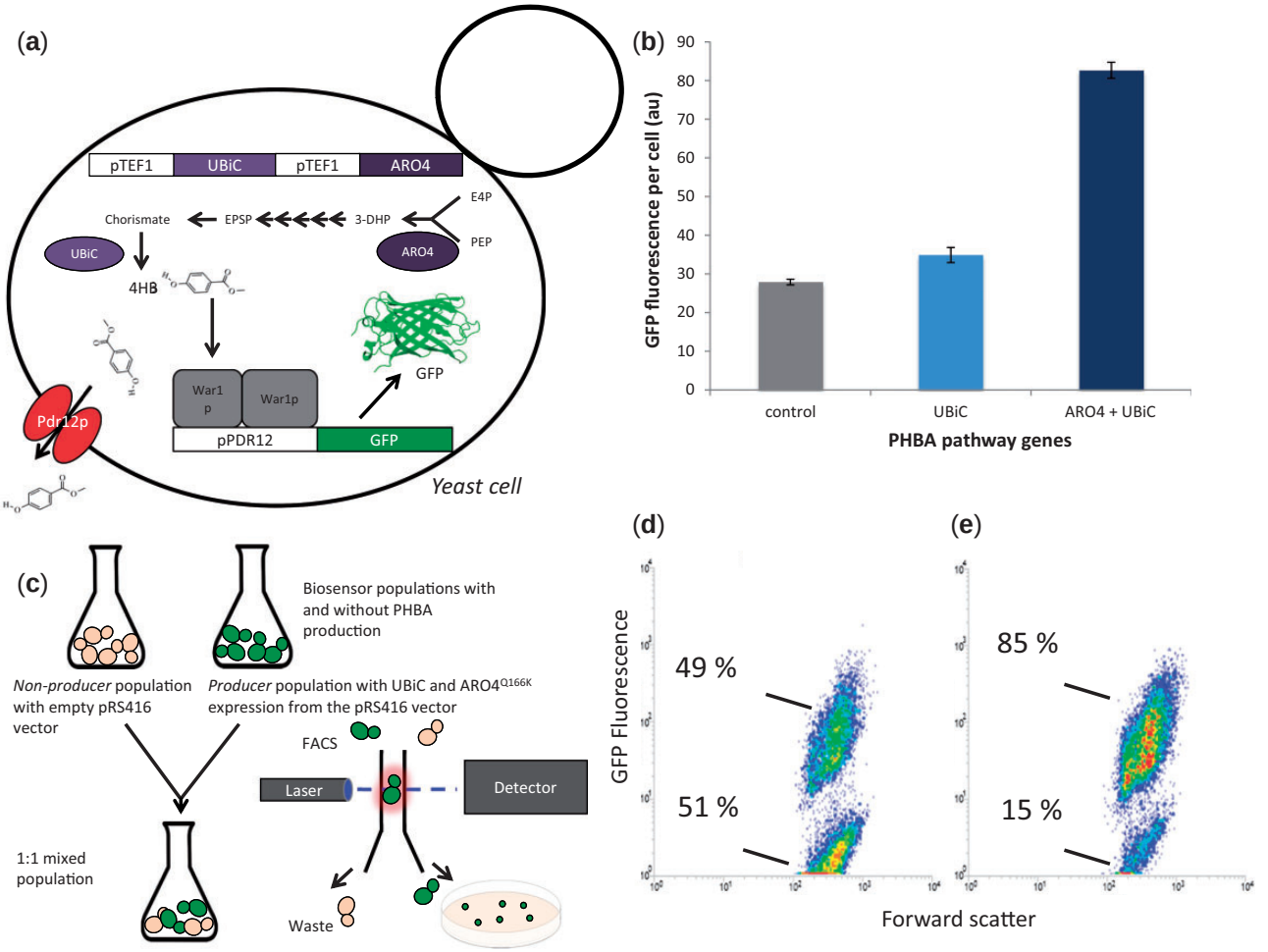


Figure 2. Detection of intracellular PHBA and shikimate pathway flux variants. (a) The expression of chorismate pyruvate lyase and 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (*UBiC* and *ARO4* genes) from strong *TEF1* promoters in yeast enables the production of PHBA (Williams et al. 2016). The *pPDR12-GFP* biosensor was tested for responsiveness to intracellular PHBA production via co-expression in a PHBA producing strain. (b) The GFP fluorescence of *pPDR12-GFP* biosensor containing strains increases with increasing PHBA production. Mean GFP values from triplicate cultures are shown with error bars representing ± 1 SD. (c) As a proof of principle, equal amounts of biosensor containing PHBA producers and non-producers were mixed in the same culture and cells were sorted onto agar plates according to GFP fluorescence. (d and e) Flow cytometry density plots are shown with GFP fluorescence on the y-axis and forward scatter on the x-axis. (d) A non-PHBA producing strain (GFP.415, Table 3). (e) The fluorescence level of a strain with *TEF1* promoter-mediated expression of *UBiC*, *ARO4*^{Q166K}, and *KanMX* (strain GFP.415.UA.KanMX, Table 3) for PHBA production.

diffuse into other cells that are non-producers (Holyoak et al. 1999), thereby preventing GFP expression that would result in the sorting of false positives. These results validate the use of the War1p-mediated *pPDR12-GFP* system as a biosensor for War1p interacting organic acid metabolites produced from engineered metabolic pathways in yeast.

An interesting feature of the *pPDR12-GFP* biosensor populations was the high noise level of the non-producer control strain, as assessed by flow cytometry (Fig. 2d). This strain was cultured without PHBA production pathway genes, and without the addition of PHBA to the growth medium, yet 49% of the cells had the same high level of GFP expression seen in the producer population (Fig. 2d). Furthermore there were two sub-populations, with most cells (51%) residing in a low-GFP state that is equivalent to yeast auto-fluorescence. When biosensor-containing cells also had *UBiC* and *ARO4*^{Q166K} genes for PHBA production the majority of cells (85%) existed in the high-GFP state (Fig. 2d). By only sorting cells in the top 3.4% of fluorescence levels, it was possible to isolate producers based on GFP expression with 66% accuracy (23/35 colonies able to grow on YPD-G418 plates). This relatively low level of sorting accuracy probably arose because the gate used to isolate producers overlapped significantly with the non-producer population. A greater separation of producer and non-producer fluorescence levels would enable the use of a gate that excluded all GFP fluorescence levels observed in the non-producer population. With these strains this is not possible (Fig. 2d–e), making the system impractical for further high-throughput screening applications. Ideally, a much higher separation of producer-GFP expression from non-producer-GFP expression would exist so that producers could be more easily isolated. This led us to consider ways of improving the dynamic range of GFP expression, and reducing the noise levels of the *pPDR12-GFP* biosensor.

3.3 Fine-tuning sensor dynamic range using positive feedback

An ideal biosensor has a large dynamic-range of output levels (e.g. GFP expression), low non-induced expression, and is effective at differentiating between a wide-range of input concentrations. The *pPDR12-GFP* biosensor was tested for dose-dependent response to externally applied PHBA with *pPDR12-GFP*

expression from low- or high-copy vectors (pRS415 or pRS425) (Fig. 3a). In each case, there was a dose-dependent increase in GFP expression, with a 2.1-fold dynamic range from the high-copy vector and a 2.9-fold dynamic range with the low-copy vector, and a response range between 10 and 75 mM PHBA. Increasing GFP copy-number via expression from the pRS425 vector had little effect on these parameters and served primarily to increase the uninduced GFP expression level (Fig. 3a). The fold-change observed in this native system was not ideal for use as a high-throughput screening biosensor, we therefore sought to implement genetic modifications in the yeast weak acid response module that would increase the dynamic range and sensitivity of the biosensor output.

Positive feedback loops are commonly used synthetic biology devices that can reduce non-specific expression while increasing dynamic range and sensitivity (Ingolia et al. 2007; Williams et al. 2013). In order to implement a positive feedback loop in the weak acid response system in yeast, we replaced the native promoter that regulates the expression of the War1p transcription factor with its target *PDR12* promoter (Fig. 3b) using CRISPR-Cas9-mediated homologous recombination (see Methods section for details). With this re-configuration the War1p transcription factor should in theory only be expressed at low-levels from 'leaky' non-induced *PDR12* promoter expression in a cell without weak acid production/exposure. Upon response to weak acid molecules, the amount of War1p available for GFP expression induction from the *PDR12* promoter should sharply increase as War1p starts to induce its own expression as part of a positive-feedback loop (Fig. 3b). When PHBA and PA dose–response experiments (Fig. 3c and d) were carried out on cells with positive feedback War1p expression, there was a significant increase in the dynamic range of GFP expression. The dynamic range increased from to 2.9- to 4.2-fold in the presence of PHBA and from 3.6- to 10-fold with PA (Fig. 3c and d). The basal non-induced GFP-expression levels were also reduced by between 7 and 19% due to positive-feedback War1p expression, as expected from the positive-feedback model (zero acid GFP values not shown in Fig. 3a–d due to the log scale on the x-axis).

In theory, positive feedback expression of War1p results in a lower basal-concentration of GFP in the absence of inducer, and a higher maximum expression-level upon induction, resulting in a higher dynamic range. While our findings are consistent with this model, it is also possible for positive-feedback loops to alter the timing and rate of gene expression (Ingolia et al. 2007; Williams et al. 2013). Therefore, another possible explanation for our observations could be that the dynamic ranges of the native and positive feedback War1p expression systems are actually the same, but that the positive-feedback system is simply induced faster. We therefore tested the *pPDR12-GFP* expression levels from both systems after PA induction over 7 hours (Fig. 4). Although the initial rate of increase in GFP expression was much higher in the positive-feedback strain (0 to 2.5 hours,

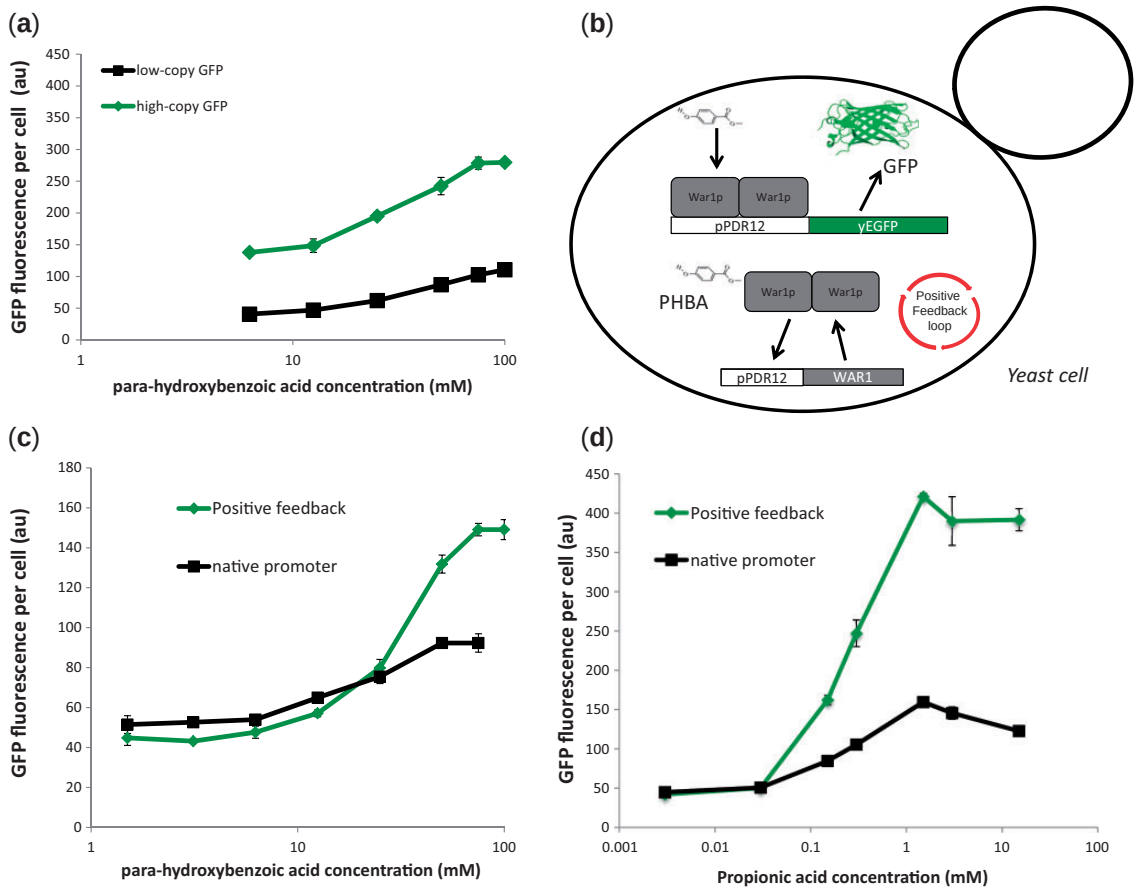


Figure 3. Tuning biosensor output using positive feedback. (a) PHBA dose response curves for strains with *pPDR12-GFP* expression form low- or high-copy vectors (pRS415 or pRS425). (b) Circuit configuration of a positive feedback biosensor. The native WAR1 promoter was replaced with the *PDR12* promoter such that the organic acid anion responsive War1p transcription factor regulates its own expression as part of a positive feedback loop, in addition to *pPDR12-GFP* expression. *pPDR12-GFP* expression levels in strains with the native *WAR1* promoter or positive-feedback *pPDR12-WAR1* expression in response to PHBA (c) or propionic acid (d). Mean GFP values and SDs from triplicate cultures are shown.

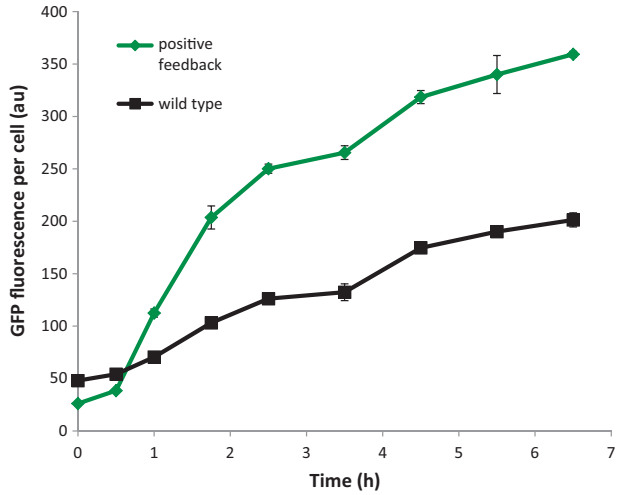


Figure 4. Time-course biosensor induction. Average GFP expression-levels of native (wild-type, black squares) and positive-feedback (green diamonds) War1p expression systems. Each strain was induced with 1.5 mM of propionic acid at time zero. Mean GFP values and standard deviations from triplicate cultures are shown.

Fig. 4), a constant difference of ~130 au was reached after 2.5 hours and the high-GFP state of the positive-feedback system was not attained by the native one within the time-frame of this experiment. These observations further demonstrate the utility of the positive feedback system for reducing basal expression levels and increasing dynamic range, and validate the use of the 3-hour sampling point that we employed for other experiments in this study.

3.4 Controlling noise using a ratiometric biosensor

High noise-levels in gene expression across a population are a universal feature of biological systems (Elowitz et al. 2002), and a constant bane to synthetic biologists. Gene expression noise can be attributed to extrinsic factors such as differences in cell size and cycle, substrate uptake rates and metabolic fluxes, as well as variations in the concentrations of other biomolecules that affect a cell's capacity to express genes such as; plasmid copy number, transcription factors, RNA polymerase, ribosomes, and ATP (Elowitz et al. 2002). Although variations in plasmid copy-number are known to contribute significantly to noise levels (Zhang et al. 2016), we chose to use the low-copy yeast vector pRS415 for biosensor expression instead of genomic integration to facilitate rapid prototyping of different biosensors. Even with genomic integration of reporter systems such as GFP, there are still significant levels of noise due to stochastic variations in the concentrations of other biomolecules between individual cells in a population (Elowitz et al. 2002). In the context of biosensor-mediated cell sorting, noisiness in gene expression can mean that the throughput of screening is greatly reduced. Although a responsive population may have a much higher average GFP level, there can be significant overlap between ligand-responsive and control populations. Populations can be gated such that only cells with GFP levels higher than a non-productive control population are sorted, but if there is too much noise then the number of cells that are available for selection is dramatically reduced. Mitigating the effects of highly variable gene expression is therefore of critical importance to biosensor-mediated high-throughput screening. One potential way to achieve this is to 'normalise' the level of biosensor output (GFP) to another constitutively expressed fluorescent protein. Because the second fluorescent protein is constitutively regulated, its expression level should serve as a proxy for aspects of cellular physiology that contribute to gene expression noise. In theory, cells can then be selected based on the ratio of biosensor-mediated GFP expression to constitutive fluorescent protein expression within a single cell.

In order to reduce extrinsic noise in biosensor output and increase screening throughput, we converted our positive-feedback biosensor into a ratiometric sensor. This was achieved by expressing a red fluorescent protein (mCherry) from the weak-constitutive *PDA1* promoter (Peng e al. 2015) alongside the weak-acid responsive *pPDR12-GFP* biosensor on the same plasmid (pPDR12-GFP- mCherry-415, Table 2), in the same strain (+FB.GFP.mCherry.415, Table 3) (Fig. 5a). By plotting mCherry fluorescence on the x-axis and GFP fluorescence on the y-axis, the ratio of GFP to mCherry in individual cells can be measured using flow-cytometry (Fig. 5b). In theory, the level of mCherry reflects a cell's general capacity for gene expression, and enables the inclusion of low GFP expressing, but organic acid responding/producing cells in high-throughput screening. The fact that there is a positive linear relationship between mCherry fluorescence and GFP fluorescence (Fig. 5b) supports the idea that a constitutively expressed fluorescent protein can be used as a kind of 'internal standard' for a cell's gene expression capacity. When a ratiometric biosensor strain was treated with and without a saturating concentration of PA (1.5 mM) and analysed using only GFP fluorescence, there was a large overlap between the groups with approximately 34% of the cells in the treated population within the range of GFP fluorescence values observed in the non-treated population (Fig. 5c). When the same comparison was made using the ratio of GFP to mCherry within each cell, the overlap between the two populations was approximately 1% (Fig. 5d). This demonstrated the efficiency of the ratiometric approach for controlling noise in the biosensor population. The same trend was observed with PHBA treatment, although there was a less pronounced effect (Figure S1 in the Supplementary Data). This is consistent with the previously observed weaker biosensor induction using PHBA relative to PA (Fig. 3c and d).

The concept of using ratiometric fluorescence normalisation to increase the signal to noise ratios of sensors has existed for some time (Demchenko et al. 2010) and has most commonly been exploited in the form of Forster Resonance Energy Transfer (FRET) systems where a different emission spectrum results from close fluorophore proximity. FRET systems have previously been used as *in vivo* metabolite biosensors (Michener et al. 2012), but are yet to be used for high-throughput screening of producer-cells. Similarly, ratiometric fluorescent protein normalisation was recently employed by Zhang et al (2016) to improve the signal-to-noise ratio of an *in vivo* NADP+/NADPH reporter with great success. However, the ratiometric approach was not used for high-throughput metabolite producer screening, and as far as we are aware, ratiometric biosensor expression has not previously been applied to high-throughput metabolite-producer screening. Given the efficiency of ratiometric fluorescent protein normalisation at reducing the signal to noise ratio of biosensor output, we sought to explore how this mode of biosensor expression affects the efficiency of high-throughput screening.

The decrease in overlap that we observed between treated and non-treated biosensor populations (Fig. 5c and d) should also result in increased high-throughput screening power, as a greater proportion of a given population is available for selection using gates that exclude control population fluorescence

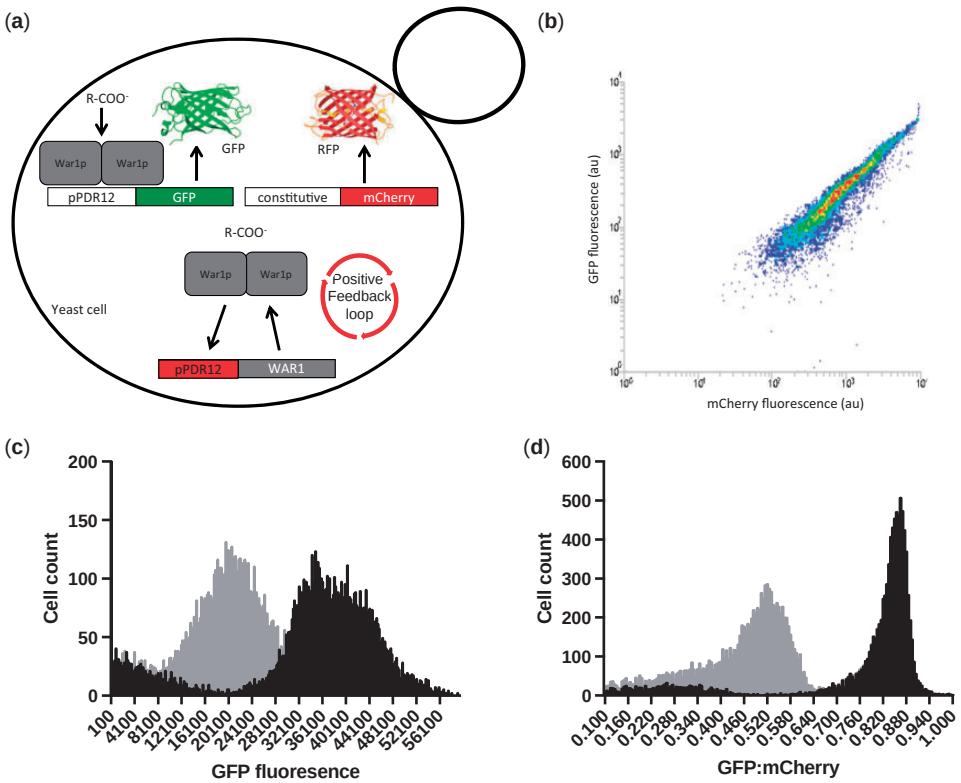


Figure 5. Enhanced population separation using a ratiometric biosensor. (a) Positive-feedback, ratiometric biosensor configuration (+FB.GFP.mCherry.415 strain, Table 3) whereby War1p-mediated biosensor output (GFP expression) is normalised to constitutively expressed mCherry in each individual cell. (b) GFP:mCherry fluorescence distribution of the positive-feedback ratiometric biosensor strain (+FB.GFP.mCherry.415, Table 3). (c) The same strain treated with (black bars) and without (grey bars) 1.5 mM propionic acid is analysed using only GFP fluorescence or (d) using the ratio of GFP to mCherry fluorescence within each cell. Data from single cultures which are representative of repeated experiments are shown here.

values. The producer-screening efficiency of the positive-feedback-ratiometric biosensor was therefore tested by attempting to sort producer cells mixed with non-producers at a variety of ratios. The success of fluorescence based PHBA-producer sorting was determined using PCR with a forward primer specific to the *TEF1* promoter and a reverse primer specific to the *CYC1* terminator region of the *UBiC* gene expression cassette, on DNA extracted from single colonies arising from sorted cells on agar plates.

Mixed populations with producer to non-producer ratios varying from 1:1 through to $1:1\times10^5$ were used to test the limits of our biosensor's sorting accuracy. When pure producer and non-producer populations from the initial non-positive-feedback, and positive-feedback biosensor strains were compared it was not possible to make a sorting gate in the producer population that did not overlap with the non-producer population, as previously observed (Fig. 2d and e). In contrast, when the GFP:mCherry fluorescence levels of producer and non-producer strains containing the positive-feedback ratiometric biosensor (strains +FB.GFP.mCherry.415.UA and +FB.GFP.mCherry.415, Table 3) were compared there was a clear separation of fluorescence levels that could be used to define a sorting gate (Fig. 6a). It should be noted that we used a high-power blue laser (200 mW, 488 nm) to excite mCherry (excitation maximum 587 nm) so it is likely that using a green laser, which is common option on many flow sorters, may significantly improve the separation of producers from non-producers. Furthermore, when mixed producer/non-producer populations were visualised using density plots after 3 hours of co-culturing, the two populations were still clearly distinguishable (Fig. 6a). This again indicates that the technique of culturing mixed populations at a pH above the pKa of the secreted organic acid product (pH 6.5 and PHBA pKa 4.54 in this case) prevents the PHBA molecules produced by one cell entering another cell in the population and potentially activating the biosensor of a non-producer. This would lead to an averaging-out of the fluorescence levels observed in the two pure populations, and was not observed here (Fig. 6a). When cells were sorted from mixed populations ranging from 1:1 to $1:10^5$ producer to non-producers, the positive-feedback ratiometric biosensor was highly efficient at enabling the correct identification of producers based on GFP:mCherry ratios (Fig. 6b). 100% of isolates from the 1:1 through to $1:10^2$ producer:non-producer populations were confirmed as producers via PCR. Sorting accuracy dropped to ~88% at $1:10^3$ and to ~66% at $1:10^4$, while no producers could be correctly identified at $1:10^5$ using this strategy (Fig. 6b). This a dramatic increase in high-throughput screening power when compared with the original non-positive feedback non-ratiometric version of the biosensor (strain GFP.415, Table 3), which only enabled 66% sorting accuracy with equal amounts of producers and non-produces in a mixed population (Fig. 2c–e). The 66% sorting accuracy we observed with the positive-feedback, ratiometric biosensor therefore represents a 5,000-fold improvement in high-throughput screening power. The magnitude of this improvement can partly be attributed to the poor-performance of the original War1p-pPDR12 system as a biosensor. It is therefore likely that positive-feedback, ratiometric biosensor expression would result in less dramatic

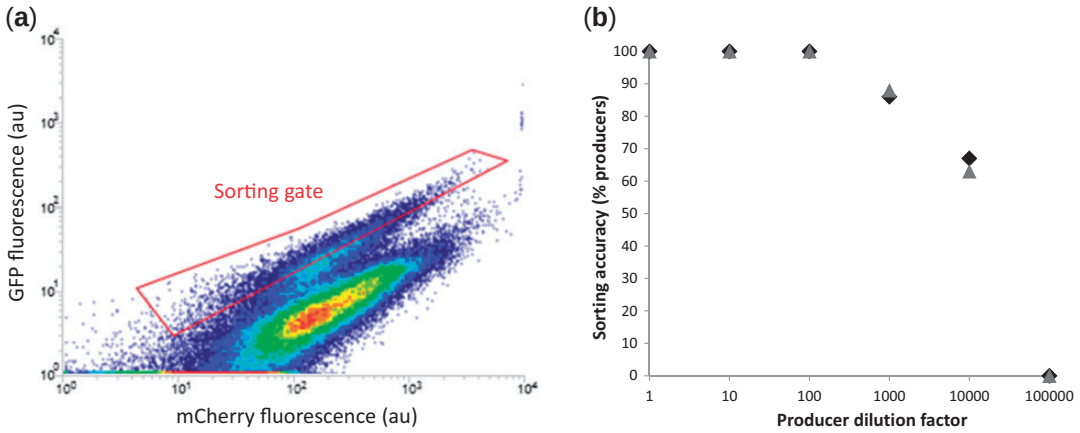


Figure 6. Positive-feedback, ratiometric biosensor mixed population cell sorting. Strains with positive-feedback *WAR1* expression, *pPDR12-GFP* expression and constitutive *PDA1* promoter-mediated *mCherry* expression, containing either the control pRS416 plasmid or the *UBiC-ARO4* expression plasmid were mixed together in different ratios to test the efficiency of biosensor mediated high-throughput screening. a) Flow cytometry density plot of PHBA producers mixed with non-producers at a ratio of 1 in 10. One hundred thousand events were recorded with increasing density indicated by the transition through blue, green, yellow, and red with mCherry (x-axis) and GFP expression (y-axis) measured simultaneously for each cell. The red coloured 'sorting gate' indicates the range of fluorescence values used to sort producers from non-producers. **b**) Producers were diluted with non-producers by the indicated factors and cultured together for 3 hours prior to sorting cells through the gate depicted in (a) onto agar plates. DNA extracted from colonies arising from individual sorted cells was used to test for the presence of the *UBiC* gene using PCR. Colonies giving rise to amplification were identified as being correctly sorted cells based on biosensor fluorescence. Between three and eight colonies were analysed for each dilution level with the proportion of producers identified expressed as a percentage (y-axis). The experiment was done in biological duplicate on different days, with raw data plotted (replicate 1 in black diamonds, replicate 2 in grey triangles).

fold-improvements of biosensors that have inherently lower noise-levels and higher dynamic ranges.

These results demonstrate the utility of the positive-feedback ratiometric biosensor approach for improving high-throughput screening efficiency via controlling for noise, reducing basal biosensor output, and increasing dynamic range. When applying this sensor to isolate producer cells from randomly mutated or evolving populations, the frequency of cells containing mutations that encode for increased metabolite production are likely to be much lower than 1 in 10^4, the limit of accuracy for our biosensor. However it has recently been demonstrated that repeated cycles of enrichment for cells with high biosensor output using FACS can be used to concentrate high producers over time and identify rare mutational events that lead to higher productivity (Mahr et al. 2015). We envisage that a similar process of iterative growth and FACS enrichment could be successfully employed to evolve PHBA or PA-producing yeast using our biosensor. The highest previously recorded titre of PHBA in yeast is ~ 1 mM, and this level was achieved by simultaneously expressing three pathway enzymes (*UBiC*, *ARO4*, *TKL1*) and dynamically repressing two enzymes (*ARO7*, *CDC19*) that compete for carbon flux (Williams et al. 2015). Although the dynamic range of our biosensor is nearly saturated via the expression of *UBiC* and *ARO4* genes (Fig. 2b), it should be possible to discover novel mutations that affect shikimate pathway flux by carrying out biosensor-mediated directed evolution experiments beginning with only *UBiC* gene expression. In theory this approach should enable the exploration of evolutionary trajectories that are independent of the traditional feedback resistant *ARO4* enzyme (Luttik et al. 2008). It is also possible that this biosensor could be used to screen PHBA production (or other organic acids) from other cells in microtiter plates as part of a co-culture system. PHBA in particular has been produced at high levels in *E. coli* (12 g/L) (Barker and Frost 2001), *Klebsiella pneumonia*, and *Pseudomonas putida* (317 mg/L) (Verhoef et al. 2010), and this biosensor has the potential to be used to screen modified or mutated strains of these producers, albeit at a much lower throughput compared with FACS.

4. Summary and Conclusions

Ligand-responsive transcriptional regulators and their cognate promoters are widely used tools in synthetic biology (Taylor et al. 2016), and are becoming increasingly valuable as metabolite biosensors for high-throughput strain screening and dynamic pathway regulation in metabolic engineering (Zhang and Keasling 2011; Liu et al. 2015; Mahr et al. 2016; Rogers et al. 2016; Williams et al. 2016). We have demonstrated that positive-feedback biosensor expression significantly reduces the basal expression level and increases the dynamic range. Furthermore ratiometric fluorescent protein 'normalisation' within single cells provides a control system for extrinsic noise, increasing the efficiency and accuracy of high-throughout strain screening. By combining positive-feedback biosensor expression with ratiometric fluorescence normalisation, high-throughput screening efficiency was improved 5,000-fold. These phenomena were demonstrated using the *WAR1* organic acid responsive transcriptional regulator in yeast, which holds great promise for use as a biosensor for isolating highly productive organic acid production strains from randomly mutated or evolving populations. Due to the fact that gene expression noise is a universal feature of biological systems (Elowitz et al. 2002; Keren et al. 2015) the design principles of positive-feedback, ratiometric biosensor expression are likely to be relevant to many other transcription-factor promoter pairs used in synthetic biology.

Author Contributions

Conceived of the project (T.C.W.), designed the experiments (T.C.W., I.T.P.), performed the experiments (T.C.W., X.X.), analysed the data (T.C.W.), trained T.C.W. in flow cytometry techniques and assisted with experimental design (M.O.), wrote the manuscript (T.C.W., I.T.P.), edited and reviewed the manuscript (T.C.W., X.X., M.O., I.S.P., I.T.P.), project support for T.C.W. and X.X. (I.S.P., I.T.P.).

Acknowledgements

The Synthetic Biology initiative at Macquarie University is financially supported by an internal grant from the University, and external grants from Bioplatforms Australia, the New South Wales (NSW) Chief Scientist and Engineer, and the NSW Government's Department of Primary Industries. Ian Paulsen is supported by an Australian Research Council Laureate Fellowship.

Data availability

All raw data are available with the online version of this article.

Supplementary data

Supplementary data are available at *Synthetic Biology* online.

Conflict of interest: None declared.

References

Abbott, D. A., Zelle, R. M., Pronk, J. T., and van Maris, A. J. A. (2009) 'Metabolic Engineering of *Saccharomyces cerevisiae* for Production of Carboxylic Acids: Current Status and Challenges', *FEMS Yeast Research*, 9: 1123–36.

Agren, R., Otero, J. M., and Nielsen, J. (2013) 'Genome-Scale Modeling Enables Metabolic Engineering of *Saccharomyces cerevisiae* for Succinic Acid Production', *Journal of Industrial Microbiology and Biotechnology*, 40: 735–47.

Almario, M. P., Reyes, L. H., and Kao, K. C. (2013) 'Evolutionary Engineering of *Saccharomyces cerevisiae* for Enhanced Tolerance to Hydrolysates of Lignocellulosic Biomass', *Biotechnology Bioengineering* 110: 2616–23.

Averesch, N. J. H., and Krömer, J. O. (2014) 'Tailoring Strain Construction Strategies for Muconic Acid Production in *S. cerevisiae* and *E. coli*', *Metabolic Engineering Communications*, 1: 19–28.

Baek, S. H., Kwon, E. Y., Kim, Y. H., and Hahn, J. S. (2015) 'Metabolic Engineering and Adaptive Evolution for Efficient Production of D-Lactic Acid in *Saccharomyces cerevisiae*', *Applied Microbiology and Biotechnology* 100: 2737–48

Barker, J. L., and Frost, J. W. (2001) 'Microbial Synthesis of p-Hydroxybenzoic Acid From Glucose', *Biotechnology Bioengineering*, 76: 376–90.

Blazeck, J. et al. (2014) 'Metabolic Engineering of *Saccharomyces cerevisiae* for Itaconic Acid Production', *Applied Microbiology and Biotechnology*, 98: 8155–64.

Borodina, I. et al. (2015) 'Establishing a Synthetic Pathway for High-Level Production of 3-Hydroxypropionic Acid in *Saccharomyces cerevisiae* via β-Alanine', *Metabolic Engineering*, 27: 57–64.

Brennan, T. C. et al. (2015) 'Evolutionary Engineering Improves Tolerance for Replacement Jet Fuels in *Saccharomyces cerevisiae*', *Applied and Environmental Microbiology*, 81: 3316–25.

Caspeta, L. et al. (2014) 'Altered Sterol Composition Renders Yeast Thermotolerant', *Science*, 346: 75–8.

Chen, Y., and Nielsen, J. (2016) 'Biobased Organic Acids Production by Metabolically Engineered Microorganisms', *Current Opinion in Biotechnology*, 37: 165–72.

Chen, M.-T., and Weiss, R. (2005) 'Artificial Cell-Cell Communication in Yeast *Saccharomyces cerevisiae* Using Signaling Elements From *Arabidopsis thaliana*', *Nature Biotechnology*, 23: 1551–5.

Christianson, T. W. et al. (1992) 'Multifunctional Yeast High-Copy-Number Shuttle Vectors, *Gene* 110: 119–22

Curran, K. A., Leavitt, J. M., Karim, A. S., and Alper, H. S. (2013) 'Metabolic Engineering of Muconic Acid Production in *Saccharomyces cerevisiae*', *Metabolic Engineering*, 15: 55–66.

Demchenko, A. P. (2010) 'The Concept of λ-Ratiometry in Fluorescence Sensing and Imaging', *Journal of Fluorescence*, 20: 1099–128.

DiCarlo, J. E. et al. (2013) 'Genome Engineering in *Saccharomyces cerevisiae* Using CRISPR-Cas Systems', *Nucleic Acids Research* 41: 4336–43.

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) 'Stochastic Gene Expression in a Single Cell', *Science*, 297: 1183–6.

Fernandes, A. R. et al. (2005) '*Saccharomyces cerevisiae* Adaptation to Weak Acids Involves the Transcription Factor Haa1p and Haa1p-Regulated Genes', *Biochemical and Biophysical Research Communications*, 337: 95–103.

Fukuda, K., Asano, K., Ouchi, K., and Takasawa, S. (1992) 'Feedback-Insensitive Mutation of 3-Deoxy-d-Arabino-Hepturosonate-7-Phosphate Synthase Caused by a Single Nucleotide Substitution of ARO4 Structural Gene in *Saccharomyces cerevisiae*', *Journal of Fermentation and Bioengineering*, 74: 117–9.

Garcia Sanchez, R. et al. (2010) 'Improved Xylose and Arabinose Utilization by an Industrial Recombinant *Saccharomyces cerevisiae* Strain Using Evolutionary Engineering', *Biotechnology for Biofuels* 3: 13.

Gallegos, M. T. et al. (1997) 'Arac/XylS Family of Transcriptional Regulators', *Microbiology and Molecular Biology Reviews*, 61: 393–410.

Gibson, D. G. et al. (2008) 'One-Step Assembly in Yeast of 25 Overlapping DNA Fragments to Form a Complete Synthetic *Mycoplasma genitalium* Genome', *Proceedings of the National Academy of Sciences* 105: 20404–9.

—— et al. (2009) 'Enzymatic Assembly of DNA Molecules Up To Several Hundred Kilobases', *Nature Methods*, 6: 343–5.

Gietz, R. D., and Schiestl, R. H. (2007) 'Quick and Easy Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method', *Nature Protocols*, 2: 35–7

Goodarzi, H. et al. (2010) 'Regulatory and Metabolic Rewiring During Laboratory Evolution of Ethanol Tolerance in *E. coli*', 6: 378.

Gregori, C. et al. (2008) 'Weak Organic Acids Trigger Conformational Changes of the Yeast Transcription Factor War1 in Vivo to Elicit Stress Adaptation, *Journal of Biological Chemistry* 283: 25752–64.

Güldener, U. et al. (1996) 'A New Efficient Gene Disruption Cassette for Repeated Use in Budding Yeast', *Nucleic Acids Research*, 24: 2519–24.

Hartmann, M. et al. (2003) 'Evolution of Feedback-Inhibited β/α Barrel Isoenzymes by Gene Duplication and a Single Mutation,' *Proceedings of the National Academy of Sciences* 100: 862–7.

Holyoak, C. D. et al. (1999) 'The *Saccharomyces cerevisiae* Weak-Acid-Inducible ABC Transporter Pdr12 Transports Fluorescein and Preservative Anions from the Cytosol by an Energy-Dependent Mechanism', *Journal of Bacteriology*, 181: 4644–52

Ingolia, N. T., and Murray, A. W. (2007) 'Positive-Feedback Loops as a Flexible Biological Module' *Current Biology* 17: 668–77.

Ishida, N. et al. (2006) 'Metabolic Engineering of *Saccharomyces cerevisiae* for Efficient Production of Pure L-(+)-Lactic Acid', *Applied Biochemistry and Biotechnology*, 129–132: 795–807.

Ito, Y., Hirasawa, T., and Shimizu, H. (2014) 'Metabolic Engineering of *Saccharomyces cerevisiae* to Improve Succinic Acid Production Based on Metabolic Profiling', *Bioscience, Biotechnology, and Biochemistry*, 78: 151–9.

Kearse, M. et al. (2012) 'Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data', *Bioinformatics*, 28: 1647–9.

Keren, L. et al. (2015) 'Noise in Gene Expression Is Coupled to Growth Rate, *Genome Research*.

Kildegaard, K. R., Wang, Z., Chen, Y., Nielsen, J., and Borodina, I. (2015) 'Production of 3-Hydroxypropionic Acid From Glucose and Xylose by Metabolically Engineered *Saccharomyces cerevisiae*', *Metabolic Engineering Communications*, 2: 132–6.

—— et al. (2014) 'Evolution Reveals a Glutathione-Dependent Mechanism of 3-Hydroxypropionic Acid Tolerance', *Metabolic Engineering*, 26: 57–66.

Korbie, D. J., and Mattick, J. S. (2008) 'Touchdown PCR for Increased Specificity and Sensitivity in PCR Amplification', *Nature Protocols*, 3: 1452–6.

Kren, A. et al. (2003) 'War1p, a Novel Transcription Factor Controlling Weak Acid Stress Response in Yeast', *Molecular Cell Biology*, 23: 1775–85.

Krömer, J. O. et al. (2012) 'Production of Aromatics in *Saccharomyces cerevisiae*—A Feasibility Study', *Journal of Biotechnology* 163: 184–93.

Nielsen, J., and Keasling, J. D. (2016) 'Engineering Cellular Metabolism', *Cell*, 164; 1185–97.

Liu, D., Evans, T., and Zhang, F. (2015) 'Applications and Advances of Metabolite Biosensors for Metabolic Engineering', *Metabolic Engineering*, 31; 35–43.

Luttik, M. A. H. et al. (2008) 'Alleviation of Feedback Inhibition in *Saccharomyces cerevisiae* Aromatic Amino Acid Biosynthesis: Quantification of Metabolic Impact', '*Metabolic Engineering*, 10: 141–53.

Mahr, R. et al. (2015) 'Biosensor-Driven Adaptive Laboratory Evolution of l-Valine Production in *Corynebacterium glutamicum*', *Metabolic Engineering*, 32: 184–94.

Mahr, R., and Frunzke, J. (2016) 'Transcription Factor-Based Biosensors in Biotechnology: Current State and Future Prospects', *Applied Microbiology and Biotechnology*, 100: 79–90.

Michener, J. K., Thodey, K., Liang, J. C., and Smolke, C. D. (2012) 'Applications of Genetically-Encoded Biosensors for the Construction and Control of Biosynthetic Pathways', *Metabolic Engineering*, 14: 212–22.

Patil, K. R., Åkesson, M., and Nielsen, J. (2004) 'Use of Genome-Scale Microbial Models for Metabolic Engineering', *Current Opinion in Biotechnology*, 15: 64–9.

Peng, B. et al. (2015) 'Controlling Heterologous Gene Expression in Yeast Cell Factories on Different Carbon Substrates and Across the Diauxic Shift: A Comparison of Yeast Promoter Activities', *Microbial Cell Factories*, 14: 91.

Piper, P. et al. (1998) 'The pdr12 ABC Transporter Is Required for the Development of Weak Organic Acid Resistance in Yeast', *The EMBO Journal*, 17: 4257–265.

Pretorius, I. S. (2016) 'Synthetic Genome Engineering Forging New Frontiers for Wine Yeast', *Critical Reviews in Biotechnology* 37: 112–36.

Quan, S. et al. (2012) 'Adaptive Evolution of the Lactose Utilization Network in Experimentally Evolved Populations of *Escherichia coli*', *PLoS Genet*. 8: e1002444.

Raab, A. M. et al. (2010) 'Metabolic Engineering of *Saccharomyces cerevisiae* for the Biotechnological Production of SUCCINIC Acid', *Metabolic Engineering*, 12: 518–25.

Ramos, J. L. et al. (2005) 'The TetR Family of Transcriptional Repressors', *Microbiology and Molecular Biology Reviews*, 69: 326–56.

Riehle, M. M., Bennett, A. F., Lenski, R. E., and Long, A. D. (2003) 'Evolutionary Changes in Heat-Inducible Gene Expression in Lines of *Escherichia coli* Adapted to High Temperature', *Physiological Genomics*, 14: 47–58.

Rogers, J. K., Taylor, N. D., and Church, G. M. (2016) 'Biosensor-Based Engineering of Biosynthetic Pathways', *Current Opinion in Biotechnology*, 42: 84–91.

Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, 1, pp. 11–258. New York, NY: Cold Spring Harbor Laboratory Press.

Sauer, M., Porro, D., Mattanovich, D., and Branduardi, P. (2008) 'Microbial Production of Organic Acids: Expanding the Markets', *Trends in Biotechnology*, 26: 100–8.

Schüller, C. et al. (2004) 'Global Phenotypic Analysis and Transcriptional Profiling Defines the Weak Acid Stress Response Regulon in *Saccharomyces cerevisiae*', *Molecular Biology of the Cell*, 15: 706–20.

Sikorski, R. S., and Hieter, P. (1989) 'A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in *Saccharomyces cerevisiae*', *Genetics*, 122: 19–27.

Shao, Z., Zhao, H., and Zhao, H. (2009) 'DNA Assembler, an In Vivo Genetic Method for Rapid Construction of Biochemical Pathways, *Nucleic Acids Research*, 37: e16–e16.

Taylor, N. D. et al. (2016) 'Engineering an Allosteric Transcription Factor to Respond to New Ligands', *Nature Methods* 13: 177–83.

Tropel, D., and van der Meer, J. R. (2004) 'Bacterial Transcriptional Regulators for Degradation Pathways of Aromatic Compounds', *Microbiology and Molecular Biology Reviews*, 68: 474–500,

Verhoef, S. et al. (2010) 'Comparative Transcriptomics and Proteomics of p-Hydroxybenzoate Producing *Pseudomonas putida* S12: Novel Responses and Implications for Strain Improvement', *Applied Microbiology and Biotechnology*, 87: 679–90.

Wiechert, W. (2002) 'Modeling and Simulation: Tools for Metabolic Engineering', *Journal of Biotechnology* 94: 37–63.

Williams, T. C. et al. (2015) 'Quorum-Sensing Linked RNA Interference for Dynamic Metabolic Pathway Control in *Saccharomyces cerevisiae*', *Metabolic Engineering*, 29: 124–34.

——, Nielsen, L. K., and Vickers, C. E. (2013) 'Engineered Quorum Sensing Using Pheromone-Mediated Cell-to-Cell Communication in *Saccharomyces cerevisiae*. *ACS Synthetic Biology*, 2: 136–49.

——, Peng, B., Vickers, C. E., and Nielsen, L. K. (2016) 'The *Saccharomyces cerevisiae* Pheromone-Response Is a Metabolically Active Stationary Phase for Bio-Production', *Metabolic Engineering Communications* 3: 142–52.

——, Pretorius, I. S., and Paulsen, I. T. (2016) 'Synthetic Evolution of Metabolic Productivity Using Biosensors', *Trends Biotechnology*, 34: 371–81.

Wisselink, H. W. et al. (2009) 'Novel Evolutionary Engineering Approach for Accelerated Utilization of Glucose, Xylose, and Arabinose Mixtures by Engineered *Saccharomyces cerevisiae* Strains', *Applied and Environmental Microbiology*, 75: 907–14.

Zelle, R. M. et al. (2008) 'Malic Acid Production by *Saccharomyces cerevisiae*: Engineering of Pyruvate Carboxylation, Oxaloacetate Reduction, and Malate Export', *Applied and Environmental Microbiology*, 74: 2766–77.

Zhang, J. et al. (2016) 'Engineering an NADPH/NADP+ Redox Biosensor in Yeast' *ACS Synthetic Biology* 5: 1546–56.

Zhang, F., and Keasling, J. (2011) 'Biosensors and Their Applications in Microbial Metabolic Engineering', *Trends in Microbiology*, 19: 323–9.

Zhang, J., Jensen, M. K., and Keasling, J. D. (2015) 'Development of Biosensors and Their Application in Metabolic Engineering', *Current Opinion in Chemical Biology*, 28: 1–8.

Zhou, H. et al. (2012) 'Xylose Isomerase Overexpression Along With Engineering of the Pentose Phosphate Pathway and Evolutionary Engineering Enable Rapid Xylose Utilization and Ethanol Production by *Saccharomyces cerevisiae*', *Metabolic Engineering*, 14: 611–22.

Chapter 4 : SCRaMbLE and biosensor-mediated selection for improved PA production in yeast

Chapter 4

SCRaMbLE and biosensor-mediated selection for improved PA production in yeast

Xin Xu [1], Thomas C. Williams [1,2 *], Isak S. Pretorius [1], Ian T. Paulsen [1 *]

[1] Department of Molecular Sciences, Macquarie University, Sydney NSW 2109, Australia

[2] CSIRO Synthetic Biology Future Science Platform, Canberra, ACT 2601, Australia

*Corresponding authors: Ian T. Paulsen and Thomas C. Williams.

E-mail addresses: ian.paulsen@mq.edu.au; tom.williams@mq.edu.au.

Abstract

Through induction of the Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution (SCRaMbLE) system in synthetic yeast genomes, diverse genome rearrangements can be generated and applied for strain improvement in yeast metabolic engineering (1, 2). In this study, SCRaMbLE was performed on haploid and diploid yeast strains containing synthetic chromosomes and a synthetic Wood-Werkman cycle to improve propionic acid (PA) production. The ratiometric biosensor combined with fluorescence activated cell sorting (FACS) were applied for high-throughput screening. After SCRaMbLE of the diploid strain, two strains with 1.7- and 2.4-fold improved titres of PA were identified by biosensor-mediated single cell sorting. However, through whole genome re-sequencing, loop-out of genes in the synthetic Wood-Werkman cycle except *scpC* were found in the parental diploid strain before the induction of SCRaMbLE, and loss of synthetic chromosomes were also identified in the two PA productive isolates. The increased production of PA was attributed to degradation of exogenous amino acids to propionyl-CoA through yeast native amino acid catabolic pathways, and the conversion of propionyl-CoA to propionate by the only remaining Wood-Werkman cycle gene, *scpC*.

Introduction

Saccharomyces cerevisiae is the most intensively studied eukaryotic model microorganism due to the availability of its whole genome sequence since 1996 (3). Considering its fast growth rate, robustness in the fermentation processes, and easy genetic manipulation, yeast has been used as a key cell factory in industrial fermentation (4). Through metabolic engineering, the yield of a desired product from natural hosts can be further improved, and novel genes and pathways can be integrated into a platform cell factory such as yeast for the production of a wide range of resources including fuels, bulk chemicals, fine chemicals, and pharmaceuticals (4, 5). However, the strain engineering processes used to improve product yields are time-consuming, labour-intensive, and limited by available biological knowledge and subsequent

rational design choices. The development of metabolic engineering has been facilitated by the advances in other fields including 'omics' techniques, systems biology and modelling approaches (5).

With the decreasing price in DNA synthesis and improvement of DNA assembly approaches, synthetic biology is emerging as a discipline that will revolutionise the engineering of biological systems (6). Synthetic biology enables the design and construction of genetic 'parts', 'modules', systems, and even whole cells (7, 8). The *Saccharomyces cerevisiae* 2.0 project (Sc2.0) aims to build the world's first synthetic eukaryotic genome. Multiple design changes have been made in the synthetic genome, within which is the whole genome rearrangement system - Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution (SCRaMbLE). In the synthetic genome, synthetic symmetrical loxPsym sites were inserted 3 bp after the stop codon of all non-essential genes and other major landmarks. The recombination events between loxPsym sites could happen in either direction via the activation of a site-specific recombinase, Cre-EBD, which localises to the nucleus in the presence of estradiol (9, 10). Compared with traditional mutagenesis methods, SCRaMbLE generates combinational genome diversifications including gene deletions, duplications, inversions and translocations (11). Diverse phenotypic changes are produced via SCRaMbLE of synthetic yeast chromosomes, which offers a large mutant pool for the selection of industrially relevant phenotypes (2).

Propionic acid (PA) is one of the top building block chemicals used as a food preservative, herbicide (12), plastic (13), pharmaceutical (14), and cosmetics (15). Currently, the refining of PA is still dependent on finite oil reserves. Microbial production of PA from renewable biomass through engineering yeast as a cell factory is a sustainable alternative to replace petrochemical refining (16). *Propionibacterium* species can produce PA as the main fermentation product through the Wood-Werkman cycle (17). However, increasing the yield of PA is difficult to achieve owing to a lack of genetic modification tools (15) and growth inhibition of propionibacteria in fermentation processes (18). Compared with native PA producers, *S. cerevisiae* is a robust host widely applied in industrial fermentation. A synthetic, redesigned

Wood-Werkman cycle has been expressed in yeast (Williams et al., unpublished data) by our group. The synthetic Wood-Werkman cycle contains the following four enzymes, the Vitamin B_{12}-dependent methyl-malonyl-CoA mutase (encoded by *mutA* and *mutB*) from *Saccharopolyspora erythraea*, the methyl malonyl-CoA epimerase (*mce*), the methyl malonyl-CoA carboxytransferase (*mtcA*, *mtcB*, *mtcC*, *mtcD*), and the propionyl-CoA succinyl-CoA transferase (*pst*) from *P. acidipropionici*, which enable the production of propionate and oxaloacetate from pyruvate and succinate, via succinyl-CoA, methyl malonyl-CoA, and propionyl-CoA (Figure 4-1) (Williams et al., unpublished data). The propionyl-CoA succinyl-CoA transferase from *Escherichia coli* (*scpC*) was also co-expressed. In their study, 0.1 mM of PA was produced through fermentation of the resulting yeast strain on 1 % glucose. All the genes in the synthetic pathway were designed with flanking loxPsym sites. The activation of SCRaMbLE system in synthetic yeast with heterologous expression of the synthetic Wood-Werkman cycle could therefore enable the multiple integration of the pathway genes into the synthetic chromosomes with genome rearrangements, potentially generating mutants with improved PA yield. SCRaMbLE can be used to generate an effectively infinite number of chromosomal rearrangements (11), yet only a small fraction of these modifications could be expected to be advantageous for any phenotype of interest. Furthermore, in addition to inducing lethality through essential gene loss, SCRaMbLE can also result in a loss of fitness through deleterous chromsomsal modifications (1, 19). A laboratory selection pressure or high-throughput screening mechanism is therefore required to isolate SCRaMbLEd yeast strains with phenotypes of interest.

In the case of selecting for improved PA production in yeast, synthetic biology tools such as biosensors can be used together with fluorescence activated cell sorting (FACS) to couple metabolite production levels to a fluorescence signal (20). In our lab, a biosensor responsive to PA based on the native yeast transcriptional regulator War1 and *PDR12* promoter has been established (Chapter 3). The biosensor can transduce various concentrations of PA into GFP fluorescence. The dynamic range of GFP output was increased dramatically through the

implementation of a positive feedback loop, and the noise of GFP expression among different cells was normalised by constitutive expression of mCherry (Chapter 3). This biosensor therefore has the potential to be applied in high-throughput screening of PA producers.

In this study, the synthetic Wood-Werkman cycle was SCRaMbLEd in the semi-synthetic yeast containing three synthetic chromosomes-syn III, syn VI and syn IXR (21). Combined with high-throughput FACS and high-performance liquid chromatography (HPLC) measurement of the titre of PA, strains with superior post-SCRaMbLE PA productivity were identified. However, whole genome sequencing showed that genes in the synthetic Wood-Werkman cycle had been looped-out in the PA productive strains, except for *scpC*. Improved PA titres were attributed to the degradation of amino acids to propionyl-CoA through native yeast amino acid catabolic pathways, and the conversion of propionyl-CoA to PA by the only remaining Wood-Werkman cycle gene, *scpC*. Although SCRaMbLE of the synthetic Wood-Werkman cycle hasn't been achieved, a superior strategy for SCRaMbLE and high-throughput screening of PA producers has been developed in our study that can be applied in future studies to identify mechanisms contributing to PA production.

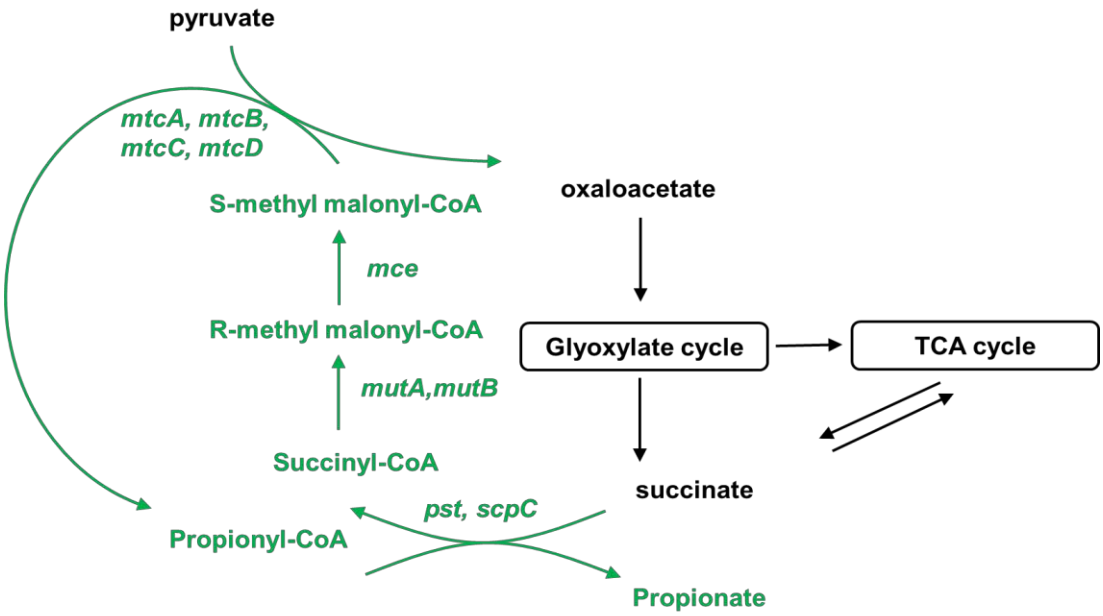


Figure 4-1. The Synthetic Wood-Werkman cycle genes, metabolites, and their interaction with native yeast metabolism. (Adopted from Williams et al., unpublished data). The synthetic Wood-Werkman cycle genes, metabolites, and reactions are depicted in green lines, while native metabolites, and reactions, and its proposed interaction with the Wood-Werkman cycle are shown in black. TCA cycle is short for the tricarboxylic acid cycle.

Methods

Growth media

For non-selective growth, yeast strains were grown in liquid yeast extract (1 % w/v) peptone (2 % w/v) glucose (2 % w/v) (YPD). To maintain auxotrophic markers, yeast strains were grown in synthetic complete (SC) media containing 1x Yeast Nitrogen Base Without Amino Acids mix (Sigma-aldrich Y0626), glucose (1 % w/v), and Synthetic complete amino acid supplement minus Histidine, Leucine, Lysine, Methionine, Uracil (USBiological Life Science, D9543) with the appropriate autotrophic amino acids supplemented at a concentration of 100 mg per litre.

Plasmid Construction

The plasmids used in this study are listed in Table 4-1. The two pathway plasmids (*PA1.b-ScpC-pRS416* and *PA2.b-pRS415*) were previously constructed in our lab. The *pPDR12-GFP-mCherry* cassette was linearised from *pPDR12-GFP-mCherry-pRS415* and cloned into *pSCW11-Cre-EBD-pRS413* using Gibson assembly of PCR amplified vector and insert DNA to create the biosensor plasmid *pGFP-mCherry-Cre-EBD-pRS413*.

Construction of haploid and diploid strains containing the heterologous pathway

The yeast lab strain BY4741 and the semi-synthetic strain yLM896 (21) containing syn III, syn VI and syn IXR were used in this study (Table 4-2). To build the haploid synthetic strain containing the PA pathway, the biosensor plasmid *pGFP-mCherry-Cre-EBD-pRS413* and 2 pathway plasmids (*PA1.b-ScpC-pRS416* and *PA2.b-pRS415*) were co-transformed into yLM896 to generate the SynPA-haploid strain. *pGFP-mCherry-Cre-EBD-pRS413*, *PA1.b-ScpC-pRS416*, and *PA2.b-pRS415* were transformed into +FB. BY4741 strain to generate the +FB. BYPA strain. The Heterozygous SynPA-diploid stain was constructed by mating the semi-synthetic strain yLM896 with +FB. BYPA.

All yeast transformations were conducted with the LiAc/SS carrier DNA/PEG method (22) and the transformants were selected for growth on SC agar plates with appropriate amino acids supplemented.

Table 4-1. Plasmids used in this study.

Name	Details	Origin
pPDR12-GFP-mCherry-pRS415	*pPDR12-yEGFP-ADH1t-pPDA1-mCherry-CYC1t*-pRS415, *URA3* marker	Williams, Xu et al., 2017
pSCW11-Cre-EBD-pRS413	*pSCW11-Cre-EBD-tCYC1*-pRS413, *HIS3* marker	Gottschling's lab
pGFP-mCherry-Cre-EBD-pRS413	*pSCW11-CRE-EBD-pPDR12-yEGFP-pPDA1-mCherry-CYC1t*-pRS413, *HIS3* marker	This study
PA1.b-ScpC-pRS416	*pTDH3-mtcA-ADH1t-pTEF1-mtcB-CYC1t-pTPI1-mtcC-STE2t-pTEF2-mtcD-MFA1t-pPGK1-mce-PHO5t-pTEF1-ScpC-TSynth25*-pRS416, *URA3* marker	This lab
PA2.b-pRS415	*pTDH3-mutA-STE2t-pTEF2-mutB-ADH1t-pTEF1-pst-PHO5t*-pRS415, *LEU2* marker	This lab

Table 4-2. Strains used in this study.

Name	Genotype	Note	Origin
BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Haploid laboratory strain, mating type 'a'	Euroscarf
yLM896	*MATα leu2Δ0 lys2Δ0 MET15 his3Δ1 ura3Δ0 synIII SYN.SUP61::ho::ura3 synVI SYN-WT.PRE4 IXL-synIXR*	Semi-synthetic haploid strain containing syn III, syn VI and syn IXR, mating type 'α'	Mitchell et al., 2017
+FB. BY4741	BY4741, *pWAR1::pPDR12*	Positive feedback strain (chromosomal replacement of the *WAR1* promoter with the *PDR12* promoter)	Williams, Xu et al., 2017
+FB. BYPA	+FB. BY4741, *pGFP-mCherry-Cre-EBD-pRS413, PA1.b-ScpC-pRS416* and *PA2.b-pRS415*	BY4741 containing positive-feedback loop, the biosensor and two PA pathway plasmids	This study
SynPA-haploid	yLM896 strain containing *GFP-mCherry-Cre-EBD-pRS413, PA1.b-ScpC-pRS416* and *PA2.b-pRS415*	Semi-synthetic haploid strain containing the biosensor and two PA pathway plasmids	This study
SynPA-diploid	Diploid of yLM896 and +FB. BYPA	Semi-synthetic diploid strain containing the positive-feedback loop, the biosensor and two PA pathway plasmids	This study

'SCRaMbLE' of haploid and diploid strain and the selection of the potential PA productive strain

To determine the suitable time length of SCRaMbLE, single colonies of SynPA-haploid strain were grown overnight at 30 °C in 5 mL of SC medium with methionine and then re-inoculated at a starting OD_{600} of 0.2 in 10 mL of medium to allow further growth for 3 hours. The 2nd

round of preculture was re-grown in fresh medium at an initial OD_{600} of 0.2, and SCRaMbLE was induced by the addition of 1 μM β-estradiol (Sigma, E2257) in triplicate cultures. The same treatment without the addition of β-estradiol was used as a non-SCRaMbLE control. After SCRaMbLE for 0 h, 2 h, 4 h, 6 h, 10 h, and 24 h, aliquots were taken, washed twice with ddH_2O, serially diluted and finally plated on YPD agar. The plates were incubated for 2 days before the colony numbers were counted.

After the determination of time length for SCRaMbLE, SynPA-haploid and SynPA-diploid strains were grown and SCRaMbLEd for 4 h using the method described above, except that the SynPA-diploid was grown in SC medium without any auxotrophic amino acids supplemented. After 4 h, the suspensions were harvested, washed twice with ddH_2O, and re-inoculated into fresh SC media with 0.1 g/L vitamin B12 for an overnight growth before fluorescence measurement.

Fluorescence measurement

A BD Influx flow cytometer was used for the determination of fluorescence signal and cell sorting. The GFP and mCherry fluorescence was measured simultaneously using a 200mW 488-nm laser with a 530/40nm emission filter and a 692/40-emission filter, respectively. BD FACS™ Sortware was used for controlling the cytometer and analysing data. The PMT voltages were set as follows: forward-angle light scatter (FSC), 13.36; side-angle light scatter (SSC), 17.83; the 530/40 emission filter (530/40), 38.1; the 692/40-emission filter (692/40), 75.0; all in log mode. For cell sorting, the piezo amplitude and the drop frequency was set at 5.46 KHz and 58.58 KHz; the stream focus was set at 16.80%; the maximum drop charge was set at 88 volts; the stream deflection was set at -99 % Max; the sort mode was 1.0 drop enrich.

Total events in the single cell region were recorded. The graph presented the GFP fluorescence in y-axes and the mCherry fluorescence in x-axes. The inducible expression of GFP can be normalised by the constitutive expression of mCherry within a cell. The GFP/mCherry ratio is positively correlated with the concentration of PA (23). Non-SCRaMbLEd cells were measured

as control. The red gate P1 in each experiment was drawn according to the top 10% fluorescent population of the non-SCRaMbLEd sample. Based on the position of the top 10% of the control population, the percentages of cells from the SCRaMbLEd population in the P1 gate were recorded. Mean GFP/mCherry ratios and the percentage of the whole population in the red gate are two parameters for evaluating the expression of normalised GFP at the population level, which in theory positively correlates to the titre of PA. After SCRaMbLE of the SynPA-haploid and SynPA-diploid strains, 50000 cells were recorded, as were the GFP/mCherry ratio and the percentages of top fluorescence populations.

Screening of 'SCRaMbLE' library for the PA productive strain

For further screening of SCRaMbLEd PA producer strains, three sorting methods were used. Firstly, the SynPA-diploid strain was subjected to iterative SCRaMbLE, sorting and outgrowth. After each round of SCRaMbLE, the cells were harvested, washed and regrown overnight, and the top 10 % fluorescent cells of the population were sorted. The green gate was drawn according to the top 10% fluorescent population of the non-SCRaMbLEd sample in the first round of determination. In each round, 50000 cells were recorded for measuring the mean GFP/mCherry ratio and the percentage of high fluorescence population (percentage of cells in the green gate), and 10000 cells were sorted for each replicate. After the sorted cells were regrown to an OD_{600} value of 4.0-6.0, the culture was reinoculated using the procedure described above before another round of SCRaMbLE. In total, four rounds of SCRaMbLE-sorting-outgrowth were conducted, and the cell suspension after the fourth sorting and outgrowth was stored for further analysis. In a second approach, the SynPA-diploid was subjected to continuous rounds of cell sorting with only one round of SCRaMbLE. After SCRaMbLE, the cells were harvested, washed and regrown overnight. 20000 cells from the overnight culture were measured and recorded by FACS, and the P2 gate was made to comprise the top 1 % of fluorescent cells of the population. Cells in the P2 gate were sorted in the continuous rounds of sorting. 50000 cells were sorted for each replicate in the 1st round of sorting. The cells obtained in the 1st round were directly subjected to the 2nd round of sorting

without outgrowth. 5000 cells were sorted and then used directly as input for a 3rd round. Finally, 500 cells were sorted and regrown in fresh SC medium until an OD_{600} of 4.0-6.0 was reached. The re-grown culture after 3 rounds of sorting was collected for the following fermentation. Thirdly, single cells were sorted after SCRaMbLE of the SynPA-diploid strain. The SCRaMbLEd cells were harvested, washed and regrown overnight and the top 0.1-1 % fluorescent cells of the population were sorted directly on to SC agar with vitamin B12. The plates were incubated for 2 days. The colonies were checked again under the blue light (Safe Imager™ 2.0, Invitrogen) and 110 high fluorescent colonies were selected for fermentation. In all the three assays, a SynPA-diploid strain that was subjected to the same process without the addition of β-estradiol was used as non-SCRaMbLE control.

Fermentation and HPLC determination of PA concentration

To measure the titre of PA, the pre-growth and the fermentation was performed in SC medium with 0.1 g/L Vitamin B_{12}. Populations and single colonies isolated using FACS were inoculated into 5 ml of the SC medium for the first round of overnight pre-growth. The overnight culture was then regrown in 5 ml of fresh medium until exponential phase for the second round of pre-culture. The cell suspension was at last inoculated into a 250 mL baffled flask containing 10 mL of medium at an initial OD of 0.1. Each preculture and fermentation was performed in triplicate for each strain, except for the initial screening of the single colonies, which were fermented using one replicate. All the fermentations were performed for 48 h. After the final OD value of each culture was recorded, the cells were harvested, and the supernatants were collected by filtering through a 0.2 μm Filter (Sartorius, Minisart Syringe Filter).

1 mL of fermentation supernatant for each sample, or sodium propionate (Sigma) standards (0.25 mM, 0.5 mM, 1.25 mM, 2.5 mM and 5 mM) were loaded into 2 mL Verex Vials (Phenomenex, AR0-3910-13). The samples and the standards were measured by HPLC using an Agilent Technologies 1260 Infinity II HPLC system with Agilent 1260 RID detector. The method of HPLC analysis has been modified from Dietmair *et al* (24). Analytes were eluted

through a guard column (SecurityGuard™ cartridges, Carbo-H, AJ0-4490) and a Phenomenex Rezex™ RHM-Monosacchride H^+ (8%) HPLC column with 4 mM H_2SO_4 at 40°C with a 0.4 ml min^{-1} flow rate for 35 min. Chromatograms were analysed by OpenLAB software.

Whole genome resequencing and data analysis

The PA productive strain SCR5, SCR22 and the parental strain SynPA-diploid were whole genome re-sequenced. After overnight growth in SC medium, the genomic DNA of the three strains was extracted with Yeast DNA Extraction Kit (Thermo Fisher scientific, Cat No. 78870). Whole genome resequencing and library preparation was performed by Macrogen Inc. using a True-Seq Nano kit with 470 bp inserts and paired-end Illumina HiSeq 2500 sequencing. The three samples all gave a quality score of Q30 above 90% and above 16 million reads and were therefore not trimmed. The reads were analysed using Geneious Pro 9.1.3 (25). Paired end reads were mapped to the S288C reference genome, synIII, synVI and synIXR chromosomes, and the sequence of 3 plasmids (*PA1.b-ScpC-pRS416, PA2.b-pRS415* and *GFP-mCherry-Cre-EBD-pRS413*). The genomes of SCR5 and SCR22 were analysed by mapping of the reads on the reference genome for detection of deletions or rearrangements, while gene duplication or amplification events were detected by the relative levels of sequence coverage.

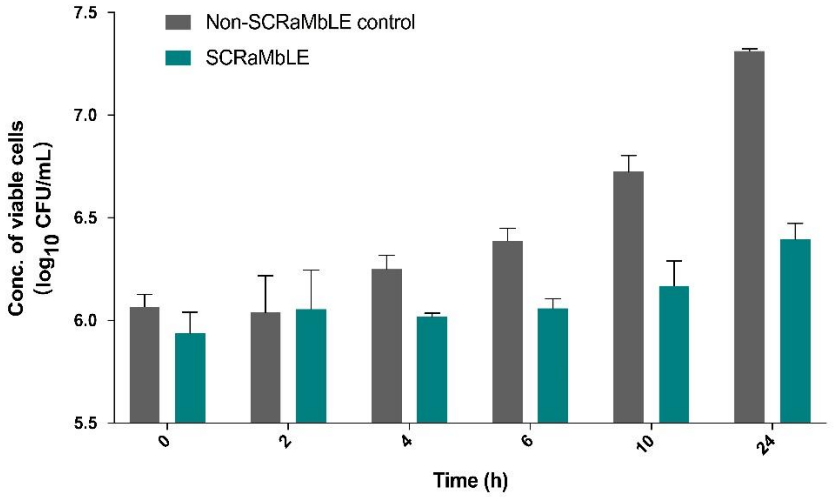
Results

The effect of SCRaMbLE on cell viability

Lengthy induction of SCRaMbLE can result in severe fitness defects and a significant decrease in the number of viable cells. Optimally, a SCRaMbLEd population will undergo viability loss as an indicator of successful SCRaMbLE, but not to the extent that too few viable cells remain in the population. To determine a desirable induction time for SCRaMbLE, the SynPA-haploid strain was treated with and without β-estradiol to activate the Cre-recombinase enzyme that targets loxPsym sites to induce SCRaMbLE rearrangements. Samples were taken at intervals within 24 h, and the number of viable cells was estimated using agar plate colony counts. There

was no obvious difference of the cell concentrations between SCRaMbLEd treatment and non-SCRaMbLEd treatment in the initial 2 h (Figure 4-2A). After SCRaMbLE for 4 h, the concentration of the non-SCRaMbLEd cells increased to 6.25 log_{10}CFU/mL, while the concentration of SCRaMbLEd cells was only 6.02 log_{10}CFU/mL (Figure 4-2A). 58.60% of cells were viable compared with the non-SCRaMbLEd control (Figure 4-2B). As the SCRaMbLE continued, the percentage of viable cells dropped. After SCRaMbLE for 24 hours, the concentration of the SCRaMbLEd cells was 6.40 log_{10}CFU/mL (Figure 4-2A), which was only 12.14% compared with the non-SCRaMbLEd control (Figure 4-2B). The decreased viable cell number likely resulted from gene deletions, which can be regarded as a sign of SCRaMbLE happening. Four hours of induction was used for SCRaMbLE in the following assays since it is sufficient for the induction of genome rearrangements, without excessive loss of population viability.

A



B

Time (h)	2	4	6	10	24
Percentage of viability (%)*	103.76	58.60	46.95	27.82	12.14

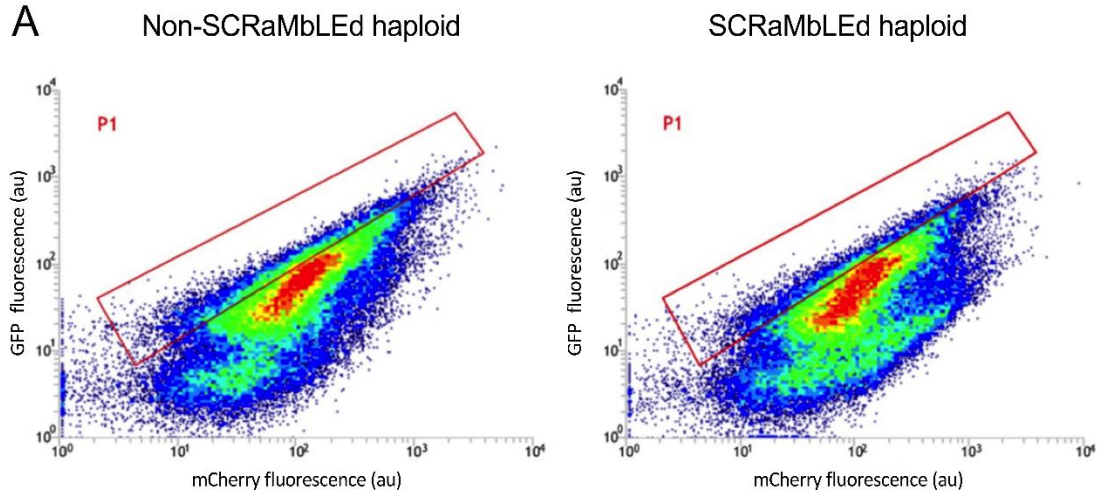
*Percentage of Viability (%) = Concentration of SCRaMbLEd cells/ Concentration of non-SCRaMbLEd cells.

Figure 4-2. Cell viability after SCRaMbLE. The SynPA-haploid was SCRaMbLEd over 24 hours. Aliquots were taken after 0 h, 2 h, 4 h, 6 h, 10 h and 24 h, washed twice, serial diluted, and finally plated on YPD agar. After 2 days of incubation, the colony numbers were counted, and the concentrations of viable cells were recorded (A). Percentages of viability after SCRaMbLE were calculated (B).

Fluorescence determination of SCRaMbLEd haploid and diploid semi-synthetic yeast strain

The fluorescence signal was measured after SCRaMbLE using flow cytometry. The PA-induced GFP fluorescence (y-axes) was normalised by the constitutive mCherry fluorescence (x-axes) to control for intrinsic noise. 50000 non-SCRaMbLEd and SCRaMbLEd SynPA-haploid cells were recorded, and the gate of the top fluorescence population was positioned according to the top 10.59% of the non-SCRaMbLEd population (Figure 4-3A). After SCRaMbLE of SynPA-haploid, the mean GFP/mCherry ratios and the percentages of the high GFP/mCherry population both decreased compared with the non-SCRaMbLEd control (Figure 4-3B). The decrease suggested the concentration of PA produced by the population was not improved after SCRaMbLE of SynPA-haploid. Thus, it was not used for any further PA producer screening.

A



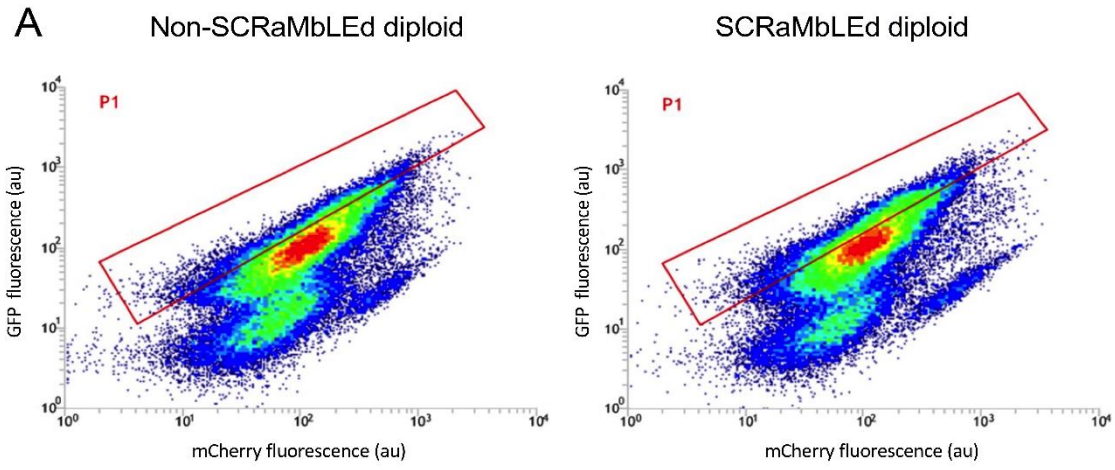
B

	Mean GFP/mCherry	Percentage of top fluorescence population (%)
Non-SCRaMbLEd haploid	0.47±0.00	10.59±0.06
SCRaMbLEd haploid	0.33±0.01	6.15±0.52

Figure 4-3. Fluorescence determination after SCRaMbLE of SynPA-haploid. Fluorescence distribution plot of non-SCRaMbLEd and SCRaMbLEd haploid strains (A). The table shows the mean GFP/mCherry ratios and the percentages of top fluorescence population (B). The red gate P1 was drawn according to the position of the top 10.59% fluorescent cells of the non-SCRaMbLEd population.

The deletion of genes affecting growth in SCRaMbLEd haploid might result in impaired fitness, leading to a decreased concentration of PA being produced. To increase the strain fitness after SCRaMbLE, the SynPA-diploid strain was generated by crossing the semi-synthetic strain yLM896 with +FB BYPA. By doing so, a copy of the wild type genome will be maintained so that essential genes cannot be lost from the genome, but synthetic chromosomes can still undergo SCRaMbLE. 50000 non-SCRaMbLEd and SCRaMbLEd SynPA-diploid cells were recorded, and the gate of the top fluorescence population was positioned to include the top 10.22% of the non-SCRaMbLEd population (Figure 4-4A). The percentage of events in this gate reached 23.26% in the SCRaMbLEd population, more than doubling the proportion of the non-SCRaMbLEd condition (Figure 4-4B), suggesting the percentage of PA-productive cells were increased. SCRaMbLEd SynPA-diploid cells also gave a higher mean GFP/mCherry ratio compared with the non-SCRaMbLEd control (Figure 4-4B), which indicated the SCRaMbLEd SynPA-diploid produced increased concentrations of PA on average. Thus, SCRaMbLE of SynPA-diploid was applied to screen for SCRaMbLEd PA producer cells.

A



B

	Mean GFP/mCherry	Percentage of top fluorescence population (%)
Non-SCRaMbLEd diploid	0.50±0.05	10.22±0.09
SCRaMbLEd diploid	0.65±0.02	23.26±2.01

Figure 4-4. Fluorescence determination after SCRaMbLE of SynPA-diploid. Fluorescence density plot of non-SCRaMbLEd and SCRaMbLEd diploid (A). The table shows the mean GFP/mCherry ratio and the percentage of top fluorescence population (B). The red gate P1 was drawn according to the position of the top 10.22% fluorescent cells of the Non-SCRaMbLEd population.

Screening of PA productive SCRaMbLEd strains

Although the average fluorescence was increased after SCRaMbLE of the SynPA-diploid, the SCRaMbLEd cells have diverse phenotypes with significantly improved PA producing cells likely present at a very low percentage within the population. To optimise the procedure for the screening of PA producers, three different approaches were attempted (Figure 4-5). Firstly, iterative SCRaMbLE-sorting-outgrowth (Figure 4-5A) was performed to further increase PA titre by accumulating the beneficial mutations after iterative rounds of SCRaMbLE; secondly, continuous sorting after one round of SCRaMbLE (Figure 4-5B) was set up to enrich the cells that produced high titres of PA stably; lastly, single cell sorting after one round of SCRaMbLE (Figure 4-5C) was conducted to sort out single cells with the highest PA production.

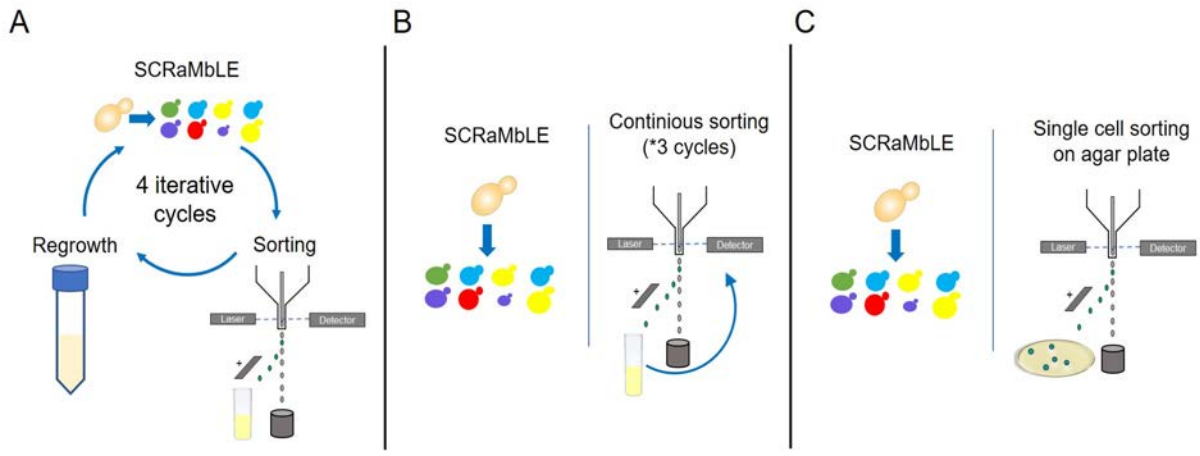


Figure 4-5. Three sorting procedures for screening of PA productive SCRaMbLEd strains: iterative SCRaMbLE-sorting-outgrowth (A); continuous sorting after one round of SCRaMbLE (B); single cell sorting after one round of SCRaMbLE (C).

Iterative SCRaMbLE, sorting, and outgrowth

Firstly, four rounds of SCRaMbLE-sorting-outgrowth were conducted. The percentage of the high fluorescent cells (cells in the green gate) increased from 26.07% in the first round to 38.29% in the final measurement (Figure 4-6A). The mean GFP/mCherry ratio increased after the 2nd round, dropped slightly in the 3rd round, and went up again in the last round (Figure 4-6B). The mean GFP/mCherry value of the non-SCRaMbLEd cells also increased in a similar trend but the value was lower than the SCRaMbLEd treatment (Figure 4-6B). The concentration of PA

in the fermentation supernatant was determined by high-performance liquid chromatography (HPLC). However, the PA concentrations of SCRaMbLEd samples were not increased compared with the non-SCRaMbLEd control through the iterative SCRaMbLE-sorting-outgrowth process (Figure 4-6C). Iterative SCRaMbLE-sorting-outgrowth failed to sort out the PA productive cells. This might result from PA producers being outgrown by the fast-growing cells during regrowth, or mutation of the biosensor encoding genes.

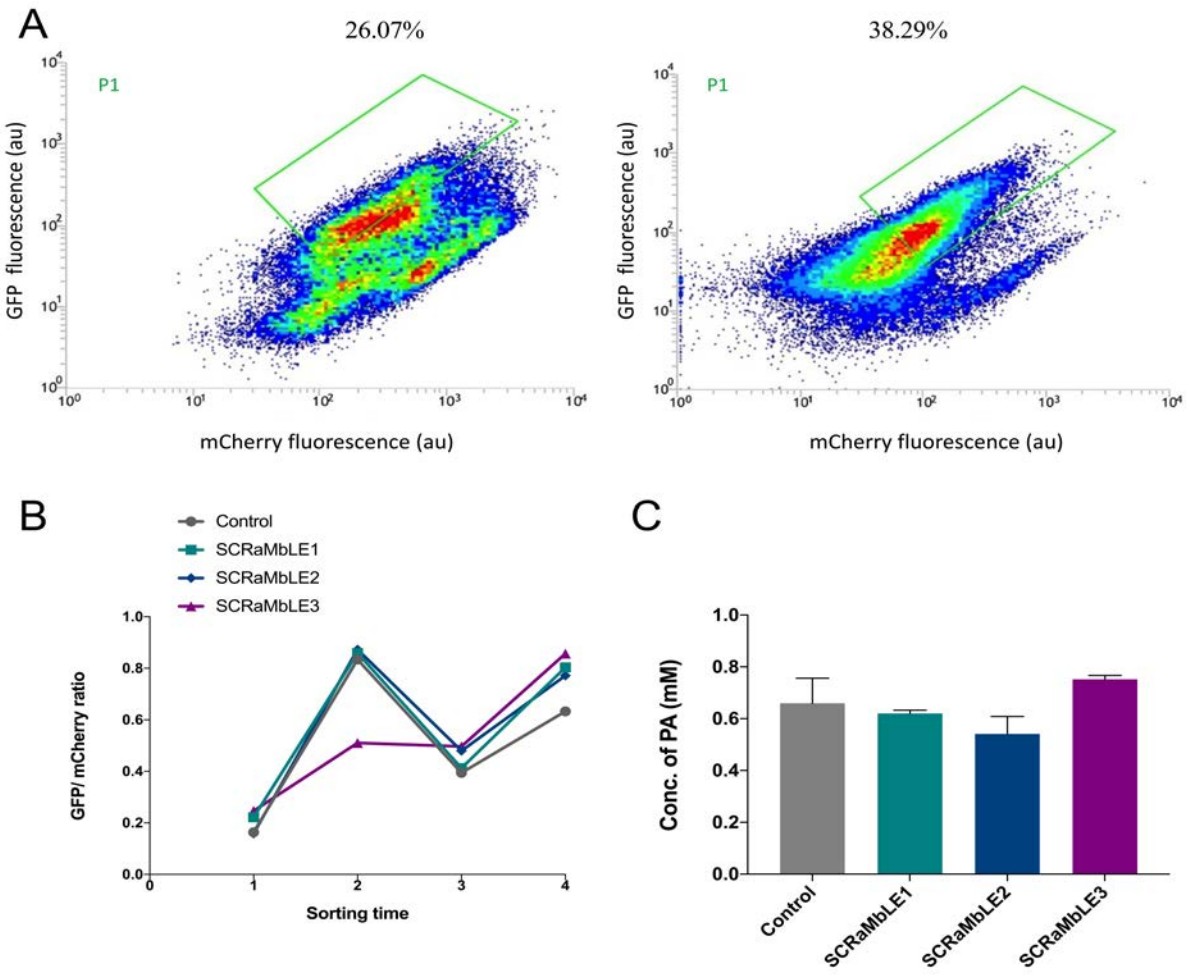


Figure 4-6. Iterative SCRaMbLE, sorting, and outgrowth. Fluorescence distribution plots of the 1st (left) and 4th (right) rounds of sorting from one SCRaMbLEd treatment. The corresponding percentages of the green gate were marked on top. The position of the green gate was drawn to comprise the top 10% fluorescent cells of the non-SCRaMbLEd control (A). The mean GFP/mCherry ratios of non-SCRaMbLEd control and the continuous SCRaMbLEd population throughout the process were determined. The value represents the mean of 50000 cells (B). The concentrations of PA in fermentation supernatants were determined using HPLC (C). Bars and error bars represent the mean and standard deviation (SD) of triplicate cultures.

Continuous sorting after one round of SCRaMbLE

Secondly, three rounds of continuous cell sorting were conducted after one round of SCRaMbLE. The percentage of cells in the high GFP/mCherry population was greatly enriched via the continuous sorting process (Figure 4-7A). However, the concentration of PA being produced was not increased either, according to the HPLC result (Figure 4-7B). Continuous sorting after one round of SCRaMbLE was not an applicable method for the screening of PA producers, which was possibly due to the influence of the low fluorescence cells. After the 3rd round of sorting, there were still around 40% of cells that were not in the high fluorescence gate. The average PA titre from the fermentation of the sorted population is likely significantly lowered by the percentage of low fluorescence cells.

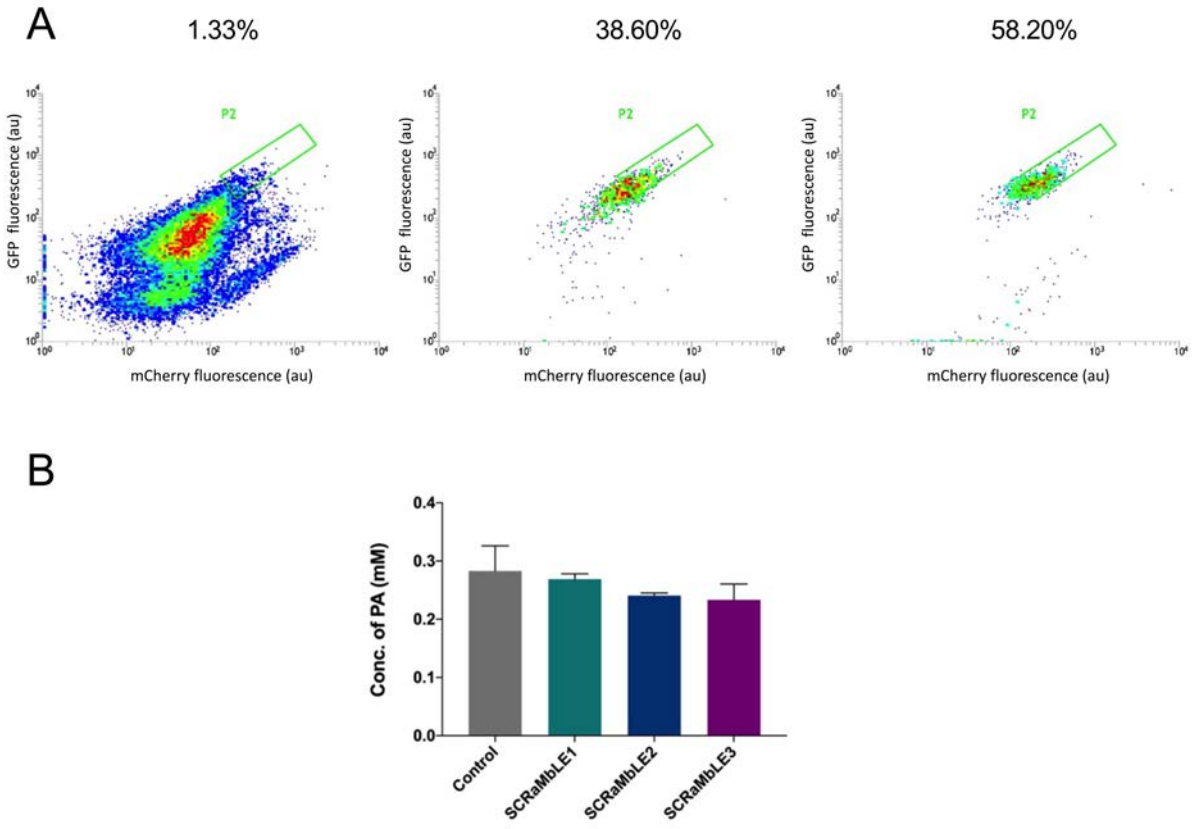


Figure 4-7. Continuous FACS sorting with one round of SCRaMbLE. Fluorescence distribution plot of one SCRaMbLEd treatment in the 1st round of sorting (the left plot), the 2nd round of sorting (the middle plot) and the last round of sorting (the right plot). The corresponding percentages of P2 gate were marked on top. The position of P2 gate was circled according to the top 1% fluorescent cells of the SCRaMbLEd population (A). The concentrations of PA in fermentation supernatants of SCRaMbLEd and control populations were determined using HPLC (B). Bars and error bars represent the mean and standard deviation (SD) of triplicate cultures.

Single cell sorting after one round of SCRaMbLE

Lastly, single cells from the top 0.1-1 % high fluorescent population were sorted directly on the SC agar with Vitamin B_{12} after SCRaMbLE. After incubation, the colonies showing high GFP fluorescence were further selected under blue light. In the initial screening of 110 strains, fermentation of SCRaMbLEd strains, such as SCR4, SCR5, SCR22, SCR88, SCR103, showed improved PA titre relative to the parental strain (Figure 4-8A). Thus, single cell sorting after one round of SCRaMbLE is a suitable method for the screening of PA productive cells. SCR5 and SCR22, were the most productive isolates from the 110 strains that were analysed in the initial screening, and were therefore used for further analysis. From triplicate fermentations, SCR5 and SCR22 produced an average of 0.94 mM and 1.32 mM of PA, representing a 1.7- and 2.4- fold increase relative to the parental strain (0.56 mM), respectively (Figure 4-8B).

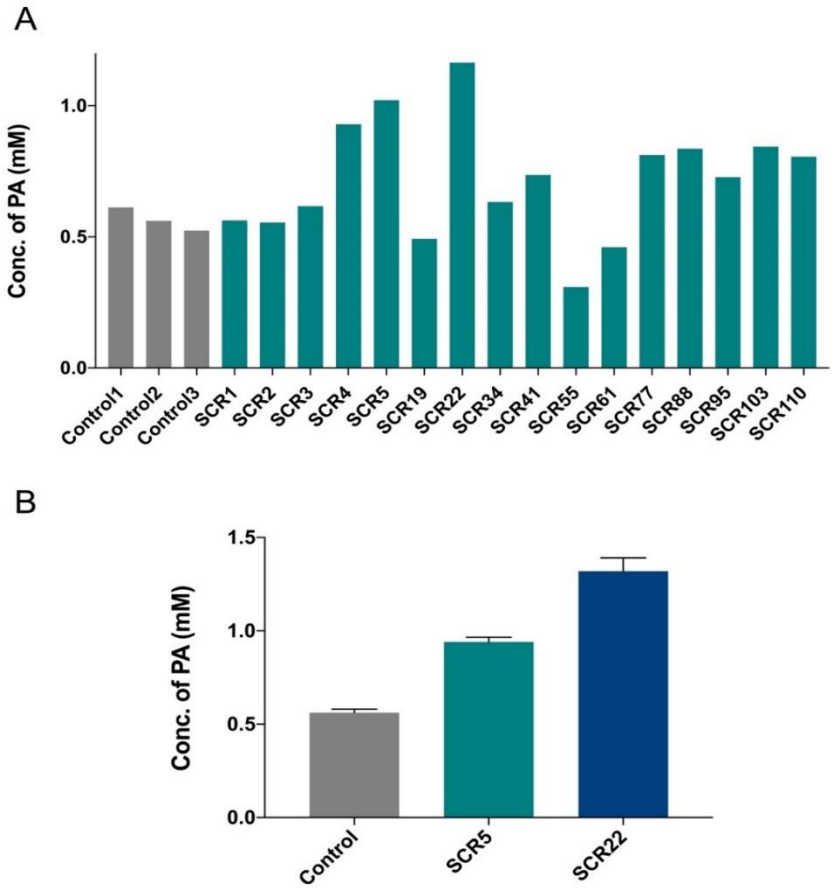


Figure 4-8. HPLC measurement of PA concentrations from fermentation of single cell sorted strains. Representative results of the initial screening of 110 colonies (A). Improved PA titres of SCRaMbLEd strains SCR5 and SCR22 determined by HPLC. Bars and error bars represent the mean and standard deviation (SD) of triplicate cultures (B).

Whole genome sequencing of the SCRaMbLEd strains

To characterise any genome rearrangements in the SCRaMbLEd strains, PA productive strains (SCR5 and SCR22) and the parental strain SynPA-diploid were sequenced and mapped to the S288C reference genome, synIII, synVI and synIXR chromosomes and the sequence of 3 plasmids (*PA1.b-ScpC-pRS416*, *PA2.b-pRS415* and *GFP-mCherry-Cre-EBD-pRS413*). We expected that an optimised ratio of the PA pathway genes and gene rearrangement in the three synthetic chromosomes would be detected through whole genome sequencing. Surprisingly, pathway gene deletions were observed in the parental strain SynPA-diploid as well as in SCR5 and SCR22. Except for *scpC*, no reads mapped to the PA pathway genes on the *PA1.b-ScpC-pRS416* plasmid (Figure 4-9A). Three pathway genes (*mutA*, *mutB* and *pst*) on *PA2.b-pRS415* were all lost in three strains (Figure 4-9B). The *GFP-mCherry-Cre-EBD* cassette remained on the biosensor plasmid (Figure 4-9C). SynIII, synVI and synIXR chromosomes were present in the starting SynPA-diploid strain. However, the synthetic chromosomes were lost in the SCR5 and SCR22 strains (Figure 4-9D). To conclude, the starting strain SynPA-diploid contained the three synthetic chromosomes but lost the PA pathway genes except *scpC* before the induction of SCRaMbLE. Neither the PA pathway genes nor the synthetic chromosomes synIII, synVI and synIXR were detected in SCR5 and SCR22.

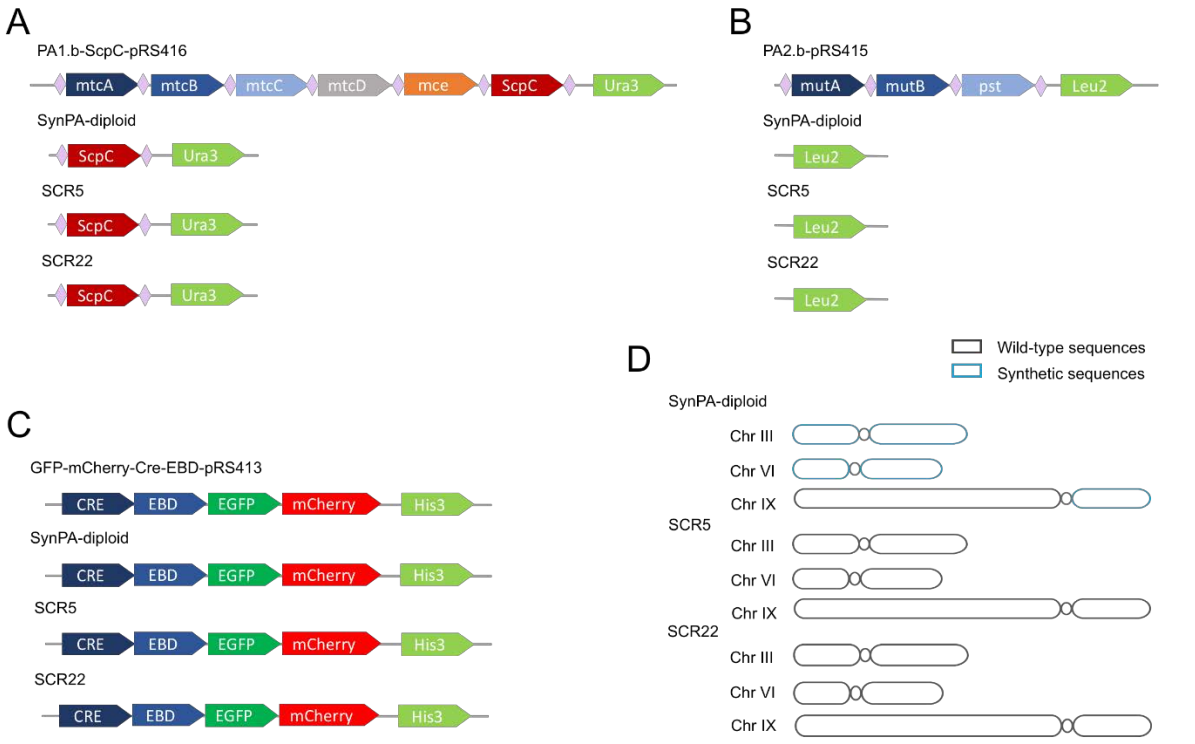


Figure 4-9. Whole-genome re-sequencing analysis of the SynPA-diploid and the PA productive strains SCR5 and SCR22. The figures represent the genotype on *PA1.b-ScpC-pRS416* (A), *PA2.b-pRS415* (B), *GFP-mCherry-Cre-EBD-pRS413* (C) and the mapping to synIII, synVI and synIXR (D).

Discussion

In this study, the potential of SCRaMbLE as a powerful tool to improve the product titre was explored by SCRaMbLEing strains carrying a novel heterologous pathway. The synthetic Wood-Werkman cycle is a complex pathway containing nine genes encoding four enzymes, which interacts with the glyoxylate cycle and the tricarboxylic acid (TCA) cycle via oxaloacetate and succinate (Figure 4-1). With the aim of optimising the expression level of each gene in the pathway and identifying any genomic modifications that could improve the titre of PA, SCRaMbLE was first performed on SynPA-haploid. The viability ratio of SynPA-haploid dropped to 58.6 % after SCRaMbLE for 4 hours (Figure 4-2). This was possibly due to the deletion of essential genes after the induction of SCRaMbLE (19), and is a sign of genome rearrangement events. Thus, four hours of induction was used for the following SCRaMbLE assays as a timepoint for triggering the genome rearrangements without too great a loss of cell viability. However, the average PA biosensor output, and the percentage of highly fluorescent

cells both decreased after SCRaMbLE of the SynPA-haploid (Figure 4-3). This decrease suggested that the concentration of PA being produced was not improved in the 50,000 cells analysed using flow-cytometry. This was likely due to the deletion of genes in SCRaMbLEd haploid led to the decreased fitness. The SynPA-diploid was then generated by crossing with the wild-type chromosomes to maintain at least one copy of essential genes, and increase the diversity of the viable cells after SCRaMbLE. Based on the fluorescent signal of the PA biosensor, the concentration of PA produced from the population was possibly increased through SCRaMbLE of the SynPA-diploid (Figure 4-4). A previous study also showed that SCRaMbLEing of haploid yeast results in a serious loss of viability (26). Jia et al. also reported that more genome rearrangements could be seen after SCRaMbLE of diploid yeast, including the deletions and duplications of both non-essential genes and the essential genes on the synthetic chromosomes (1). In their study, SCRaMbLEing of haploid yeast carrying the carotenoid pathway increased the titre 1.53- to 1.74-fold compared with the parental haploid strain, while SCRaMbLEing the diploid strain led to a superior phenotype, which increased the production 6.29- to 7.81-fold compared with the non-SCRaMbLEd diploid strain (1).

To isolate PA-producer strains after SCRaMbLE of the SynPA-diploid, three sorting methods were conducted (Figure 4-5). Although the mean GFP/mCherry values of SCRaMbLEd and non-SCRaMbLEd cells both increased with iterative SCRaMbLE, sorting, and outgrowth (Figure 4-6B), the titre of PA being produced was not improved (Figure 4-6C). This might be explained by the evolution of ‘cheaters’ (20) due to continuous rounds of transfer and outgrowth. Basically, the cheaters are the cells that have accumulated mutations in biosensor genes to produce a higher fluorescent signal even if the titre of PA has not increased. To avoid the evolution of cheaters, continuous sorting after one round of SCRaMbLE and regrowth was conducted. Although the percentage of cells in the highly fluorescent population was enriched, around half of the population remained outside of the high fluorescence gate after three rounds of sorting (Figure 4-7A). The PA titre did not improve either (Figure 4-7B), which could be possibly affected by the low fluorescent cells within the population. The fermentation of

populations after FACS could be affected by the existence of low PA producing cells and 'cheaters'. Finally, single cells from the top 0.1-1 % high fluorescent gate were sorted directly onto agar plates and strains with improved PA production were identified (Figure 4-8). The organic acid biosensor combined with FACS was supposed to establish a high-throughput sorting process. However, in our study, the sorting was likely influenced by a large percentage of cheater cells. In addition, the concentration of PA produced initially in the haploid strain was only 0.1 mM, which is around the lower limit of detection for the PA biosensor (23). If the titre has not been dramatically increased, the PA sensor might not be sensitive enough to detect the improved producer cells. The strains with significantly improved PA titre should be in an extremely low ratio within the population. The sorting of the top fluorescent population could be a mixture of a small number of real PA producers and a large amount of cheater cells and cells with no increased PA titre. Instead of sorting out population, the single cell sorting method is a more reliable procedure to obtain PA productive strains.

Whole genome resequencing was performed on the parental strain and two PA-productive strains after SCRaMbLE. Instead of the identification of the SCRaMbLEd PA pathway genes, surprisingly, all PA pathway genes except *scpC* had been looped-out before the induction of SCRaMbLE (Figure 4-9A & B). All of the genes in the PA pathway were designed to be flanked by loxPsym sites, and the synthetic genome also contains many loxPsym sites. The pathway genes might be looped out due to the homologous recombination system in yeast (27). In addition, pSCW11-Cre-EBD was used to control the process of SCRaMbLE in our study. The pSCW11 is a daughter-cell-specific promoter, which in theory only produces a pulse of recombinase expression once in each cell's lifetime (9). The process of SCRaMbLE is supposed to be tightly controlled by the addition of β-estradiol. However, growth defects of the synthetic yeast strains containing pSCW11-Cre-EBD were observed without the addition of β-estradiol suggesting the leaky expression of the system (11, 19, 28). This leaky expression problem might explain the deletion of the pathway genes via a loop-out process. In other studies, loss of the pSCW11-Cre-EBD was induced to maintain genome stability after SCRaMbLE (2). A

galactose-driven Cre-EBD had been used for the precise control of iterative cycles of SCRaMbLE (1). In future work, a more stable control system for SCRaMbLE should be used to replace the pSCW11-Cre-EBD.

In addition to the deletion of the PA pathway genes, the absence of the three synthetic chromosomes was another unexpected result (Figure 4-9D). Chromosome loss is a major class of genetic alterations that results in spontaneous loss of heterozygosity in diploid *S. cerevisiae* cells (29). One possible explanation could be that during iterative rounds of growth and replication, spontaneous, or possibly selective processes resulted in loss of the synthetic chromosomes. After SCRaMbLE, the deletions of the large fragments on the synthetic chromosomes might also accelerate loss of the synthetic chromosomes. This result suggests that although SCRaMbLE of heterozygous diploids has the advantage of maintaining cell viability and generating more diverse phenotypes, it might result in complete synthetic chromosome loss.

In spite the fact that the Wood-Werkman cycle is superior to other pathways in terms of the maximum theoretical PA and energy yields (Gonzalez-Garcia et al., 2017), the pathway could still result in a significant metabolic burden for yeast from the heterologous over-expression of the nine PA pathway genes. This is consistent with the fact that slow growth during culture of the SynPA-haploid and SynPA-diploid on synthetic minimal medium was observed. To increase strain fitness and growth, a complete amino acid mixture was supplemented. Unfortunately, propionate production has been observed in some microorganisms through fermentation with complex media derived from amino acid catabolic pathways, specifically via the degradation of valine, threonine, isoleucine and methionine (15). Given this, the supplementation of amino acids in our study offered a rich source of amino acids, which could be catalysed into propionyl-CoA by the native yeast amino acid catabolic pathways, and subsequently converted into propionate with the expression of *scpC* from the synthetic Wood-Werkman cycle. This probably led to the observed PA production in the absence of Wood-Werkman cycle genes other than *scpC*. In *E. coli*, the threonine catabolic pathway has been

explored as a source of propionyl-CoA for the production of 1-propanol (15), 3-hydroxyvalerate (30) and erythromycin (31). However, PA production through amino acid catabolic pathways is not a promising route due to the high cost of supplementing amino acids in industrial fermentations. The complete amino acid powder should therefore not be used in future studies to prevent from the undesired PA production. Alternative metabolic pathways, for example, 1, 2-propanediol associated pathways (32, 33) and the 3-hydroxypropanoate (3HP) cycle (34, 35), could also be expressed and SCRaMbLEd in semi-synthetic yeast for PA production.

Conclusion

SCRaMbLE was applied to semi-synthetic haploid and diploid yeast strains with a heterologous pathway for PA production and a biosensor for PA so that FACS could be used to isolate SCRaMbLEd yeast strains with improved PA production. Compared to SCRaMbLE-ing haploids, SCRaMbLE of diploids is a more effective strategy for maintaining viability and generating genetic diversity, whereas it brings the risk of completely losing synthetic chromosomes. Single cell sorting after one round of SCRaMbLE was a superior strategy for sorting PA producers, while iterative sorting methods gave only highly fluorescent cheater populations. Through whole genome-re-sequencing, deletions of genes in the synthetic Wood-Werkman cycle except *scpC* were identified in the control strain before SCRaMbLE, which might have resulted from the leaky expression of the pSCW11-Cre-EBD system. The three synthetic chromosomes were also lost in the two productive strains after SCRaMbLE, probably due to growth advantages associated with their loss. Given the absence of the PA pathway genes except *scpC* in the productive SCRaMbLE-sort isolates, PA production can be attributed to amino acid catabolic pathways degrading the supplemented amino acid mixture. In the future, a more precise control system of SCRaMbLE should replace the current SCRaMbLE system, such as recently developed systems that regulated by light (36), and different promoters for CRE expression (1), to reduce the leaky expression of Cre recombinase and increase the genome stability. During post-SCRaMbLE culturing and strain screening processes, PCR

assays of the pathway genes and the synthetic chromosomes need to be performed at intervals to confirm the presence of the corresponding genes. Additionally, the supplementation of extra amino acids should be avoided in future studies to prevent PA production from amino acid catabolic pathways. Alternative PA pathways could also be attempted to reduce the metabolic burden of yeast and to get a higher yield. In spite of the results in our study, SCRaMbLE has been demonstrated as a powerful tool to increase the productivity of the heterologous pathways and mining the underlying mechanisms (1, 2, 37, 38). As the Sc2.0 strain reaches completion, there will be a greater number of loxPsym sites available for generating much more recombination events. Thus, the potential of SCRaMbLE of the synthetic genome to improve the yield of PA is still to be explored.

Reference

1. Jia B, Wu Y, Li B-Z, Mitchell LA, Liu H, Pan S, et al. Precise control of SCRaMbLE in synthetic haploid and diploid yeast. Nature communications. 2018;9(1):1933.

2. Blount B, Gowers GF, Ho J, Ledesma-Amaro R, Jovicevic D, McKiernan R, et al. Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome. Nature communications. 2018;9(1):1932.

3. Goffeau A, Barrell BG, Bussey H, Davis R, Dujon B, Feldmann H, et al. Life with 6000 genes. Science. 1996;274(5287):546-67.

4. Hong K-K, Nielsen J. Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. Cellular and Molecular Life Sciences. 2012;69(16):2671-90.

5. Krivoruchko A, Siewers V, Nielsen J. Opportunities for yeast metabolic engineering: lessons from synthetic biology. Biotechnology journal. 2011;6(3):262-76.

6. Jensen MK, Keasling JD. Recent applications of synthetic biology tools for yeast metabolic engineering. FEMS Yeast Res. 2014;15(1):1-11.

7. Purnick PE, Weiss R. The second wave of synthetic biology: from modules to systems. Nature reviews Molecular cell biology. 2009;10(6):410.

8. Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new engineering rules for an emerging discipline. Molecular systems biology. 2006;2(1).

9. Lindstrom DL, Gottschling DE. The mother enrichment program: a genetic system for facile replicative life span analysis in Saccharomyces cerevisiae. Genetics. 2009.

10. Hoess RH, Wierzbicki A, Abremski K. The role of the loxP spacer region in PI site-specific recombination. Nucleic acids research. 1986;14(5):2287-300.

11. Shen Y, Stracquadanio G, Wang Y, Yang K, Mitchell LA, Xue Y, et al. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. Genome research. 2016;26(1):36-49.

12. Zidwick MJ, Chen J-S, Rogers P. Organic acid and solvent production: propionic and butyric acids and ethanol. The Prokaryotes: Springer; 2013. p. 135-67.

13. Álvarez-Chávez CR, Edwards S, Moure-Eraso R, Geiser K. Sustainability of bio-based plastics: general comparative analysis and recommendations for improvement. Journal of Cleaner Production. 2012;23(1):47-56.

14. Hebert RF. Stable indole-3-propionate salts of S-adenosyl-L-methionine. Google Patents; 2017.

15. Gonzalez-Garcia RA, McCubbin T, Navone L, Stowers C, Nielsen LK, Marcellin E. Microbial Propionic Acid Production. Fermentation. 2017;3(2):21.

16. Woskow SA, Glatz BA. Propionic acid production by a propionic acid-tolerant strain of Propionibacterium acidipropionici in batch and semicontinuous fermentation. Applied and Environmental Microbiology. 1991;57(10):2821-8.

17. Swick RW, Wood HG. The role of transcarboxylation in propionic acid fermentation. Proceedings of the National Academy of Sciences. 1960;46(1):28-41.

18. Kandasamy V, Vaidyanathan H, Djurdjevic I, Jayamani E, Ramachandran K, Buckel W, et al. Engineering Escherichia coli with acrylate pathway genes for propionic acid synthesis and its impact on mixed-acid fermentation. Applied microbiology and biotechnology. 2013;97(3):1191-200.

19. Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature. 2011;477(7365):471.

20. Williams TC, Pretorius IS, Paulsen IT. Synthetic evolution of metabolic productivity using biosensors. Trends in biotechnology. 2016;34(5):371-81.

21. Mitchell LA, Wang A, Stracquadanio G, Kuang Z, Wang X, Yang K, et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. Science. 2017;355(6329):eaaf4831.

22. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature protocols. 2007;2(1):31.

23. Williams TC, Xu X, Ostrowski M, Pretorius IS, Paulsen IT. Positive-feedback, ratiometric biosensor expression improves high-throughput metabolite-producer screening efficiency in yeast. Synthetic Biology. 2017;2(1):ysw002.

24. Dietmair S, Timmins NE, Gray PP, Nielsen LK, Krömer JO. Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. Analytical biochemistry. 2010;404(2):155-64.

25. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-9.

26. Lin Q, Qi H, Wu Y, Yuan Y. Robust orthogonal recombination system for versatile genomic elements rearrangement in yeast Saccharomyces Cerevisiae. Scientific reports. 2015;5:15249.

27. Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. Proceedings of the National Academy of Sciences. 1981;78(10):6354-8.

28. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, et al. Total synthesis of a functional designer eukaryotic chromosome. science. 2014;344(6179):55-8.

29. Hiraoka M, Watanabe K-i, Umezu K, Maki H. Spontaneous loss of heterozygosity in diploid Saccharomyces cerevisiae cells. Genetics. 2000;156(4):1531-48.

30. Tseng H-C, Harwell CL, Martin CH, Prather KL. Biosynthesis of chiral 3-hydroxyvalerate from single propionate-unrelated carbon sources in metabolically engineered E. coli. Microbial cell factories. 2010;9(1):96.

31. Jiang M, Pfeifer BA. Metabolic and pathway engineering to influence native and altered erythromycin production through E. coli. Metabolic engineering. 2013;19:42-9.

32. Staib L, Fuchs TM. Regulation of fucose and 1, 2-propanediol utilization by Salmonella enterica serovar Typhimurium. Frontiers in microbiology. 2015;6:1116.

33. Cameron DC, Cooney CL. A Novel Fermentation: The Production of R (–)–1, 2–Propanediol and Acetol by Clostridium thermosaccharolyticum. Nature Biotechnology. 1986;4(7):651.

34. d Mattozzi M, Ziesack M, Voges MJ, Silver PA, Way JC. Expression of the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: Toward horizontal transfer of autotrophic growth. Metabolic engineering. 2013;16:130-9.

35. Berg IA, Kockelkorn D, Buckel W, Fuchs G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. Science. 2007;318(5857):1782-6.

36. Hochrein L, Mitchell LA, Schulz K, Messerschmidt K, Mueller-Roeber B. L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. Nature communications. 2018;9(1):1931.

37. Liu W, Luo Z, Wang Y, Pham NT, Tuck L, Pérez-Pi I, et al. Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods. Nature communications. 2018;9(1):1936.

38. Wu Y, Zhu R-Y, Mitchell LA, Ma L, Liu R, Zhao M, et al. In vitro DNA SCRaMbLE. Nature communications. 2018;9(1):1935.

Chapter 5 : Evolutionary engineering in *Saccharomyces cerevisiae* reveals a *TRK1*-dependent potassium influx mechanism for propionic acid tolerance

Biotechnology for Biofuels

RESEARCH Open Access

Evolutionary engineering in *Saccharomyces cerevisiae* reveals a *TRK1*-dependent potassium influx mechanism for propionic acid tolerance

Xin Xu[1], Thomas C. Williams[1,2*], Christina Divne[3], Isak S. Pretorius[1] and Ian T. Paulsen[1*]

Abstract

Background: Propionic acid (PA), a key platform chemical produced as a by-product during petroleum refining, has been widely used as a food preservative and an important chemical intermediate in many industries. Microbial PA production through engineering yeast as a cell factory is a potentially sustainable alternative to replace petroleum refining. However, PA inhibits yeast growth at concentrations well below the titers typically required for a commercial bioprocess.

Results: Adaptive laboratory evolution (ALE) with PA concentrations ranging from 15 to 45 mM enabled the isolation of yeast strains with more than threefold improved tolerance to PA. Through whole genome sequencing and CRISPR–Cas9-mediated reverse engineering, unique mutations in *TRK1*, which encodes a high-affinity potassium transporter, were revealed as the cause of increased propionic acid tolerance. Potassium supplementation growth assays showed that mutated *TRK1* alleles and extracellular potassium supplementation not only conferred tolerance to PA stress but also to multiple organic acids.

Conclusion: Our study has demonstrated the use of ALE as a powerful tool to improve yeast tolerance to PA. Potassium transport and maintenance is not only critical in yeast tolerance to PA but also boosts tolerance to multiple organic acids. These results demonstrate high-affinity potassium transport as a new principle for improving organic acid tolerance in strain engineering.

Keywords: Adaptive laboratory evolution, Organic acid tolerance, Propionic acid, *TRK1*, Potassium uptake, Yeast

Background

Propionic acid (PA), a key building-block chemical, has been widely used as a food preservative and a chemical intermediate in plastics, pharmaceutical, cosmetic, paint, and herbicide industries [1–4]. According to the U.S. Department of Energy, PA is one of the top 30 candidate platform chemicals in use [5]. The global production of PA reached 400 kilotons in 2014 with USD 1.07 billion in revenue and was predicted to increase to USD 1.55 billion by 2020 [6]. However, as with most industrial chemicals, PA is currently derived from finite oil reserves in an environmentally-destructive manner. Microbial fermentation is a sustainable solution to petrochemical refining with the advantage that it is environmentally friendly, and can utilize renewable biomass as a feedstock [7]. Propionibacteria can produce PA from sugars via the Wood–Werkman cycle [8]. Different culture conditions and fermentation modes have been explored to increase the PA yield in propionibacteria [9]. However, further improving PA production in propionibacteria is challenging due to a lack of genetic engineering tools and strain characterization [10]. In addition, production is also constrained by slow growth rates, growth inhibition in acidic conditions, and high product purification costs [11]. Due to these limitations, it is of great interest to engineer platform microorganisms for the heterologous production of PA. *Escherichia coli* has been engineered with the

*Correspondence: tom.williams@mq.edu.au; ian.paulsen@mq.edu.au
[1] Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia
Full list of author information is available at the end of the article

acrylate pathway of *Clostridium propionicum* to synthesize PA but the titer was only 3.7 ± 0.2 mM [11].

In contrast to native PA producers, the yeast *Saccharomyces cerevisiae* is a robust cell factory that can grow at relatively low pH and can be easily manipulated using advanced genetic tools. Yeast has been engineered for the biotechnological production of various organic acids, such as lactic acid [12], succinic acid [13], 3-hydroxypropionic acid (3-HP) [14], and muconic acid [15]. PA has also been detected previously as a by-product in fermentation of *S. cerevisiae* [16]. Yeast is therefore a promising candidate for engineering PA production from sugars, and potentially from cellulosic biomass. However, product toxicity is a problem equal in significance to product yield optimization in microbial organic acid production. At external pH below the pKa of a weak acid, the undissociated (protonated) form of the acid can pass through the plasma membrane freely. In the near-neutral cytoplasm, it dissociates and releases the protons and counterions. The protons lead to intracellular acidification that affects internal pH homeostasis, lipid organization, and the function of cellular membranes [17–19]. In addition, the accumulation of anions is also toxic to yeast cells. To reduce stress, yeast cells increase proton export via plasma membrane and vacuolar H^+-ATPases to maintain pH homeostasis in response to multiple organic acids [20–23]. Through transcriptomic analysis, several transcriptional regulators have been identified that mediate the response to organic acid stress in yeast. Overexpression of the Haa1p transcription factor enhanced acetic acid tolerance in yeast [24]. Multidrug resistance transporters and remodelling of the cellular envelope are also involved in weak acid detoxification [20]. For example, the ATP-binding cassette (ABC) transporters Pdr12p and Pdr5p have been proposed to be implicated in the efflux of the toxic counterions of hydrophilic and lipophilic weak acids [18, 20, 25, 26]. *SPI1*, encoding a glycosylphosphatidylinositol (GPI)-anchored cell wall protein, was identified to play a key role in yeast response to 2,4-dichlorophenoxyacetic acid [27], octanoic acid and benzoic acid [28].

In previous studies, the response of yeast to PA has been studied to investigate the resistance mechanism to organic acid food preservatives. Genome-wide screening of the yeast knock-out library identified that the transcription factor Rim101p, which is involved in maintenance of pH homeostasis and cell wall remodelling, is required for yeast resistance to PA stress [29]. Subsequently, metabolomics revealed changes in the abundance of amino acids, ATP, NAD^+, glycerol, and trehalose in yeast exposed to PA [30]. However, despite the regulation by *RIM101*, the underlying mechanism of the PA response is yet to be fully understood, which impedes the development of yeast cell factories for PA production. Furthermore, despite these innate PA response mechanisms, yeast tolerance to PA is far below the levels required for industrial fermentation.

Adaptive laboratory evolution (ALE) is a widely used method to gain insights into the mechanisms of evolution and adaptive changes that accumulate under defined growth conditions for prolonged periods of selection [31]. ALE has, therefore, proven to be a powerful tool in the field of metabolic engineering, both for the elucidation of new design principles and the engineering of superior production strains. Applications of ALE include improving the growth rate [32] and product yield [33], adaption of strains to utilize non-native substrates or produce non-native products [34–36], and increasing the tolerance of strains towards a specific environmental stress [37, 38]. Combined with systems biology approaches, the relationship between genomic changes and the adaptive phenotype can be discovered by whole genome re-sequencing [39]. ALE has also been applied successfully in generating tolerant yeast strains to weak acids such as acetic acid [40], 3-HP [41], and lactic acid [42]. Despite the attractiveness of yeast-based PA production, no study has been performed to engineer the cell factory yeast with prolonged PA tolerance and mechanisms for improving yeast tolerance to PA have not been explored using ALE. Here, we demonstrate improved PA tolerance in yeast using ALE, and decipher the molecular mechanism of tolerance by whole genome re-sequencing and functional analysis.

Results

Adaptive laboratory evolution of propionic acid tolerance

Before the ALE, an initial growth test of *S. cerevisiae* CEN.PK 113-7D with PA concentrations ranging from 0 to 25 mM was conducted to identify inhibitory concentrations of PA. At 15 mM of PA, the growth rate of CEN.PK 113-7D was nearly halved, and at 25 mM of PA, the growth of the parental strain was significantly affected (Additional file 1: Fig. S1). Thus, 15 mM of PA was used as the starting concentration for ALE. Three different conditions were used for ALE in parallel: minimal medium (pH 5), buffered minimal medium (pH 3.5), and PA treated (pH 3.5). The minimal medium and buffered minimal medium acted as controls for mutations arising from genetic drift in minimal medium and from tolerance to low-pH medium, respectively (Fig. 1a). The concentration of PA was increased to 20 mM, 25 mM, 35 mM, 40 mM and 45 mM during ALE (Fig. 1a). Finally, at 45 mM, no further growth improvement was observed, and the experiment was stopped after 64 days. Fluctuations in cell density during the evolution were recorded (Additional file 1: Fig. S2) using optical density at 600 nm

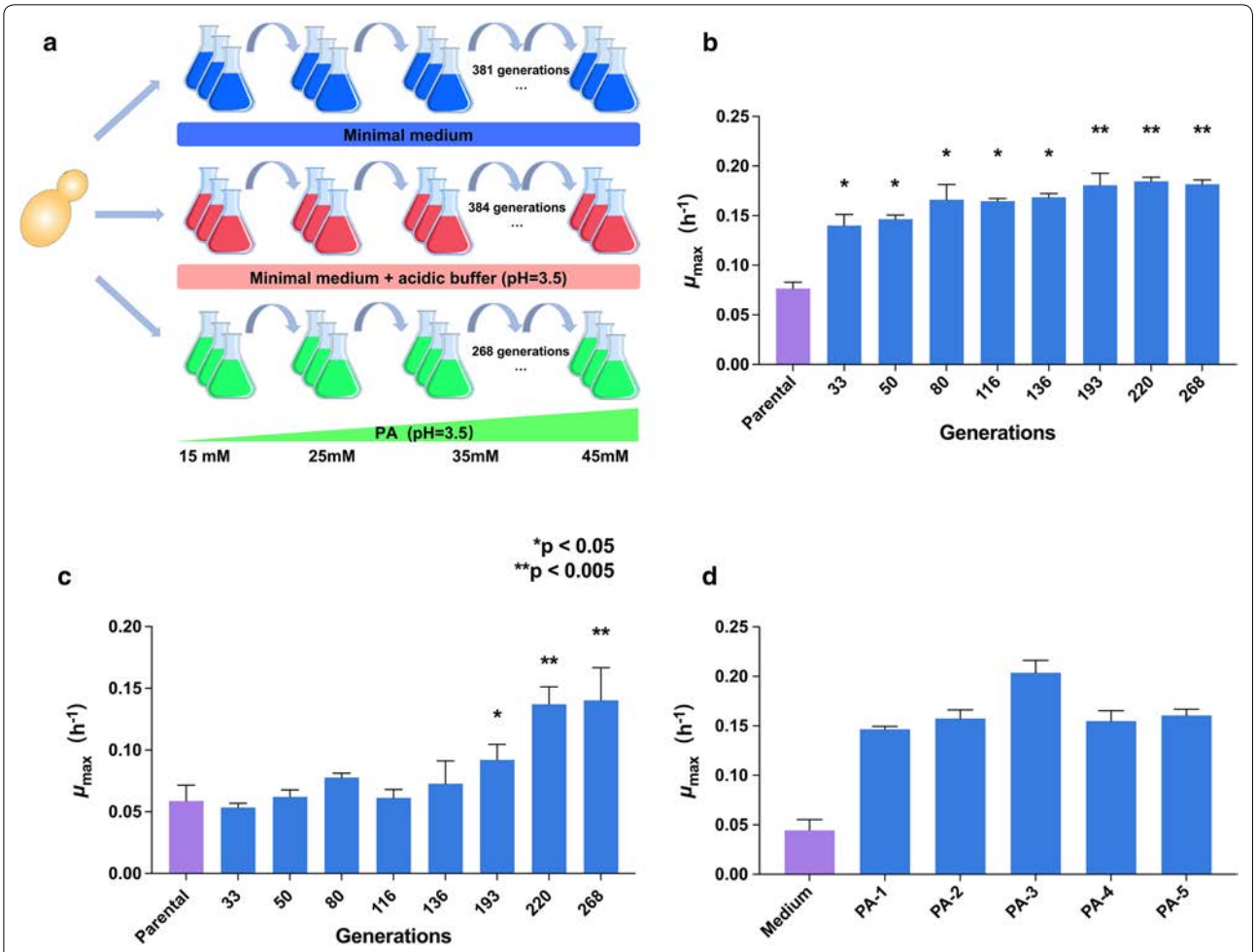


Fig. 1 Overview of the ALE process and improved PA tolerance. Three conditions were used for the ALE experiment: minimal medium (pH 5); buffered minimal medium (pH 3.5); buffered minimal medium with increasing concentrations of PA (pH 3.5). Three independent cultures were grown under each condition (**a**). Growth rates of PA-evolved populations sampled through ALE were determined in buffered minimal medium containing 25 mM PA (**b**) and 35 mM PA (**c**). Growth rates of isolates from end-point PA-evolved populations (PA-1, PA-2, PA-3, PA-4, and PA-5) were determined in buffered minimal medium containing 35 mM PA. "Medium" represents the isolate from end-time minimal medium-evolved population (**d**). Bars and error bars represent the mean and standard deviation (SD) of triplicate cultures

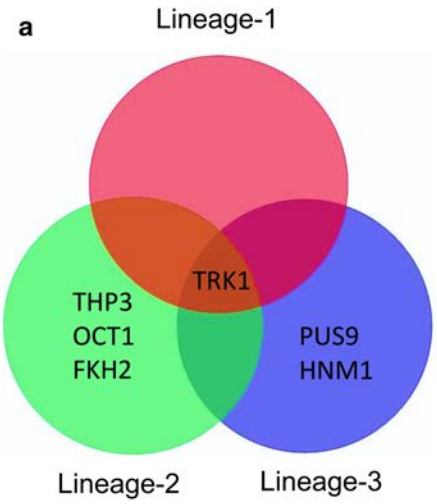
(OD_{600nm}). The CEN.PK 113-7D strain was cultured for approximately 381 generations in minimal medium (pH 5), for 384 generations in buffered minimal medium (pH 3.5), and for 268 generations in the PA-stressed conditions (pH 3.5) (Fig. 1a). After 33 generations, the PA evolved populations ($\mu_{max} \approx 0.14\ h^{-1}$) showed a slightly improved fitness relative to the parental strains ($\mu_{max} \approx 0.08\ h^{-1}$) in buffered minimal medium supplemented with 25 mM PA (Fig. 1b). After 193 generations, the maximum specific growth rates of the evolved populations treated with 25 mM and 35 mM PA were 0.18 h^{-1} and 0.09 h^{-1}, respectively, which were significantly faster than the parental strain (Fig. 1b, c). At the end of the ALE experiment, the growth rate of evolved populations treated with 35 mM PA was 2.4-fold greater than the parental strain (Fig. 1c). Five strains were isolated from end-point PA-evolved populations and designated as PA-1, PA-2, PA-3, PA-4, and PA-5. The PA-evolved strains showed 3.3–4.6-fold improved tolerance relative to the strains isolated from the minimal-medium control population (pH 5) (Fig. 1d).

Whole genome sequencing of evolved populations and isolates

To link the PA tolerance phenotype with the genotype, whole genome re-sequencing was performed on the CEN.PK 113-7D parental strain, end-point medium control, end-point pH 3.5 buffered medium control, PA-evolved

populations sampled at different generations and five PA-evolved isolates from the end-point populations. Reads from the sequenced genomes were mapped to the reference genome of CEN.PK 113-7D. Non-synonymous mutations in the parental strain and mutations in *PEX19*, *SNO4*, *KEL3* and *ALD6* that were also present in minimal medium or buffered minimal medium conditions were excluded. Non-synonymous mutations in the *TRK1* gene, encoding a high-affinity potassium transporter, were common to all PA-evolved lineages and PA-evolved isolates, suggesting that it might be the key determinant of the tolerance phenotype (Fig. 2a). Different point mutations encoding single-amino acid substitutions in *TRK1* were identified in the three independent evolved lineages: a nucleotide at position 3223 was changed from C to A (H1074N at the amino acid level) in lineage-1, a nucleotide at position 2981 was changed from G to C (R993P) in lineage-2, and a nucleotide at position 3509 was changed from C to A (A1169E) in lineage-3 (Fig. 2b). The mutations appeared after 80 generations of evolution, accounting for 25.0–47.8% of the reads from the whole evolved population (Fig. 2b). The variant frequencies became higher with increasing PA concentrations throughout the ALE experiment. At the end of the evolution experiment, the strains that contained the mutated *TRK1* alleles occupied from 71.3 to 99.1% of the whole PA-evolved populations, according to read–variant frequency (Fig. 2b). In lineage-2, non-synonymous single-nucleotide mutations were also found in *THP3, OCT1,* and *FKH2*. In lineage-3, a non-synonymous single-nucleotide mutation was also found in *PUS9*, and a premature stop-codon was identified in *HNM1*. The mutations and the functions of these genes are shown in Additional file 1: Table S3. No deletions, insertions, or duplications were observed in coding sequence (CDS) regions.

a



b

Lineage	Mutation position (Amino acid level)	Variant Frequency (%)		
		G80	G220	G268
1	H1074N	25.0	99.0	88.1
2	R993P	43.2	99.2	99.1
3	A1169E	47.8	83.7	71.3

Fig. 2 Identification of single-nucleotide polymorphism (SNPs) in PA-evolved lineages. After sequencing, Geneious Pro 9.1.3 software was used to map the reads to the CEN.PK113-7D reference genome and subsequently to perform variant calling. Non-synonymous polymorphisms in CDS regions were identified. Mutations that were also present in the parental strain, minimal medium or buffered minimal medium conditions were excluded. The Venn diagram shows that the genes with non-synonymous mutations that were identified from three end-point PA-evolved populations (**a**). The table presents variant frequencies of Trk1p in three PA-evolved lineages sampled at different generations. G represents the generation number at which the population was sequenced (**b**)

Reverse engineering of *TRK1* mutations

To determine whether the single-nucleotide mutations in *TRK1* confer tolerance to PA, three different point mutations were independently reverse engineered into the parental strain to replace the original sequence. The mutations in the resulting strains, ($TRK1^{H1074N}$, $TRK1^{R993P}$, and $TRK1^{A1169E}$) were confirmed using Sanger sequencing. The three reverse engineered strains showed 2.4–3-fold increased growth rates relative to the parental strain when treated with 35 mM PA at a level not significantly different to the evolved strains (Fig. 3a), fully recovering the evolved PA tolerance phenotype. To test if the *TRK1* mutations had any additive or synergistic affects, the evolved isolate PA-3 was engineered with each of the two other mutations to create the double mutant strains $TRK1^{R993P, H1074N}$ and $TRK1^{R993P, A1169E}$. However, no further growth-rate increase was detected under PA stress (Additional file 1: Fig. S3). Overexpression of *TRK1* was achieved by expressing the *TRK1* gene with the strong constitutive *PGK1* promoter and the *CYC1* terminator on a centromeric plasmid. The overexpression strain had an increased growth rate when treated with 35 mM PA, while the *trk1Δ* strain did not grow at all (Fig. 3b).

Interpretation of *TRK1* mutations

TRK1, which encodes a plasma membrane protein required for high-affinity potassium transport [43], belongs to the Trk/Ktr/HKT superfamily of ion transporters. Its structural features include four MPM motifs, where "M" denotes transmembrane helix, and "P" refers to the re-entrant pore loop preceding the selectivity filter. Each motif (Fig. 4) contains an M1 helix, a pore helix (MP), the ion-selectivity filter motif (black asterisks in Fig. 4), and helices M2a and M2b [44]. To facilitate understanding of the identified mutations, we set out to generate a homology model of *S. cerevisiae* Trk1p (*Sc*Trk1p) (Additional file 1: Fig. S4) based on a structural alignment of the two homologous transporters, i.e., *Bacillus subtilis* KtrB (*Bs*KtrB) [45] and *Vibrio parahaemolyticus* TrkH (*Vp*TrkH) [46]. The sequence identity with *Sc*Trk1p is only 22% in *Bs*KtrB and 11% in *Vp*TrkH; however, some regions are relatively well conserved, especially the helices and pore filter regions. Based on the alignment, the mutation from lineage-1 (H1074N) is positioned in helix M1 of the D motif. This position corresponds to Lys350 in *Bs*KtrB, which plays a role in securing the C-terminal tail by interactions with Lys315 and Asn119; the mutation from lineage-2 (R993P) would be at the end of the intramembrane loop (IML) and the start of helix M2b in motif C, which is likely to induce rigidity in the backbone conformation and would affect the structure and accessibility of the cytoplasmic pore. This position may also form interactions with the C-tail of another Trk1p monomer. The mutation from lineage-3 (A1169E) coincides with the first residue of the C-terminal tail. The first 15 residues of the C-tail in *Bs*KtrB ensure homodimer formation and functional association with *Bs*KtrA. Replacing the corresponding alanine in Trk1p (Ala1169) with a glutamate side chain would enable ionic interactions or hydrogen bonds that are potentially stronger than the hydrophobic interactions offered by alanine, and possibly increase oligomer stabilization [47, 48].

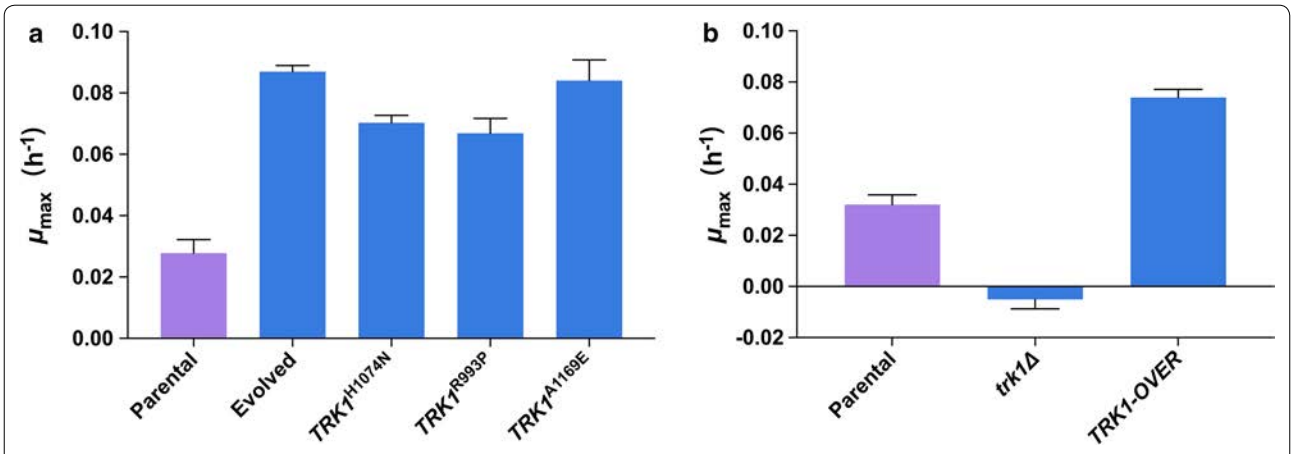


Fig. 3 Fitness test of *TRK1* mutants treated with 35 mM PA. Three different *TRK1* point mutations were reverse engineered into the parental strain CEN.PK 113-7D to generate the $TRK1^{H1074N}$, $TRK1^{R993P}$, and $TRK1^{A1169E}$ strains. Growth rates of these three strains carrying *TRK1* point mutations, the end-point PA-evolved isolate PA-1 (evolved), and the parental strain were determined in buffered minimal medium containing 35 mM PA (**a**). Growth rates of a *TRK1* deletion strain (*trk1Δ*), a *TRK1* overexpression strain (*TRK1-OVER*) and the parental strain were measured in buffered minimal medium containing 35 mM PA (**b**). Bars and error bars represent the mean and SD of triplicate cultures

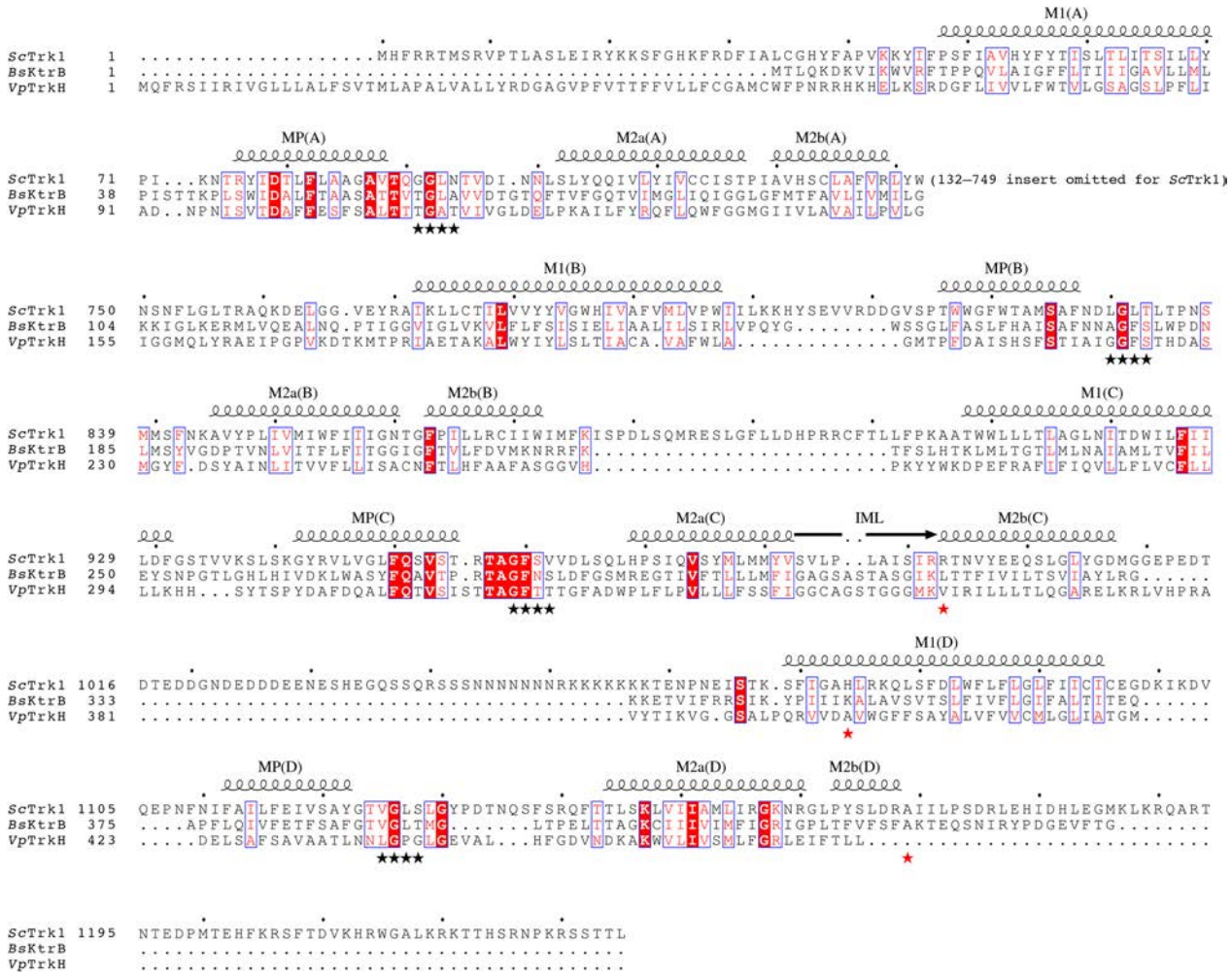


Fig. 4 Structure alignment of *Sc*Trk1 with its homologs with known structures. The alignment was based on a structural superposition of the crystal structures of *Bs*KtrB (PDB code 4J7C [PMID: 23598340]) and *Vp*TrkH (PDB code 3PJZ [PMID: 21317882]). The selectivity filters are marked by black asterisks, the three mutations by red asterisks, and the intramembrane loop is denoted as IML. For clarity, the extensive insert (residues 132-749) in the *Sc*Trk1 sequence has been omitted. Red filled boxes indicate strict identity, and blue-framed boxes represent similarity across all aligned sequences

Effect of potassium concentration on PA tolerance

To explore the effect of potassium concentrations and the function of *TRK1* on yeast tolerance to PA, WT, *trk1Δ* and three evolved isolates (PA-1, PA-3, and PA-4) were spotted on yeast nitrogen base (YNB) agar supplemented with increasing concentrations of potassium (0.1 mM, 1 mM, 10 mM, and 100 mM) and with 0 mM, 15 mM and 25 mM PA. On YNB agar with 0 mM PA, the three evolved strains and the WT strain had similar growth profiles at a broad potassium concentration range (from 0.1 to 100 mM) (Fig. 5). With PA added into the agar, the three evolved strains grew much faster than the WT strain, especially at 0.1 mM, 1 mM and 10 mM potassium. *Trk1Δ* generally grew slower, and it showed an obvious growth defect when PA was supplemented. Through adding increasing concentrations of potassium, growth could be restored. The growth assay was also conducted in liquid medium, which showed the same trend of improved growth with potassium supplementation under PA stress, and with *TRK1* mutation (Additional file 1: Fig. S5). These results indicate the function of *TRK1,* which encodes the high-affinity potassium transporter, is essential for the tolerance to PA in yeast, and that the evolved *TRK1* mutations enable a higher uptake of potassium under PA stress.

The effect of potassium on tolerance to other organic acids

We further explored whether the tolerance mechanism can be applied to other organic acids, including the main inhibitor in lignocellulose hydrolysate, acetic acid, and

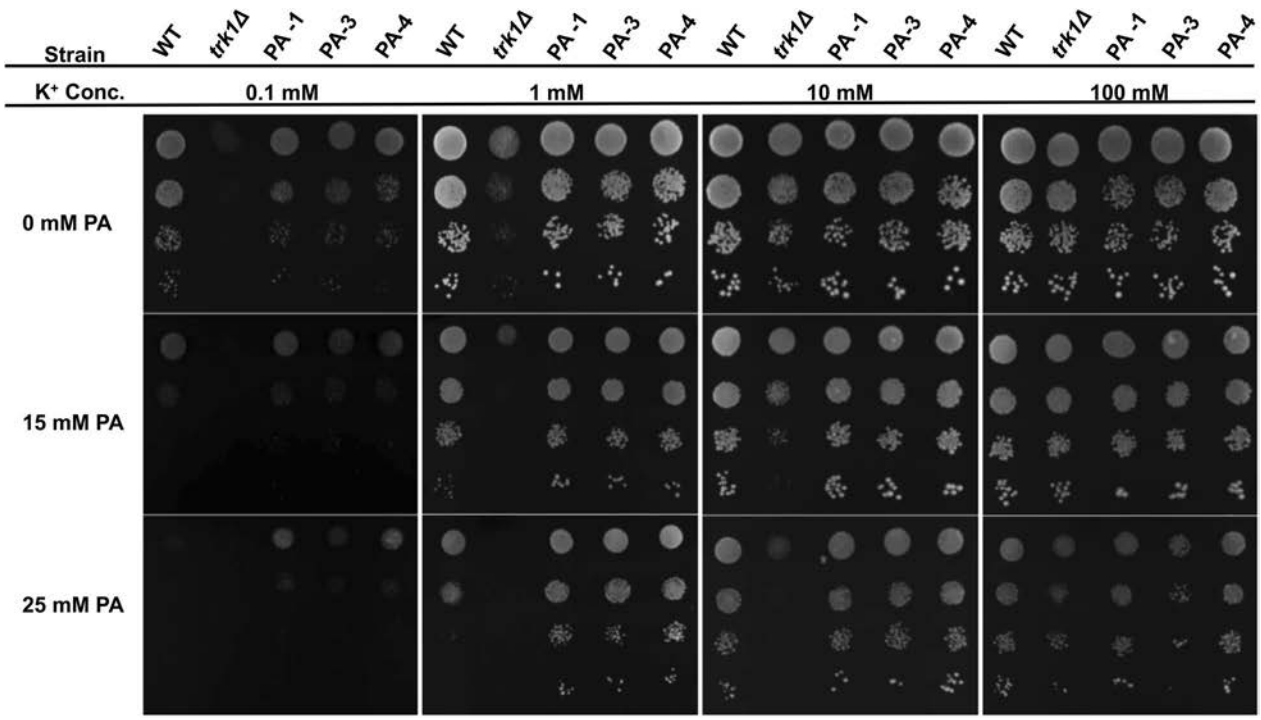


Fig. 5 The effect of potassium concentrations and *TRK1* on the tolerance of yeast strains to PA. WT, *trk1Δ* and three end-point-evolved isolates (PA-1, PA-3, and PA-4) were grown to an OD_{600} = 0.2, serially diluted to 1:0, 1:10, 1:100, 1:1000, and then spotted on YNB agar supplemented with increasing concentrations of potassium (0.1 mM, 1 mM, 10 mM, and 100 mM) with 0 mM, 15 mM, and 25 mM PA. The plates were incubated for 3 days at 30 °C

three other valuable organic acids also used as food preservatives (lactic acid, benzoic acid, and sorbic acid). WT, *trk1Δ* and the evolved isolate PA-3 were tested on agar containing 1 mM and 100 mM potassium, without additional acids and with 83 mM acetic acid, 111 mM lactic acid, 2 mM benzoic acid, and 2 mM sorbic acid. In the no acid control, the evolved strain showed the same fitness as the WT at 1 mM and 100 mM potassium, while at 1 mM potassium the *trk1Δ* had a growth defect (Fig. 6). *TRK1* deletion resulted in sensitivity to all the organic acids tested. At 1 mM potassium, the evolved strain had much improved growth under the organic acid stress relative to the parental strain, especially when treated with 83 mM acetic acid, 2 mM benzoic acid, and sorbic acid. At 100 mM potassium, the evolved strain still had increased fitness relative to the WT under acetic acid and lactic acid stress, while with benzoic acid and sorbic acid, the growth of the WT and *trk1Δ* strains was recovered.

Discussion

This study demonstrates ALE combined with whole genome re-sequencing and functional analysis can be used to unveil non-intuitive mechanisms of PA tolerance in yeast. During ALE, the concentration of PA was increased gradually to allow the accumulation of beneficial mutations as well as maintain the growth of evolved populations (Fig. 1a). The evolved strains with improved tolerance require a pre-adaption to low concentrations of PA before treatment with 35 mM PA. This could be interpreted as the cells requiring activation of the expression of inducible genes that contribute to the improved tolerance phenotype. In a previous study, *S. cerevisiae* acquired improved tolerance through laboratory evolution after continuous growth with increasing concentrations of acetic acid. However, the tolerance was rapidly lost upon storage at −80 °C and growth without acetic acid [49]. Batch fermentation with alternating transfer to medium with and without acetic acid yielded evolved strains with constitutive tolerance. *ASG1*, *ADH3*, *SKS1* and *GIS4* were identified for the improvement of acetic acid tolerance. This on-and-off strategy allows the selection of tolerant strains that can initiate growth without pre-adaption, whereas it may also result in decreased fitness during the off phase [40].

The unique mutations in response to PA stress were obtained after excluding mutations that also existed in the parental strain, the medium-evolved condition, and low-pH medium-evolved condition (Fig. 2a).

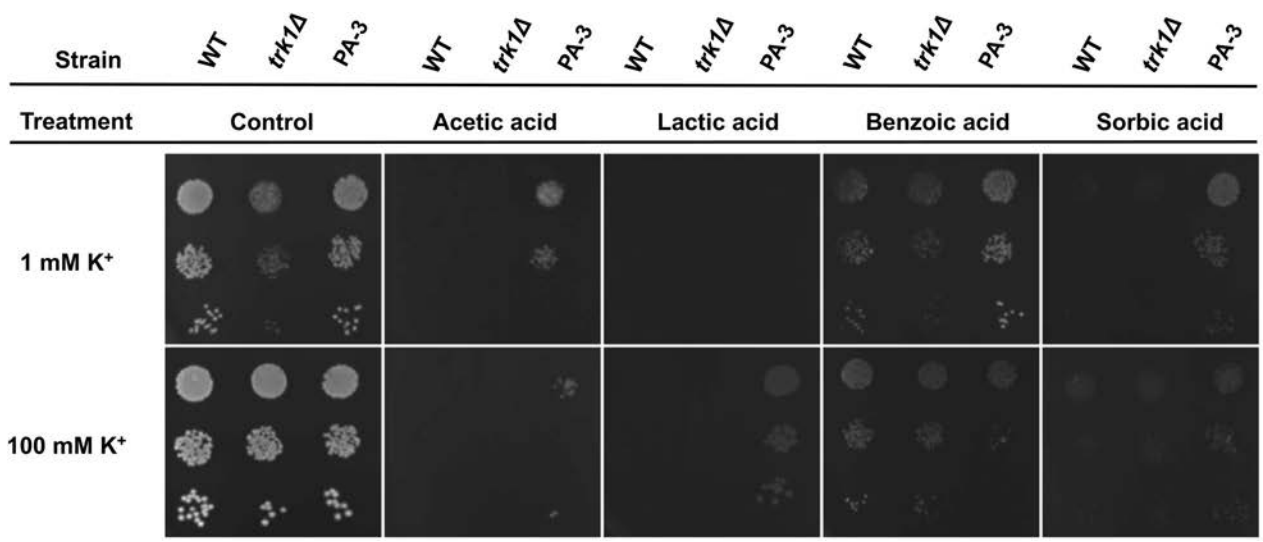


Fig. 6 The effect of potassium concentrations and *TRK1* on the tolerance to other organic acids. WT, *trk1Δ* and the end-point-evolved isolate (PA-3) were grown to an OD_{600} = 0.2, serially diluted to 1:10, 1:100, 1:1000, and then spotted on YNB agar containing 1 mM and 100 mM potassium, without additional acids and with 83 mM acetic acid, 111 mM lactic acid, 2 mM benzoic acid, and 2 mM sorbic acid. The plates were incubated for 3 days at 30 °C

Non-synonymous mutations were found in six genes, of which *TRK1* were shared by all PA-challenged lineages and isolates (Fig. 2a). *TRK1* encodes for the high-affinity transporter for potassium uptake [43]. Three different point mutations encoding amino acid substitutions in *TRK1* were identified in the three independent evolved lineages. By reverse engineering each of the three mutations, the resulting strains fully recovered the PA tolerance phenotype, demonstrating *TRK1* is the genetic determinant (Fig. 3a). PA tolerance was not further enhanced by engineering combinations of two *TRK1* mutations in a single strain, demonstrating that the effects of the three different *TRK1* mutations on PA tolerance are not additive (Additional file 1: Fig. S3). Overexpression of the native *TRK1* gene improved the growth rate in the PA-challenged condition, showing that a higher expression level of *TRK1* also enables PA tolerance (Fig. 3b).

The other five mutated genes identified from the evolved populations only appeared in single PA-evolved lineages. A non-synonymous single-nucleotide mutation in *FKH2* and a truncation of *HNM1* were identified in evolved lineage-2 and lineage-3, respectively. Although these gene mutations were not shared among different lineages, it is noticeable that *HNM1* and *FKH2* were observed to be stress responsive in previous studies. *HNM1*, encoding a plasma membrane transporter for choline, was significantly downregulated in exposure to a multifunctional drug, pyrrolidine dithiocarbamate [50]. Fkh2, a Forkhead family transcription factor regulating the cell cycle [51], is involved in oxidative, heat [52] and acetic acid stresses [53].

A refined structural model of Trk1 and experimental validations showed that glycines within the selectivity filter are important for protein function and correct folding/membrane targeting [44]. Specifically, mutating these glycines into bulky residues leads to a growth defect in low concentrations of KCl. When highly conserved residues, such as D79 and K1147 within the $M2_D$ helix and the P_A loop, were mutated to alter the charges, severe protein folding and function defects were observed as well [44]. Our sequence alignment of *Sc*Trk1p against *Bs*KtrB and *Vp*TrkH agrees well with the previous study for the more conserved regions, while there are discrepancies for other regions [44]. In our study, the point mutations H1074N, A1169E and R993P were located in helix M1, in the C-terminal tail and within or close to the intramembrane loop, respectively (Fig. 4). All appeared to play an important role in interactions associated with the C-tail, and probably increase oligomer stabilization. Given that the mutated residues are also associated with interactions that affect the pore gate, we would assume that the mutations also affect transporter function. Our results provide new insights to understand and modulate Trk1p function.

Through potassium supplementation growth assays, it was evident that insufficient potassium uptake by *trk1Δ* and the parental strain lead to susceptibility to PA stress (Fig. 5). Supplementation of higher concentrations of potassium recovered the growth and

improved yeast tolerance to PA. The evolved strains showed much improved growth under PA stress, especially at low potassium concentrations. Our results suggest that the mutated *TRK1* alleles of the evolved strains improve the uptake and maintenance of intracellular potassium ions, which facilitate the detoxification of PA. When other organic acids were tested, the evolved *TRK1* mutation also conferred tolerance to acetic acid, lactic acid, benzoic acid, and sorbic acid (Fig. 6). Increasing the concentration of extracellular potassium to 100 mM enabled the growth of the WT and *trk1Δ* strains under benzoic acid and sorbic acid stresses, while it decreased the fitness of the evolved mutant slightly, indicating there is a suitable potassium concentration range for tolerance to these organic acids.

Deletion of genes related with potassium import (Trk1p, Nha1p, Arl1p) has been reported to result in susceptibility to acetic acid [54]. The food spoilage yeasts *Zygosaccharomyces bailii* and *S. cerevisiae* have been observed to accumulate potassium during long-term adaption to benzoic acid [55]. *TRK1* deletion in *S. cerevisiae* confers sensitivity to benzoic acid [55] and sorbic acid [56]. Our potassium supplementation assay is consistent with the previous studies above and also revealed the significance of potassium homeostasis in *S. cerevisiae* tolerance to lactic acid for the first time. Interestingly, although genome-wide screening of the *S. cerevisiae* knock-out mutant library demonstrated that vacuolar function and the *RIM101* pathway are important for *S. cerevisiae* resistance to PA; none of the genes identified in our study were identified previously [29]. This is not surprising given that our aims and experimental conditions were different. The previous study focused on the identification of gene deletions resulting in PA susceptibility upon short exposure to PA with the auxotrophic BY4741 strain and the Euroscarf collection of mutant strains [29]. However, our study was aimed at improving prolonged PA tolerance for future PA production in yeast, and the prototrophic CEN.PK 113-7D strain was used for ALE, which is a more robust and industrially relevant strain with thousands of genetic differences to the BY4741 strain [57].

In this study, a long-term PA tolerance mechanism was identified through ALE, and yeast tolerance to PA was demonstrated to positively correlate with the extracellular concentration of potassium and the expression level or activity of Trk1p for the first time. The tolerance mechanism identified in our study will pave the way for future engineering of PA production in yeast. *TRK1* mutations or extracellular potassium supplementation not only confers tolerance to PA but also boosts tolerance to multiple organic acids, many of which are valuable products or potent growth-inhibitors found in lignocellulose hydrolysate. In particular, acetic acid is formed from the degradation of hemicelluloses during lignocellulose pre-treatment [58]. These evolved *TRK1* mutations are, therefore, highly significant for future biorefining of organic acids and lignocellulose-based biofuel production.

One of the major weak acid toxicity mechanisms is the acidification of the cytoplasm. The export of protons relies on the activation of *PMA1*, which encodes for an H^+-ATPase, and the H^+ efflux is coupled with K^+ influx. It has been reported that potassium uptake by the Trk system promotes the export of protons and alkalinizes cytosolic pH [59]. Furthermore, the undissociated form of PA impairs plasma membrane integrity, which leads to the disruption of ionic gradients and the increase of ion leakage. We hypothesized that the supplementation of higher concentrations of extracellular potassium or the mutations of *TRK1* might function in two major ways: first, they promote the export of protons, reducing the toxicity of intracellular acidification; second, they strengthened transmembrane ion gradients (proton and potassium gradients), helping to stabilize the membrane potential and improve ion homeostasis. Potassium uptake by yeast has also been linked to protein synthesis, activation of enzymes [60, 61], regulation of oxidative phosphorylation [62], DNA replication, and the cell cycle [59, 63], which all play critical roles in weak acid tolerance [55]. In addition to tolerance to weak acids, it has been observed that KCl and KOH supplementation boosted ethanol production from yeast by improving tolerance to ethanol. The tolerance was not limited to ethanol, but also observed for higher chain alcohols, which results from strengthening the opposing potassium and proton electrochemical membrane gradient [64]. Trk1p also increases tolerance to cations by preventing cation import. Heterologous expression of *TRK1* from *Zygosaccharomyces rouxii* confers high lithium tolerance in *S. cerevisiae* [65]. The functions of Trk1p have not been fully discovered and it may play an important role in tolerance to other stress conditions. These results, along with our own, demonstrate the complex role that Trk1p plays in stress tolerance in yeast, including organic acid tolerance and alcohol tolerance, with *TRK1* knock-out, overexpression, and amino acid substitutions giving different phenotypes in different stress conditions. *TRK1* and its homologs are emerging as critical components that must be considered in weak acid tolerance engineering.

Conclusion

The results presented in this paper show that ALE can be used for the accumulation of beneficial mutations and the selection of yeast strains that are more tolerant

to PA. Three convergent mutations in *TRK1* were identified using whole genome sequencing of independent laboratory evolved lineages and reverse engineering of the ancestral strain. The PA detoxification mechanism is dependent on potassium uptake and accumulation enabling maintenance of pH homeostasis and stabilization of membrane potential, and this mechanism can be applied to multiple organic acids. In the future, the toxicity associated with the bioproduction of organic acids could be reduced by engineering *TRK1,* or by increasing the potassium concentration in the fermentation medium. Conceivably, these two approaches could be employed simultaneously to further increase organic acid tolerance in a production scenario, although the use of mutated *TRK1* is preferable given that it avoids the cost of additional potassium supplementation, as well as the biological burden of *TRK1* overexpression. This work, therefore, reveals a new principle for strain engineering in the production of PA and multiple organic acids, as well as reducing the toxicity of organic acids in the fermentation of lignocellulosic hydrolysates in yeast.

Methods

Strains and media

Yeast strains were grown in liquid yeast peptone dextrose (YPD) broth at 30 °C, 200 rpm, streaked out on YPD agar, and incubated at 30 °C for 1 day to maintain the strains. Strains transformed with the *hphMX* marker gene were selected on YPD agar plates supplemented with 300 μg/mL hygromycin. Minimal medium containing 1× yeast nitrogen base without amino acids mix (Sigma-Aldrich Y0626), with 1% glucose (pH 5) was used for ALE. For the buffered control and PA-treated conditions, 174 mL of 0.5 M citric acid and 140 mL of 0.5 M Na_2HPO_4 solution were added per liter of minimal medium (pH 3.5). For potassium supplementation assays, the YNB agar was made of TRANSLUCENT K^+-free medium (Formedium, CYN7501), with 1% of glucose, 2.4% agar and appropriate concentrations of potassium.

All the cloning experiments were conducted with *E. coli* DH5α. The *E. coli* strains were grown in Luria–Bertani (LB) broth at 37 °C with shaking at 200 rpm. *E. coli* transformants were plated on LB agar plate containing ampicillin and incubated at 37 °C overnight.

ALE of *S. cerevisiae* strain CEN.PK113-7D

The haploid *S. cerevisiae* strain CEN.PK113-7D was used as the original parental strain for the ALE experiment. To determine the inhibitory PA concentration for evolution, an initial growth test of the parental strain in buffered minimal medium with 0–25 mM PA was conducted. For ALE, one single colony of CEN.PK 113-7D was inoculated into 5 mL YPD and grown overnight. The overnight culture was re-inoculated into 20 mL YPD at a starting OD_{600} of 0.5 and grown for another 4 h. The cells obtained from two rounds of pre-growth were used to inoculate flasks containing 10 mL of medium at an initial OD_{600} of 0.1. ALE was conducted in three different conditions, and three replicates were set in each condition. The evolution conditions were set as follows: minimal medium (pH 5); minimal medium with citric acid and Na_2HPO_4 buffered to pH 3.5; minimal medium with citric acid and Na_2HPO_4 buffered to pH 3.5 with PA at a starting concentration of 15 mM. The cultures were grown at 30 °C, 200 rpm until they reached the exponential phase before being transferred into fresh medium daily at an initial OD_{600} value of 0.1–0.2. Changes in cell density were recorded throughout the ALE. The concentration of PA was increased gradually when there was an improvement of growth and finally increased to 45 mM at the end of the evolution.

The number of generations through the evolution was estimated by adding up the $\log_2^{(\text{final OD}_{600}/\text{initial OD}_{600})}$ value of each transfer. Glycerol stocks of evolved populations were made at intervals throughout the evolution. Microscopy was conducted weekly to determine whether there was any contamination in the cell culture. At the end of the evolution, single clones were isolated by serial dilution from each of the evolution lineages and plating on YNB agar.

Fitness test under propionic acid stress

Aliquots (30 μL) of glycerol stocks of evolved populations kept at intervals were inoculated into 1 mL of minimal medium (pH 3.5) and incubated overnight with shaking before re-inoculation into 1 mL of buffered minimal medium with 15 mM PA at initial OD of 0.2 for another overnight culture. The second round of preculture was inoculated into 5 mL of buffered minimal medium with 25 mM and 35 mM of PA (pH 3.5). The experiments were repeated three times, and in each time, it was conducted in triplicates. The OD_{600} values were measured, and maximum specific growth rates were calculated.

The growth rate test of PA-evolved strains isolated at the end of the evolution was conducted with the same method above. Eight colonies were randomly picked from each of end-time PA-evolved cultures, and a total of five top performing isolates (PA-1, PA-2, PA-3, PA-4, and PA-5) were selected for further analysis.

Whole genome sequencing and variant calling

Twenty samples submitted for whole genome sequencing were as follows: the parental strain, five PA-evolved single isolates, the intermediate PA-evolved populations sampled at two time-points, and end-time populations evolved in minimal medium, in buffered minimal

medium, and in buffered minimal medium containing PA. The strains were grown overnight in YPD, and their genomic DNA was extracted with the Yeast DNA Extraction Kit (Thermo Fisher scientific, Cat No. 78870). Sequencing and library preparation were carried out by Macrogen Inc. using a True-Seq Nano kit with 470-bp inserts and paired-end Illumina HiSeq 2500 sequencing. According to the raw data report, all samples gave a quality score of Q30 above 90% and above 9 million reads and were, therefore, not trimmed. Paired end reads were analyzed using Geneious Pro 9.1.3 [66] by mapping to the CEN.PK 113-7D reference genome. This gave an average coverage of > 50×. Variant calling was performed at a minimum coverage of 10× and a minimum variant frequency of 0.05. Maximum variant p value was set at 10^{-5} and minimum strand-bias p value was set at 10^{-5} when exceeding 65% bias. Non-synonymous polymorphisms in CDS regions were identified. The mutations existing in the parental strain and in the control conditions were excluded from further analysis.

Reverse engineering

The parental strain CEN.PK 113-7D was used for reverse engineering. The plasmids and primers used in this study are listed in Additional file 1: Tables S1 and S2. *E. coli* DH5α was used for cloning using standard techniques. Yeast DNA extraction was carried out with the LiOAc–SDS method [67] or with the genomic DNA extraction kit mentioned above. Yeast transformation was performed using the LiAc/SS carrier DNA/PEG method [68].

Re-construction of *TRK1* mutations in the CEN.PK 113-7D parental strain and the PA-evolved isolate PA-3 was conducted via CRISPR–Cas9-mediated targeting and homologous recombination. First, each of the three *TRK1* mutations was reverse engineered into the parental strain, respectively. Partial open reading frames (ORFs) of *TRK1* variants containing each of the three mutated sequences were PCR amplified with TRK1 primers. A plasmid (Cas9-gRNA-423) containing the guide RNA expression cassette, the *Streptococcus pyogenes Cas9* gene [69], and the *hphMX* marker was used for CRISPR–Cas9-mediated genome modification. The three gRNAs for targeting different *TRK1* regions were designed by annotating all potential PAM sites and guide RNAs using Geneious Pro 9.1.3 [66]. Plasmid Cas9-gRNA-pRS423 [70] containing a previous gRNA was PCR amplified with each of the three pCRISPR-TRK1-gRNA primers that have 20 nucleotides (nt) new gRNA targeting DNA sequences as 5′ extensions. The primer extensions of CRISPR-TRK1-gRNA were complementary so that the linearized PCR product could be circularized to create Cas9-TRK1-crRNA-pRS423 using Yeast Assembly [71]. The partial *TRK1* fragments and linearized PCR product of Cas9-TRK1-crRNA-pRS423 were co-transformed into CEN.PK 113-7D, respectively, and plated onto YPD agar containing 300 μg/mL Hygromycin B. Second, partial ORFs of *TRK1* variants containing each of two other *TRK1* mutations were PCR amplified with TRK1-3223 and TRK1-3509 primers, respectively. The PCR products were used to reconstruct *TRK1* mutations in the PA-evolved isolate PA-3 via CRISPR–Cas9-mediated targeting and homologous recombination as the same method above. All the site-directed mutated strains were identified and confirmed by Sanger sequencing with TRK1 seq primers.

The *trk1Δ* strain was generated by replacing the ORF region of *TRK1* with the *hphMX* cassette. The *hphMX* cassette was amplified using Hph-TRK primers from pRS426-HphMX with overhangs homologous to *TRK1*. Upstream 500-bp and downstream 500-bp overhangs of *TRK1* were PCR amplified from gDNA with TRK1-upstream and downstream primers, which contained regions homologous to *hphMX*. The two overhangs have homologous sequences to both *TRK1* flanking regions and *hphMX* cassette, and were co-transformed with the Hph-TRK1 amplicon for deletion of *TRK1*. The *trk1Δ* mutant was confirmed by PCR with TRK1-Hph-knock-seq primers.

To enable overexpression of native *TRK1*, the ORF region of *TRK1* and the linearized pRS426 backbone were PCR amplified from yeast genomic DNA and pPGK1-CYC1t-pRS426, respectively, with overhangs to each other. The PCR products were then purified and ligated using Gibson Assembly [72]. The ligation products were transformed into *E. coli*, and the newly generated plasmids pPGK1-TRK1-CYC1t-pRS426 were extracted and characterized using TRK1-pRS426-check primers. The pPGK1-TRK1-CYC1t-HphMX cassette and linearized pRS413 backbone were amplified with corresponding primers and ligated through Gibson Assembly. After transforming into *E. coli*, the plasmids were extracted and characterized using ′pPGK1-TRK1-CYC1t-pRS413 check′ primers. Plasmid pPGK1-TRK1-CYC1t-pRS413 was transformed into the CEN.PK 113-7D parental strain for overexpression of *TRK1*.

The fitness tests of the reverse engineering strains were conducted with the same method as described above.

Sequence alignment and structure-based homology modelling of Trk1

The coordinates for *Bs*KtrB (PDB code 4J7C [45]) and *Vp*TrkH protein structures (PDB code 3PJZ [46]) were superimposed using the secondary-structure matching (SSM) algorithm in COOT [73]. The sequences for *Bs*KtrB (UniProt O32081) and *Vp*TrkH (UniProt Q87TN7) were then aligned manually to agree with the topological

alignment. The *ScTrk1*p sequence (UniProt P12685) was added manually to the structure-based sequence alignment and adjusted based on sequence similarity/identity and topological features of the aligned crystal structures as well as results from consensus transmembrane-helix assignment generated by TOPCONS [74]. A starting homology model was generated automatically with the SWISS-MODEL server (https://swissmodel.expasy.org), and further rebuilt manually in COOT to agree with our manually optimized structure-based sequence alignment, and to impose good backbone and side-chain torsion angle geometry, and relieve unfavorable interactions. Figure 4 is generated by displaying the manual alignment using ESPript 3.0 (http://espript.ibcp.fr) [75].

Potassium uptake measurement

To determine the effect of potassium concentrations and the function of *TRK1* on the tolerance to PA, the WT, *trk1Δ,* and three evolved isolates (PA-1, PA-3, and PA-4) were grown overnight in Translucent K^+ free YNB medium supplemented with 0.5 mM KCl. The suspension was washed twice, adjusted to an $OD_{600} = 0.2$ and serially diluted to 1:0, 1:10, 1:100, 1:1000, then spotted with 10 µL on YNB agar (pH 4.7) supplemented with increasing concentrations of potassium (0.1 mM, 1 mM, 10 mM, and 100 mM) with 0 mM, 15 mM (pH 4.0), and 25 mM PA (pH 4.0). To investigate the effect of potassium supplementation and *TRK1* on the tolerance to other organic acids, the WT, *trk1Δ,* and the evolved isolate PA-3 were grown and diluted as the same method above, then spotted on YNB agar containing 1 mM and 100 mM potassium, without additional acids and with 83 mM acetic acid (pH 3.2), 111 mM lactic acid (pH 2.5), 2 mM benzoic acid (pH 3.7), and 2 mM sorbic acid (pH 3.0). The plates were incubated for 3 days at 30 °C. Growth rate tests of WT, *trk1Δ* and PA-3 were also performed in liquid Translucent K^+-free YNB medium supplemented with increasing concentrations of potassium treated with 0 mM and 25 mM PA.

Additional file

Additional file 1: Fig. S1. PA effect to the growth of *S. cerevisiae*. **Fig S2.** The fluctuations of yeast growth through adaptive laboratory evolution. **Fig. S3.** Fitness test of *TRK1* mutants containing different combinations of two mutations in 35 mM PA. **Fig. S4.** Cartoon showing the overall fold of the *Sc*Trk1 channel. **Fig. S5.** The effect of potassium concentrations and *TRK1* on the tolerance of yeast strains to PA in liquid culture. **Table S1.** List of plasmids used in this study. **Table S2.** List of primers used in this study. **Table S3.** Genotypic changes in the PA evolved populations.

Abbreviations

PA: propionic acid; ALE: adaptive laboratory evolution; 3-HP: 3-hydroxypropionic acid; ABC: ATP-binding cassette; GPI: glycosylphosphatidylinositol; WT: wild type; OD_{600}: optical density at 600 nm; SD: standard deviation; SNPs: single-nucleotide polymorphism; CDS: coding sequence; ORF: open reading frame; *Sc*Trk1p: *Saccharomyces cerevisiae* Trk1p; *Bs*KtrB: *Bacillus subtilis* KtrB; *Vp*TrkH: *Vibrio parahaemolyticus* TrkH; IML: the intramembrane loop; MP: pore helix; SSM: secondary-structure matching; YNB: yeast nitrogen base; YPD: yeast peptone dextrose; LB: Luria–Bertani.

Authors' contributions

This study was designed by XX, TCW, and ITP. The experiments were conducted by XX. The data were analyzed by XX, TCW, and CD. The manuscript was written by XX, TCW, and CD, and reviewed and edited by XX, TCW, CD, ITP, and ISP. All authors read and approved the final manuscript.

Author details

[1] Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia. [2] CSIRO Synthetic Biology Future Science Platform, Canberra, ACT 2601, Australia. [3] KTH School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, 106 91 Stockholm, Sweden.

Acknowledgements

We would like to acknowledge Monica Espinosa and Dr. Niël Van Wyk for providing the Cas9-gRNA-pRS423-HphMX plasmid, Dr. Heinrich Kroukamp for providing the pPGK1-CYC1t-pRS426-HphMX plasmid and Dominic Logel for preliminary analysis of the Trk1p structure.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included in this published article (and its additional file).

Consent for publication

All authors have approved Biotechnology for Biofuels for publication.

Ethics approval and consent to participate

Not applicable.

Funding

The research was supported by the Synthetic Biology initiative at Macquarie University, which is funded by an internal grant from the University, and external grants from Bioplatforms Australia, the New South Wales (NSW) Chief Scientist and Engineer, and the NSW Government's Department of Primary Industries. XX is supported by a Co-funded China Scholarship Council-Macquarie University Research Excellence Scholarship (iCFCSCMQ). TCW is supported by the CSIRO Synthetic Biology Future Science Platform. ITP is supported by an Australian Research Council Laureate Fellowship.

Publisher's Note

Received: 22 November 2018 Accepted: 8 April 2019
Published online: 23 April 2019

References

1. Zidwick MJ, Chen J-S, Rogers P. Organic acid and solvent production: propionic and butyric acids and ethanol. The prokaryotes. Berlin: Springer; 2013. p. 135–67.
2. Wemmenhove E, van Valenberg HJ, Zwietering MH, van Hooijdonk TC, Wells-Bennik MH. Minimal inhibitory concentrations of undissociated lactic, acetic, citric and propionic acid for Listeria monocytogenes under conditions relevant to cheese. Food Microbiol. 2016;58:63–7.

3. Álvarez-Chávez CR, Edwards S, Moure-Eraso R, Geiser K. Sustainability of bio-based plastics: general comparative analysis and recommendations for improvement. J Cleaner Prod. 2012;23(1):47–56.
4. Hebert RF. Stable indole-3-propionate salts of *S*-adenosyl-L-methionine. Google Patents; 2017.
5. Vörös A, Horváth B, Hunyadkürti J, McDowell A, Barnard E, Patrick S, et al. Complete genome sequences of three *Propionibacterium acnes* isolates from the type IA2 cluster. J Bacteriol. 2012;194(6):1621–2.
6. Propionic Acid Market for Animal Feed & Grain Preservatives, Calcium & Sodium Propionates, Cellulose Acetate Propionate and Other Applications: global industry perspective, comprehensive analysis, size, share, growth, segment, trends and forecast, 2014–2020. https://www.marketresearchstore.com/report/propionic-acid-market-for-animal-feed-grain-z39993. Assessed 23 Aug 2018.
7. Woskow SA, Glatz BA. Propionic acid production by a propionic acid-tolerant strain of Propionibacterium acidipropionici in batch and semi-continuous fermentation. Appl Environ Microbiol. 1991;57(10):2821–8.
8. Swick RW, Wood HG. The role of transcarboxylation in propionic acid fermentation. Proc Natl Acad Sci. 1960;46(1):28–41.
9. Liu L, Zhu Y, Li J, Wang M, Lee P, Du G, et al. Microbial production of propionic acid from propionibacteria: current state, challenges and perspectives. Crit Rev Biotechnol. 2012;32(4):374–81.
10. Gonzalez-Garcia RA, McCubbin T, Navone L, Stowers C, Nielsen LK, Marcellin E. Microbial propionic acid production. Fermentation. 2017;3(2):21.
11. Kandasamy V, Vaidyanathan H, Djurdjevic I, Jayamani E, Ramachandran K, Buckel W, et al. Engineering *Escherichia coli* with acrylate pathway genes for propionic acid synthesis and its impact on mixed-acid fermentation. Appl Microbiol Biotechnol. 2013;97(3):1191–200.
12. Valli M, Sauer M, Branduardi P, Borth N, Porro D, Mattanovich D. Improvement of lactic acid production in *Saccharomyces cerevisiae* by cell sorting for high intracellular pH. Appl Environ Microbiol. 2006;72(8):5492–9.
13. Raab AM, Gebhardt G, Bolotina N, Weuster-Botz D, Lang C. Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. Metab Eng. 2010;12(6):518–25.
14. Borodina I, Kildegaard KR, Jensen NB, Blicher TH, Maury J, Sherstyk S, et al. Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in *Saccharomyces cerevisiae* via β-alanine. Metab Eng. 2015;27:57–64.
15. Curran KA, Leavitt JM, Karim AS, Alper HS. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab Eng. 2013;15:55–66.
16. Eglinton JM, Heinrich AJ, Pollnitz AP, Langridge P, Henschke PA, de Barros Lopes M. Decreasing acetic acid accumulation by a glycerol overproducing strain of *Saccharomyces cerevisiae* by deleting the ALD6 aldehyde dehydrogenase gene. Yeast. 2002;19(4):295–301.
17. Fernandes A, Mira N, Vargas R, Canelhas I, Sá-Correia I. *Saccharomyces cerevisiae* adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. Biochem Biophys Res Commun. 2005;337(1):95–103.
18. Piper P, Mahe Y, Thompson S, Pandjaitan R, Holyoak C, Egner R, et al. The Pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast. EMBO J. 1998;17(15):4257–65.
19. Stratford M, Anslow P. Comparison of the inhibitory action on *Saccharomyces cerevisiae* of weak-acid preservatives, uncouplers, and medium-chain fatty acids. FEMS Microbiol Lett. 1996;142(1):53–8.
20. Mira NP, Teixeira MC, Sá-Correia I. Adaptive response and tolerance to weak acids in *Saccharomyces cerevisiae*: a genome-wide view. OMICS. 2010;14(5):525–40.
21. Carmelo V, Santos H, Sa-Correia I. Effect of extracellular acidification on the activity of plasma membrane ATPase and on the cytosolic and vacuolar pH of *Saccharomyces cerevisiae*. Biochem Biophys Acta. 1997;1325(1):63–70.
22. Fernandes A, Durao P, Santos P, Sá-Correia I. Activation and significance of vacuolar H+-ATPase in *Saccharomyces cerevisiae* adaptation and resistance to the herbicide 2,4-dichlorophenoxyacetic acid. Biochem Biophys Res Commun. 2003;312(4):1317–24.
23. Holyoak C, Stratford M, McMullin Z, Cole M, Crimmins K, Brown A, et al. Activity of the plasma membrane H (+)-ATPase and optimal glycolytic flux are required for rapid adaptation and growth of *Saccharomyces cerevisiae* in the presence of the weak-acid preservative sorbic acid. Appl Environ Microbiol. 1996;62(9):3158–64.
24. Tanaka K, Ishii Y, Ogawa J, Shima J. Enhancement of acetic acid tolerance in *Saccharomyces cerevisiae* by overexpression of the HAA1 gene, encoding a transcriptional activator. Appl Environ Microbiol. 2012;78(22):8161–3.
25. Alenquer M, Tenreiro S, Sá-Correia I. Adaptive response to the antimalarial drug artesunate in yeast involves Pdr1p/Pdr3p-mediated transcriptional activation of the resistance determinants TPO1 and PDR5. FEMS Yeast Res. 2006;6(8):1130–9.
26. Teixeira MC, Sá-Correia I. Saccharomyces cerevisiae resistance to chlorinated phenoxyacetic acid herbicides involves Pdr1p-mediated transcriptional activation of TPO1 and PDR5 genes. Biochem Biophys Res Commun. 2002;292(2):530–7.
27. Simoes T, Teixeira M, Fernandes A, Sá-Correia I. Adaptation of *Saccharomyces cerevisiae* to the herbicide 2, 4-dichlorophenoxyacetic acid, mediated by Msn2p-and Msn4p-regulated genes: important role of SPI1. Appl Environ Microbiol. 2003;69(7):4019–28.
28. Simoes T, Mira N, Fernandes A, Sá-Correia I. The SPI1 gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. Appl Environ Microbiol. 2006;72(11):7168–75.
29. Mira NP, Lourenço AB, Fernandes AR, Becker JD, Sa-Correia I. The RIM101 pathway has a role in *Saccharomyces cerevisiae* adaptive response and resistance to propionic acid and other weak acids. FEMS Yeast Res. 2009;9(2):202–16.
30. Lourenço AB, Ascenso JR, Sá-Correia I. Metabolic insights into the yeast response to propionic acid based on high resolution 1 H NMR spectroscopy. Metabolomics. 2011;7(4):457–68.
31. Dragosits M, Mattanovich D. Adaptive laboratory evolution—principles and applications for biotechnology. Microb Cell Fact. 2013;12(1):64.
32. Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature. 2002;420(6912):186.
33. Reyes LH, Gomez JM, Kao KC. Improving carotenoids production in yeast via adaptive laboratory evolution. Metab Eng. 2014;21:26–33.
34. Atsumi S, Wu TY, Machado IM, Huang WC, Chen PY, Pellegrini M, et al. Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. Mol Syst Biol. 2010;6(1):449.
35. Lee JY, Yang KS, Jang SA, Sung BH, Kim SC. Engineering butanol-tolerance in escherichia coli with artificial transcription factor libraries. Biotechnol Bioeng. 2011;108(4):742–9.
36. Fischer CR, Tseng H-C, Tai M, Prather KL, Stephanopoulos G. Assessment of heterologous butyrate and butanol pathway activity by measurement of intracellular pathway intermediates in recombinant *Escherichia coli*. Appl Microbiol Biotechnol. 2010;88(1):265–75.
37. Portnoy VA, Bezdan D, Zengler K. Adaptive laboratory evolution—harnessing the power of biology for metabolic engineering. Curr Opin Biotechnol. 2011;22(4):590–4.
38. Brennan TC, Williams TC, Schulz BL, Palfreyman RW, Krömer JO, Nielsen LK. Evolutionary engineering improves tolerance for replacement jet fuels in *Saccharomyces cerevisiae*. Appl Environ Microbiol. 2015;81(10):3316–25.
39. Williams TC, Pretorius IS, Paulsen IT. Synthetic evolution of metabolic productivity using biosensors. Trends Biotechnol. 2016;34(5):371–81.
40. González-Ramos D, de Vries ARG, Grijseels SS, Berkum MC, Swinnen S, Broek M, et al. A new laboratory evolution approach to select for constitutive acetic acid tolerance in *Saccharomyces cerevisiae* and identification of causal mutations. Biotechnol Biofuels. 2016;9(1):173.
41. Kildegaard KR, Hallström BM, Blicher TH, Sonnenschein N, Jensen NB, Sherstyk S, et al. Evolution reveals a glutathione-dependent mechanism of 3-hydroxypropionic acid tolerance. Metab Eng. 2014;26:57–66.
42. Fletcher E, Feizi A, Bisschops MM, Hallström BM, Khoomrung S, Siewers V, et al. Evolutionary engineering reveals divergent paths when yeast is adapted to different acidic environments. Metab Eng. 2017;39:19–28.
43. Gaber RF, Styles CA, Fink GR. TRK1 encodes a plasma membrane protein required for high-affinity potassium transport in *Saccharomyces cerevisiae*. Mol Cell Biol. 1988;8(7):2848–59.
44. Zayats V, Stockner T, Pandey SK, Woerz K, Ettrich R, Ludwig J. A refined atomic scale model of the *Saccharomyces cerevisiae* K+-translocation protein Trk1p combined with experimental evidence confirms the role of selectivity filter glycines and other key residues. Biochimica et Biophysica Acta (BBA). 2015;1848(5):1183–95.

45. Vieira-Pires RS, Szollosi A, Morais-Cabral JH. The structure of the KtrAB potassium transporter. Nature. 2013;496(7445):323.
46. Cao Y, Jin X, Huang H, Derebe MG, Levin EJ, Kabaleeswaran V, et al. Crystal structure of a potassium ion transporter, TrkH. Nature. 2011;471(7338):336.
47. Kuroda T, Bihler H, Bashi E, Slayman C, Rivetta A. Chloride channel function in the yeast TRK-potassium transporters. J Membr Biol. 2004;198(3):177–92.
48. Rivetta A, Kuroda T, Slayman C. Anion currents in yeast K+ transporters (TRK) characterize a structural homologue of ligand-gated ion channels. Pflügers Arch-Eur J Physiol. 2011;462(2):315–30.
49. Wright J, Bellissimi E, de Hulster E, Wagner A, Pronk JT, van Maris AJ. Batch and continuous culture-based selection strategies for acetic acid tolerance in xylose-fermenting *Saccharomyces cerevisiae*. FEMS Yeast Res. 2011;11(3):299–306.
50. Landstetter N, Glaser W, Gregori C, Seipelt J, Kuchler K. Functional genomics of drug-induced ion homeostasis identifies a novel regulatory crosstalk of iron and zinc regulons in yeast. OMICS. 2010;14(6):651–63.
51. Hollenhorst PC, Bose ME, Mielke MR, Müller U, Fox CA. Forkhead genes in transcriptional silencing, cell morphology and the cell cycle: overlapping and distinct functions for FKH1 and FKH2 in *Saccharomyces cerevisiae*. Genetics. 2000;154(4):1533–48.
52. Postnikoff SD, Malo ME, Wong B, Harkness TA. The yeast forkhead transcription factors fkh1 and fkh2 regulate lifespan and stress response together with the anaphase-promoting complex. PLoS Genet. 2012;8(3):e1002583.
53. Mira NP, Henriques SF, Keller G, Teixeira MC, Matos RG, Arraiano CM, et al. Identification of a DNA-binding site for the transcription factor Haa1, required for *Saccharomyces cerevisiae* response to acetic acid stress. Nucleic Acids Res. 2011;39(16):6896–907.
54. Mira NP, Palma M, Guerreiro JF, Sá-Correia I. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. Microb Cell Fact. 2010;9(1):79.
55. Macpherson N, Shabala L, Rooney H, Jarman MG, Davies JM. Plasma membrane H+ and K+ transporters are involved in the weak-acid preservative response of disparate food spoilage yeasts. Microbiology. 2005;151(6):1995–2003.
56. Mollapour M, Fong D, Balakrishnan K, Harris N, Thompson S, Schüller C, et al. Screening the yeast deletant mutant collection for hypersensitivity and hyper-resistance to sorbate, a weak organic acid food preservative. Yeast. 2004;21(11):927–46.
57. Nijkamp JF, van den Broek M, Datema E, de Kok S, Bosman L, Luttik MA, et al. De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. Microb Cell Fact. 2012;11(1):36.
58. Jönsson LJ, Alriksson B, Nilvebrant N-O. Bioconversion of lignocellulose: inhibitors and detoxification. Biotechnol Biofuels. 2013;6(1):16.
59. Yenush L, Mulet JM, Ariño J, Serrano R. The Ppz protein phosphatases are key regulators of K+ and pH homeostasis: implications for salt tolerance, cell wall integrity and cell cycle progression. EMBO J. 2002;21(5):920–9.
60. Ariño J, Ramos J, Sychrová H. Alkali metal cation transport and homeostasis in yeasts. Microbiol Mol Biol Rev. 2010;74(1):95–120.
61. Kahm M, Navarrete C, Llopis-Torregrosa V, Herrera R, Barreto L, Yenush L, et al. Potassium starvation in yeast: mechanisms of homeostasis revealed by mathematical modeling. PLoS Comput Biol. 2012;8(6):e1002548.
62. Aiking H, Sterkenburg A, Tempest D. Influence of specific growth limitation and dilution rate on the phosphorylation efficiency and cytochrome content of mitochondria of *Candida utilis* NCYC 321. Arch Microbiol. 1977;113(1–2):65–72.
63. Merchan S, Pedelini L, Hueso G, Calzada A, Serrano R, Yenush L. Genetic alterations leading to increases in internal potassium concentrations are detrimental for DNA integrity in *Saccharomyces cerevisiae*. Genes Cells. 2011;16(2):152–65.
64. Lam FH, Ghaderi A, Fink GR, Stephanopoulos G. Engineering alcohol tolerance in yeast. Science. 2014;346(6205):71–5.
65. Zimmermannova O, Salazar A, Sychrova H, Ramos J. Zygosaccharomyces rouxii Trk1 is an efficient potassium transporter providing yeast cells with high lithium tolerance. FEMS Yeast Res. 2015;15(4):fov029.
66. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
67. Lõoke M, Kristjuhan K, Kristjuhan A. Extraction of genomic DNA from yeasts for PCR-based applications. Biotechniques. 2011;50(5):325.
68. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2007;2(1):31.
69. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR–Cas systems. Nucleic Acids Res. 2013;41(7):4336–43.
70. Williams TC, Xu X, Ostrowski M, Pretorius IS, Paulsen IT. Positive-feedback, ratiometric biosensor expression improves high-throughput metabolite-producer screening efficiency in yeast. Synth Biol. 2017;2(1):ysw002.
71. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Proc Natl Acad Sci. 2008;105(51):20404–9.
72. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA III, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343.
73. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr. 2004;60(12):2126–32.
74. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res. 2015;43(W1):W401–7.
75. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014;42(W1):W320–4.

Chapter 6 : General discussion and future directions

6.1 General discussion

Metabolic engineering has enabled sustainable bio-production of a wide range of molecules such as chemicals, fuels and pharmaceuticals. However, strain engineering is usually a process that costs much time and effort. The rapid development of synthetic biology has enabled novel techniques that have the potential to reduce the time and cost of strain development, and provide solutions that are not readily available in natural systems. For example, synthetic genomics can be used to introduce novel forms of genetic variation, and biosensor-mediated high-throughput screening can be used to isolate productive strains from genetically diverse populations. It is therefore of great value to develop and apply these novel synthetic biology tools to revolutionise the time-consuming strain engineering process.

PA is one of the top 30 candidate building block chemicals (1), widely used as a food preservative and a chemical intermediate in many industries (2-5). The global production of PA is estimated to be ~450,000 tonnes annually with a 2.7% annual growth, and is predicted to reach a revenue of USD 1.55 billion by 2020 (6). However, PA is still derived from finite oil reserves in an unsustainable manner. Microbial PA production through engineering the cell factory yeast is a sustainable alternative strategy. The overall aim of the project is to establish novel tools for the generation and selection of genetic diversity to improve PA production and PA tolerance. To achieve this purpose, I contributed to the construction and the debugging of the synthetic yeast genome (Sc2.0 synXIV, Chapters 2) and the development of an organic acid biosensor for high-throughput screening of PA producing yeast cells (Chapter 3). With the synthetic genome platform and the PA biosensor, a superior strategy for SCRaMbLE and biosensor facilitated FACS has been developed and successfully used to isolate two strains with improved PA titres. However, the genes of the synthetic Wood-Werkman cycle except *scpC* were lost in the PA productive strains, and the improved production of PA unintentionally was likely produced from native yeast amino acid catabolic pathways coupled with the expression of *scpC* (Chapter 4). In addition, the PA tolerance in yeast was improved by more than 3-fold

through ALE and the underlying tolerance mechanism has been identified (Chapter 5). These approaches revealed novel principles that will shed light on future microbial organic acid engineering.

In Chapter 2, I contributed to the synthesis of the synthetic *S. cerevisiae* chromosome XIV. The full 753,097 bp synXIV has been completed and shown to function in yeast. Throughout the construction process, unexpected mutations occurred possibly during consecutive rounds of transformation, and whole genome sequenceing was used throughout the construction to reveal genome discrepancies with the designed sequences. The genetic change made *in silico* also resulted in growth defect, which is in accordance with previous studies (7, 8). Pooled PCRTag mapping (PoPM) (7) and whole genome sequencing revealed a major growth defect resulted from a loxPsym site insertion in the 3' UTR of *MRPL19*, which interfered with the import of this mitochondria ribosomal protein. In the debugging of SynX, a loxPsym site insertion in the 3' UTR also led to a major growth defect by disrupting the promoter region of *ATP2*, encoding a subunit of the mitochondrial F1F0 ATP synthase (7). LoxPsym site insertions facilitate whole genome rearrangements, however, the insertions are also likely to interfere with the function of adjacent genes. Great care needs to be employed in the future to the debugging processes. The fitness of synXIV was further improved by ALE, which resulted in an increased copy number of the rDNA locus on chromosome XII and expression of the *TAR1* protein in response to defective mitochondria. ALE was applied for the first time in debugging, and proven to be an effective approach to improve the fitness of the semi-synthetic strain, as well as uncovering causal genetics. After debugging and corrections, the phenotype and fitness of the final synXIV strain was comparable with the wild type. Our study has demonstrated again that design, build, test and learn cycles were necessary to construct a semi-synthetic strain with high fitness. Frequent whole genome sequencing will become essential for the confirmation of sequences and debugging in large scale synthetic genome projects.

260 loxPsym sites have been implemented in synXIV. These sites will facilitate the generation of novel genome arrangements upon the induction of SCRaMbLE. As a genome evolution

system, SCRaMbLE can serve as a platform for screening of industrially favourable phenotypes. To apply SCRaMbLE for the screening of PA producers, I contributed to the development of an organic acid biosensor responsive to PHBA and PA in Chapter 3. The biosensor relies on the native weak acid responsive transcription factor War1p and its target promoter *PDR12*, which can be used to transduce concentrations of PHBA and PA into a GFP fluorescence signal. Although the transcription factor based biosensor was responsive to PHBA and PA, it had a high noise level and a limited dynamic range. A positive feedback loop was then engineered by controlling *WAR1* expression with its target *PDR12* promoter to increase the dynamic range of the biosensor. A red fluorescent mCherry protein was also expressed constitutively to normalise the difference in GFP expression within each individual cell in a population to control for intrinsic noise. The positive feedback ratiometric biosensor has demonstrated novel design principles that are likely applicable for the improvement of other transcription-factor based biosensors. Combined with FACS, the positive feedback ratiometric biosensor enabled the accurate sorting of PHBA-producing yeast cells that were mixed in a population with PHBA non-producers at a ratio of 1:10,000. Thus, this biosensor is very promising for application in high-throughput screening of yeast cells that produce the highest level of PHBA or PA throughout a mutant library, which will reveal novel engineering principles for organic acid production.

In Chapter 4, the potential of SCRaMbLE has been explored in yeast metabolic engineering. Specifically, SCRaMbLE of the synthetic Wood-Werkman cycle was induced in a haploid synthetic strain (with syn III, syn VI and syn IXR) and a synthetic-wild-type hybrid diploid to improve the heterologous production of PA. Our results showed SCRaMbLE of a synthetic-wild-type hybrid potentially generated more genome diversity with desired phenotypes than SCRaMbLE of a synthetic haploid, which is consistent with previous reports of reduced SCRaMbLE-mediated lethality in synthetic-wild-type polyploid hybrids (9, 10). The biosensor we previously established was applied for high-throughput screening of PA productive SCRaMbLEd cells. Although the biosensor is sensitive to extracellular supplementation of

different concentrations of PA, it is difficult to sort out the SCRaMbLEd cells producing high PA titers due to its extremely low abundance and the rapid evolution of cheater cells. High fluorescent population sorting was confounded by the outgrowth of cheater cells. Single cell sorting after one round of SCRaMbLE was an effective SCRaMbLE-sorting procedure, which sorted out producers with 1.7-to 2.4-fold improved titres. However, through sequencing of the non-SCRaMbLEd strain and PA producers, we found quite surprising results. The genes in the synthetic Wood-Werkman cycle except *scpC* were looped-out in the starting strain without the induction of SCRaMbLE. This was probably due to the leakiness of Cre expression and activity from the pSCW11-Cre-EBD expression system. In addition, there is no energetic or selective advantage from PA production via the synthetic Wood-Werkman cycle. Therefore, the burden of heterologous expression may have favoured gene loss. In the future the synthetic Wood-Werkman cycle genes could be ratained if the pathway were designed to supply the electron transport chain. The three synthetic chromosomes were not SCRaMbLEd either since they were all lost in the two producer strains. Synthetic chromosome loss is a major issue that needs to be considered when SCRaMbLE-ing heterozygous diploids, which might occur randomly or for improved growth. In future studies, PCRtag analysis should be used to confirm the heterozygosity after outgrowth and screening of SCRaMbLEd diploids. The higher PA yield might have resulted from the degradation of amino acids via amino acid catabolic pathways such as the threonine, methionine, isoleucine and valine pathways. The propionyl-CoA generated from amino acid degradation could then be converted to PA by the remaining *scpC* gene. Our study is the first attempt to apply biosensors in the synthetic genome platform to improve a heterologous pathway. Due to the evolution of cheater cells, a cheater elimination mechanism would be required to facilitate the screening of real PA producers in future applications of biosensor based FACS. This could be, for example, utilising inducible expression of the pathway with alternating rounds of positive and negative selection (11). Despite the undesired results, the study contributed to the establishment of the methodology of

SCRaMbLE. In the future, a precise control system needs to be used to reduce the leakiness of the SCRaMbLE system especially when used to diversify a complex biosynthesis pathway.

PA toxicity to yeast is a major limitation in terms of its bio-production. In chapter 5, ALE was conducted to improve PA tolerance in yeast. In this study, a key element for the success of ALE is to increase PA concentration gradually. At the start, a low PA concentration allowed a sufficient number of cell replications and the generation of beneficial mutations. The increase of PA concentration enables the variants containing causal mutations to take over the population. Strains with more than 3-fold improved tolerance to PA were isolated after more than 260 generations of ALE. Non-synonymous mutations in *TRK1,* which encodes a high-affinity potassium transporter, were confirmed to be the cause of the increased PA tolerance using whole-genome re-sequencing and CRISPR-cas9 mediated SNP reconstruction in the parental strain. Based on the sequence alignment against *BsKtrB* and *VpTrkH,* the three mutations likely increase oligomer stabilization via interactions with the C-tail, and affect transporter function via interactions that affect the pore gate. However, given that the homology between *ScTrk1*, *BsKtrB*, and *VpTrkH* is relatively low, the availability of an experimentally guided *ScTrk1p* structure would facilitate further understanding of the functions of these mutations. Potassium uptake assays showed that extracellular potassium supplementation as well as *TRK1* mutated alleles not only increased PA tolerance but also conferred tolerance to multiple organic acids. This tolerance mechanism can therefore be applied to improve future organic acid and lignocellulosic ethanol production processes in yeast. Given that potassium transport had also been reported to be relevant to ethanol tolerance and high lithium tolerance (12, 13), potassium concentrations and engineering of Trk1p may also play an important role in other stress conditions. Although PA response to yeast had been studied through genome-wide screening of the yeast knock-out library (14), our study identified novel genes, and improved prolonged PA tolerance in yeast for the first time. This study has demonstrated once again that ALE combined with ‘omics’ techniques is a powerful method for obtaining and understanding industrially relevant stress tolerance phenotypes.

6.2 Future directions

6.2.1 A more precise control of SCRaMbLE

In this thesis, loop-out of pathway genes in the synthetic Wood-Werkman cycle have been observed before the induction of SCRaMbLE, which possibly results from the leakiness of the Cre-EBD expression system used to control SCRaMbLE. This assumption is supported by several studies that growth defects have been observed in the synthetic strains with integration of pSCW11-Cre-EBD even without induction of SCRaMbLE (15-17). Hochrein et al. also observed high basal expression without induction when using pSCW11-Cre-EBD. In the future, a control system that enables tighter regulation of SCRaMbLE should be applied. A light induced Cre-recombinase has been used to control the SCRaMbLE system (L-SCRaMbLE) (18). Using L-SCRaMbLE, the extent of recombination can be controlled by adjusting the induction time and concentration of the chromophore phycocyanobilin. Compared with the β-estradiol induced pSCW11-Cre-EBD system, L-SCRaMbLE generated much less basal activity and more diverse recombination events. A genetic AND gate switch has also been developed by using a galactose-inducible promoter *pGAL1* to control Cre (9). Thus, the SCRaMbLE system could only be induced by adding galactose and estradiol simultaneously. In comparison with *pSCW11,* the *pGAL1* induced system facilitated more Cre-EBD protein localization in the nucleus and exhibited a higher expression level upon induction. In addition, it also showed much less leaky expression. Both of these systems could be used to replace the current pSCW11-Cre-EBD for a more precise control of SCRaMbLE.

6.2.2 A reporter system as an indicator of real SCRaMbLEd events

Because of the effectively infinite number of different SCRaMbLE gene rearrangements, a SCRaMbLEd genome with the target phenotype is always extremely rare. To make things worse, there are still a considerable number of non-SCRaMbLEd cells within the post-SCRaMbLE population, which might result from survival and outgrowth of mother cells, or mutations in the Cre gene (19). The non-SCRaMbLEd cells can potentially interfere with the

identification of SCRaMbLEd genomes of interest if they are too prevalent in the post-SCRaMbLE population. To tackle this problem, a reporter system can be implemented to exclude the non-SCRaMbLEd cells and ensure the presence of genome rearrangements. For example, a marker gene flanked by loxPsym sites could be integrated into the synthetic genome. After SCRaMbLE, by selecting for loss of the marker, we can enrich the proportion of cells with rearranged genomes, which facilitates the selection of genomes of interest. A system that enables the identification of SCRaMbLEd cells, termed 'ReSCuES' was recently developed to identify SCRaMbLEd synXII strains (19). The system consists of '*URA3*-*LEU2*' cassette in a convergent direction, flanked by two inward loxP sequences. The markers will be inverted upon the induction of SCRaMbLE. Through the counter selection of *URA3* and *LEU2*, ReSCuES had greatly enriched the number of SCRaMbLEd genomes with a phenotype of interest. In addition to auxotropic markers or antibiotic resistance markers, fluorescent proteins could also be used for generating a reporter system. Combined with FACS, the fluorescent reporter system will enable high-throughput screening of SCRaMbLEd genomes.

6.2.3 Alternative pathways for SCRaMbLE of PA production in yeast

Although the Wood-Werkman cycle is a superior pathway based on calculations of theoretical PA and energy yields (20), it involves the heterologous expression of nine pathway genes in yeast, which causes a burden for the protein expression machinery and possible metabolism. In addition, maintenance of all the pathway genes during SCRaMbLE is problematic, and the supplementation of vitamin B12 for the function of methyl-malonyl-CoA mutase also makes the route less economically attractive. It is therefore worth considering alternate metabolic pathways for the production of PA in yeast.

Aside from the Wood-Werkman cycle, several metabolic pathways have also been reported to facilitate PA production, for example, 1, 2-propanediol associated pathways (21, 22), amino acid catabolic pathways, the acrylate pathway and the 3HP cycle (23, 24).

1, 2-propanediol associated pathways have been reported in *Salmonella typhimurium* (22) and *Roseburia inulinivorans* (25), in which deoxy sugars or the glycolytic intermediate glycerone phosphate can be catabolised into 1, 2-propanediol and finally into propionate. The pathways have a high theoretical PA yield. However, deoxy sugars are not favourable substrate for yeast catabolism (26), and diol dehydratase, which converts 1,2-propodiol to propionaldehyde, is also vitamin B12-dependant (27). These characteristics make these pathways less attractive for engineering in yeast. Amino acid catabolic pathways were not attractive either due to the cost of supplying amino acids to an industrial-scale fermentation. The acrylate pathway has been reported in *Clostridium propionicum, Megasphaera elsdenii* and *Prevotella ruminicol* (20). The acrylate pathway is attractive due to the ATP neutral conversion from lactate to propionate (20) and no requirement for vitamin B12-dependant enzymes. A synthetic acrylate pathway involves seven genes, encoding propionate CoA-transferase, lactoyl-CoA dehydratase and acryloyl-CoA reductase, has been engineered in *E. coli* (28). The resulting strain produced 3.7 mM PA through fermentation of glucose. The 3HP cycle is another attractive route identified in *Chloroflexacae* that enables carbon dioxide fixation (23, 24). The cycle contains acetyl-CoA carboxylase, malonyl-CoA reductase, and propionyl-CoA synthase, which converts acetyl-CoA into malonyl-CoA, 3HP, and propionyl-CoA, respectively (24, 29). *ACC1* encodes acetyl-CoA carboxylase in yeast (30). Heterologous expression of malonyl-CoA reductase (mcr_{Ca}) (31) and propionyl-coa synthase (*pcs*) (29) from *Chloroflexus aurantiacus* and propionyl-CoA succinate-CoA transferase (*scpC*) from *E. coli* will theoretically enable PA production in yeast.

The 3HP cycle has several advantages to be engineered in yeast: firstly, it contains no vitamin B12 dependent enzyme; secondly, only expression of three heterologous genes are necessary for PA production; lastly, the 3HP cycle has already been optimised in yeast for producing 3HP (32), which is the precursor for PA production. Chen et al. reported over-expression of *ACC1*, *ADH2* (encoding alcohol dehydrogenase), *ALD6* (encoding NADP-dependent aldehyde dehydrogenase) and acs_{SE}^{L641P} (encoding an acetylation-insensitive acetyl-CoA synthetase from *Salmonella enterica*) improved 3HP yield by increasing the malonyl-CoA supply. In

addition, overexpression of *gapN* (encoding non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans*) also contributed to 3HP yield by complementing NADPH supply (32). These genes can be overexpressed or SCRaMbLEd to further increase the PA yield. Thus, a synthetic 3HP cycle has been designed by flanking the three 3HP cycle genes (*mcr*$_{Ca}$, *pcs* and *scpC*) with delta sites that are homologous to the hundreds of retrotransposons repeats in the yeast genome, and inside the cassette, each of the genes was flanked by loxPsym sites. The genes encoding the precursors (*ACC1*, *ADH2*, *ALD6*, *acs*$_{SE}^{L641P}$) and NADPH (*gapN*) were also flanked by delta sites and loxPsym sites to further improve the PA production. The delta sites could facilitate integration of the genes into the yeast genome at multiple loci using the high-efficiency DiCRISPR system (33). After integration of the genes, diverse rearranged genomes could be generated through the induction of the SCRaMbLE system. Given the fact that SCRaMbLE of the synthetic Wood-Werkman cycle didn't generate genome rearrangements for PA production, the synthetic 3HP cycle could be a promising alternative route for expression and SCRaMbLE. Combined with the PA biosensor-facilitated FACS and nanopore sequencing, PA-productive synthetic genomes will be sorted out, and the underlying mechanism that leads to improved production could be identified.

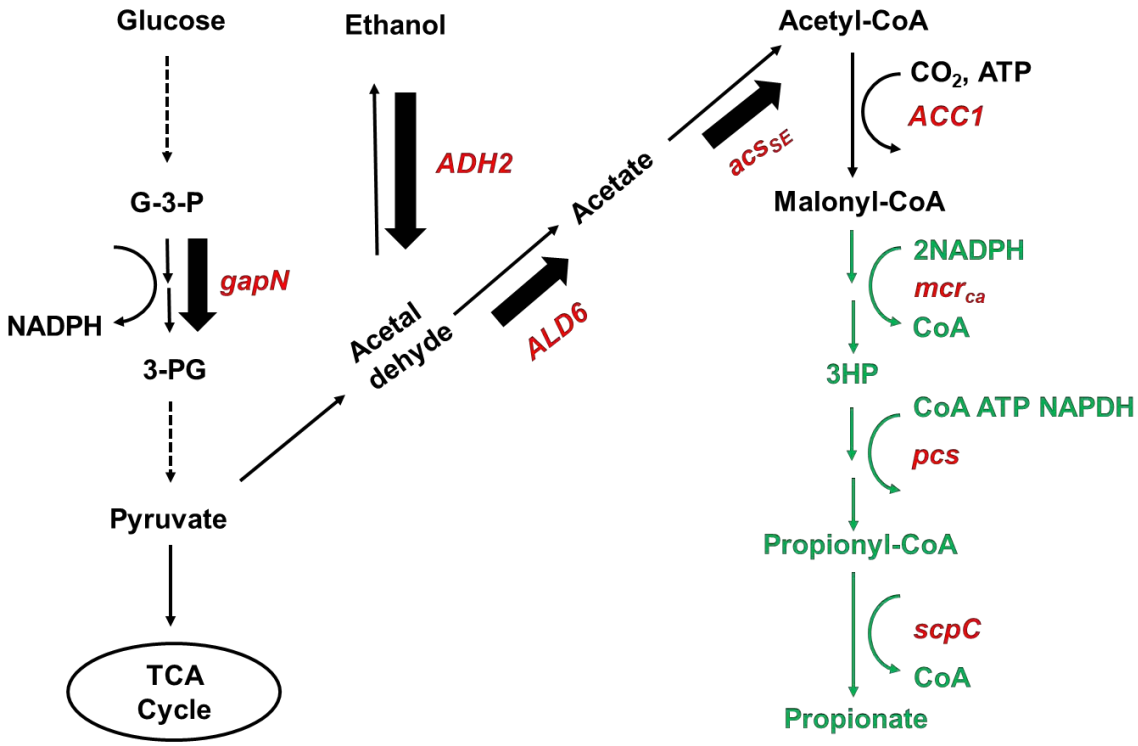


Figure 6-1. Redesigned synthetic 3HP pathway for PA production in *S. cerevisiae*. The native reactions and intermediates are shown in black, while the heterologous reactions and intermediates are shown in green. The reactions that could be further overexpressed or SCRaMbLEd to increase the precursor supply are shown using thick black arrows. The enzymes for engineering the pathway are marked in red. G-3-P: glyceraldehyde-3-phosphate; 3-PG: 3-phospho-D-glycerate.

6.2.4 Engineering PA tolerance in PA-productive yeast strains

Engineering improved product tolerance has been reported to increase the product yields in several studies: Lam et al. reported elevated potassium and pH generated yeast tolerant to multiple alcohols, and this increased tolerance also boosted ethanol production by enhancing the population viability (12); Mutations in the transcription factor Spt15p increased yeast tolerance to ethanol and improved conversion efficiency from glucose to ethanol, which led to an improved ethanol yield (34); Genome shuffling improved acid tolerance in *Lactobacillus rhamnosus*, and the tolerant strains had an increased volumetric productivity of L-lactic acid (35).

The present study has demonstrated that yeast tolerance to PA can be improved simply by engineering the beneficial single-nucleotide mutations in *TRK1* or extracellular supplementation of potassium in the medium. However, the influence of increased PA tolerance

on yeast PA yield hasn't been explored yet. The enhanced tolerance possibly promotes the export of protons from the cytoplasm and stabilises the membrane potential. This likely enables the maintenance of pH homeostatsis and therefore population viability, which is very likely to boost the yield. Thus far, the PA yield we achieved in yeast was only 1.32 mM, which is far below the MIC to influence growth. When a better PA producer is engineered in the future, the significance of tolerance engineering in PA production could be tested by engineering *TRK1* with the beneficial mutations.

6.2.5 Application of SCRaMbLE for industrially favourable tolerance phenotypes

No more than three synthetic chromosomes have ever been SCRaMbLE'd simultaneously in a single strain, including in this thesis. As more synthetic chromosomes are completed, they are consolidated into a single strain containing all sixteen synthetic chromosomes. The synIII, synVI and synXIR used in our study contains more than 200 loxPsym sites, while the ultimate synthetic strain will have more almost 4000 loxPsym sites with much greater potential to generate more diverse genomes (36). Inevitably, a larger number of loxPsym sites will also result in more genome instability and increase lethal events during SCRaMbLE. Thus, it is necessary to apply a tight regulated Cre-recombinase for the precise control of SCRaMbLE system. SCRaMbLE of strains containing more synthetic chromosomes could then be used to generate organic acid tolerance, as well as other industrially relevant tolerance phenotypes, such as ethanol, sugar, heat, cold, salt, and oxidative stress tolerance. In addition, a reporter system could also be applied to isolate strains with genome rearrangements and to exclude those that have only physiological (non-genetic) changes. Combined with omics analysis of superior strains and reverse engineering, SCRaMbLE will reveal novel genome design principles for a multitude of yeast phenotypes that are of interest for both basic and applied research.

Overall, this thesis has contributed to the construction of the synthetic yeast genome (Sc2.0) and the utilisation of the SCRaMbLE genome evolution system in metabolic engineering. Particularly, novel tools have been established for the generation and high-throughput selection

of genetic diversity to improve PA production and PA tolerance, which has provided a generalised strategy for industrial strain improvement with synthetic genomics approaches. With much more loxPsym sites implemented in the ultimate synthetic yeast genome, SCRaMbLE will be a powerful platform to provide fast, inexpensive, and non-intuitive solutions for current problematic biological engineering issues.

6.3 Reference

1. Vörös A, Horváth B, Hunyadkürti J, McDowell A, Barnard E, Patrick S, et al. Complete genome sequences of three Propionibacterium acnes isolates from the type IA2 cluster. Journal of bacteriology. 2012;194(6):1621-2.

2. Álvarez-Chávez CR, Edwards S, Moure-Eraso R, Geiser K. Sustainability of bio-based plastics: general comparative analysis and recommendations for improvement. Journal of Cleaner Production. 2012;23(1):47-56.

3. Hebert RF. Stable indole-3-propionate salts of S-adenosyl-L-methionine. Google Patents; 2017.

4. Wemmenhove E, van Valenberg HJ, Zwietering MH, van Hooijdonk TC, Wells-Bennik MH. Minimal inhibitory concentrations of undissociated lactic, acetic, citric and propionic acid for Listeria monocytogenes under conditions relevant to cheese. Food microbiology. 2016;58:63-7.

5. Zidwick MJ, Chen J-S, Rogers P. Organic acid and solvent production: propionic and butyric acids and ethanol. The Prokaryotes: Springer; 2013. p. 135-67.

6. Propionic Acid Market for Animal Feed & Grain Preservatives, Calcium & Sodium Propionates, Cellulose Acetate Propionate and Other Applications: Global Industry Perspective, Comprehensive Analysis, Size, Share, Growth, Segment, Trends and Forecast, 2014 - 2020 [Available from: https://www.marketresearchstore.com/report/propionic-acid-market-for-animal-feed-grain-z39993.

7. Wu Y, Li B-Z, Zhao M, Mitchell LA, Xie Z-X, Lin Q-H, et al. Bug mapping and fitness testing of chemically synthesized chromosome X. Science. 2017;355(6329):eaaf4706.

8. Zhang W, Zhao G, Luo Z, Lin Y, Wang L, Guo Y, et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. Science. 2017;355(6329):eaaf3981.

9. Jia B, Wu Y, Li B-Z, Mitchell LA, Liu H, Pan S, et al. Precise control of SCRaMbLE in synthetic haploid and diploid yeast. Nature communications. 2018;9(1):1933.

10. Shen M, Wu Y, Yang K, Li Y, Xu H. Heterozygous diploid and interspecies SCRaMbLEing Nat. Nature communications; 2018.

11. Raman S, Rogers JK, Taylor ND, Church GM. Evolution-guided optimization of biosynthetic pathways. Proceedings of the National Academy of Sciences. 2014;111(50):17803-8.

12. Lam FH, Ghaderi A, Fink GR, Stephanopoulos G. Engineering alcohol tolerance in yeast. Science. 2014;346(6205):71-5.

13. Zimmermannova O, Salazar A, Sychrova H, Ramos J. Zygosaccharomyces rouxii Trk1 is an efficient potassium transporter providing yeast cells with high lithium tolerance. FEMS yeast research. 2015;15(4).

14. Mira NP, Lourenço AB, Fernandes AR, Becker JD, Sa-Correia I. The RIM101 pathway has a role in Saccharomyces cerevisiae adaptive response and resistance to propionic acid and other weak acids. FEMS yeast research. 2009;9(2):202-16.

15. Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature. 2011;477(7365):471.

16. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, et al. Total synthesis of a functional designer eukaryotic chromosome. science. 2014;344(6179):55-8.

17. Shen Y, Stracquadanio G, Wang Y, Yang K, Mitchell LA, Xue Y, et al. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. Genome research. 2016;26(1):36-49.

18. Hochrein L, Mitchell LA, Schulz K, Messerschmidt K, Mueller-Roeber B. L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. Nature communications. 2018;9(1):1931.

19. Luo Z, Wang L, Wang Y, Zhang W, Guo Y, Shen Y, et al. Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. Nature communications. 2018;9(1):1930.

20. Gonzalez-Garcia RA, McCubbin T, Navone L, Stowers C, Nielsen LK, Marcellin E. Microbial Propionic Acid Production. Fermentation. 2017;3(2):21.

21. Cameron DC, Cooney CL. A Novel Fermentation: The Production of R (–)–1, 2–Propanediol and Acetol by Clostridium thermosaccharolyticum. Nature Biotechnology. 1986;4(7):651.

22. Staib L, Fuchs TM. Regulation of fucose and 1, 2-propanediol utilization by Salmonella enterica serovar Typhimurium. Frontiers in microbiology. 2015;6:1116.

23. Berg IA, Kockelkorn D, Buckel W, Fuchs G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. Science. 2007;318(5857):1782-6.

24. d Mattozzi M, Ziesack M, Voges MJ, Silver PA, Way JC. Expression of the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: Toward horizontal transfer of autotrophic growth. Metabolic engineering. 2013;16:130-9.

25. Scott KP, Martin JC, Campbell G, Mayer C-D, Flint HJ. Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "Roseburia inulinivorans". Journal of bacteriology. 2006;188(12):4340-9.

26. Saxena R, Anand P, Saran S, Isar J, Agarwal L. Microbial production and applications of 1, 2-propanediol. Indian journal of microbiology. 2010;50(1):2-11.

27. Abeles RH, Lee HA. An intramolecular oxidation-reduction requiring a cobamide coenzyme. Journal of Biological Chemistry. 1961;236(8):2347-50.

28. Kandasamy V, Vaidyanathan H, Djurdjevic I, Jayamani E, Ramachandran K, Buckel W, et al. Engineering Escherichia coli with acrylate pathway genes for propionic acid synthesis and its impact on mixed-acid fermentation. Applied microbiology and biotechnology. 2013;97(3):1191-200.

29. Alber BE, Fuchs G. Propionyl-coenzyme A synthase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation. Journal of Biological Chemistry. 2002;277(14):12137-43.

30. Hasslacher M, Ivessa A, Paltauf F, Kohlwein S. Acetyl-CoA carboxylase from yeast is an essential enzyme and is regulated by factors that control phospholipid metabolism. Journal of Biological Chemistry. 1993;268(15):10946-52.

31. Hügler M, Menendez C, Schägger H, Fuchs G. Malonyl-coenzyme A reductase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation. Journal of bacteriology. 2002;184(9):2404-10.

32. Chen Y, Bao J, Kim I-K, Siewers V, Nielsen J. Coupled incremental precursor and co-factor supply improves 3-hydroxypropionic acid production in Saccharomyces cerevisiae. Metabolic engineering. 2014;22:104-9.

33. Shi S, Liang Y, Zhang MM, Ang EL, Zhao H. A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in Saccharomyces cerevisiae. Metabolic engineering. 2016;33:19-27.

34. Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G. Engineering yeast transcription machinery for improved ethanol tolerance and production. Science. 2006;314(5805):1565-8.

35. Wang Y, Li Y, Pei X, Yu L, Feng Y. Genome-shuffling improved acid tolerance and L-lactic acid volumetric productivity in Lactobacillus rhamnosus. Journal of Biotechnology. 2007;129(3):510-5.

36. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, et al. Design of a synthetic yeast genome. Science. 2017;355(6329):1040-4.

Appendix I: Supplementary material for Chapter 2 (Synthesis and laboratory evolution of *Saccharomyces cerevisiae* synthetic chromosome XIV)

Appendix I

Synthetic chromosome XIV sequence discrepancies and repairs

Whole genome sequencing of synXIV strain revealed a number of missing Sc2.0 features and point mutations that deviated from the intended synXIV sequence. A subset of these were selected for repair according to their relative importance to the project, and ease of re-introduction. All missing TAA stop codons were repaired (Table S2-1), as this modification will serve to free up the TAA codon to encode for non-natural amino acids in the future. All frame-shift mutations in open reading frames were repaired to enable functional expression of the relevant proteins. Some missing LoxP sites were corrected if they were nearby other features already being repaired, but were otherwise left as-is due to the fact that SCRaMbLE has a high degree of redundancy (Table S2-2). PCR-tags are only used to verify the correct insertion of megachunks during the construction phase and were therefore left unaltered if missing, unless they were nearby another fix. Synonymous point mutations in open reading frames were also left unaltered (Table S2-2).

Table S2-1. SynXIV corrected sequence discrepancies

Discrepancy Number	Discrepancy type	Chromosome location	Gene(s)	Protein affect
1	T to C substitution	A3, 12735 bp	YNL329C	R276G
2	T insertion	A3, 13932 bp	YNL328C	Frame-shift
3	T to G substitution	A3, 13944 bp	YNL328C	N98H
4	C to T substitution	A3, 18438 bp	YNL326C	R302Q
5	G to T substitution	A3, 18512 bp	YNL326C	F277L
6	A to G substitution	A3, 18709 bp	YNL326C	Y212H
7	T insertion	A4, 19902 bp	YNL325C	Frame-shift
8	A to G substitution	A4, 22042 bp	YNL325C	synonymous
9	G to A transition	A4, 25505 bp	intergenic	-
10	G to A substitution	A4, 26370 bp	YNL321W	synonymous
11	Missing TAA stop codon	B4, 52811 bp	YNL304W	synonymous
12	Missing LoxP site	D4, 121257 bp	YNL270C	-
13	Missing TAA stop codon and LoxP site	E1, 125772 bp	YNL268W	synonymous
14	T to A substitution	E1, 125952 bp	intergenic	-
15	Missing TAA stop codon	M3, 400292 bp	YNL114C	synonymous
16	Missing TAA stop codon	M3, 400807 bp	YNL113C	synonymous
17	Missing *Bsu36I* restriction site	M3-M4, 401355 bp	YNL112W	synonymous
18	Missing LoxP site	V4, 690721 bp	YNR051C	-
19	Missing TAA stop codon	V4, 690758 bp	YNR051C	synonymous
20	Missing PCR-tag	V4, 690776 bp	YNR051C	synonymous
21	Missing PCR-tags, LoxP, and TAA stop codons	W3-W4, 710310 bp – 717155 bp	YNR059W, YNR060W YNR061C, YNR062C	synonymous

Table S2-2. SynXIV unaltered sequence discrepancies

Discrepancy type	Chromosome location	Gene(s)	Protein affect
C-T substitution	A1-A2 boundary 5863 bp	YNL332W	synonymous
G-A substitution	A2-A3 boundary 12071 bp	YNL329C	S497L
Missing *SfiI* restriction site	D2-D3, 108471 bp	YNL273W	synonymous
Missing PCR-tag	D3, 108741 – 108719 bp	YNL273W	synonymous
Missing PCR-tag and *SfiI* restriction site	E1-E2 boundary, 131105 bp and 131162 bp	YNL264C	synonymous
Missing *BsoBI* restriction site	E2-E3, 139933 bp	YNL262W	synonymous
C to T substitution	E3, 140826 bp	YNL261W	T126I
T to C substitution	E4, 150055 bp	YNL256W	synonymous
T to C substitution	F3, 174665 bp	YNL243W	synonymous
Missing PCR-tag	F3, 174923 bp	YNL243W	synonymous
G to A substitution	F3-F4, 175553 bp	YNL243W	synonymous
Missing PCR-tag	F4, 176150 bp	YNL242W	synonymous
Missing PCR-tag	F4, 180372 bp	YNL242W	synonymous
T to A substitution	M3, 396611 bp	YNL116W	L223I
Missing PCR-tag	N1, 410750 bp	YNL106C	synonymous
Missing LoxP	N4, 438943 bp	YNL092W	-
Missing *BsoBI* restriction site	N4-O1, 439383 bp	YNL091W	synonymous
T to C substitution	P3, 495214 bp	YNL063W	synonymous
T to C substitution	P3, 495229 bp	YNL063W	synonymous
Missing PCR-tag	P3, 495487 bp	YNL063W	synonymous
T to C substitution	P3, 495691 bp	YNL063W	synonymous
Missing *BsoBI* restriction site	P3-P4, 499045 bp	YNL062C	synonymous
T to C substitution	R1, 539674 bp	YNL037C	T99A
Missing PCR-tag	S4, 595657 bp	YNL005C	synonymous
Missing LoxP	T1, 604952 bp	YNR001C	-
Missing *SfiI* restriction site	T1-T2, 603448 bp	YNR001C	synonymous
C to A substitution	U3, 646352 bp	YNR024W	synonymous
Missing PCR-tag	U5, 658820 bp	YNR031C	synonymous
Missing PCR-tag	X1, 725470 bp	YNR065C	synonymous
Missing LoxP site	X3, 748800 bp	YNR072W	synonymous
Missing PCR-tag	X3, 747793 bp	YNR073C	synonymous
Missing *BstEII* restriction site	X3-X4, 749858 bp	YNR073C	synonymous

Table S2-3. Plasmids used in this study.

Names	Details	Origin
Cas9-gRNA-pRS423	*pTEF1-CAS9-CYC1t-SNR52-CAN1.Y-crRNA-CYC1t* -pRS423 with *HIS3* marker	Williams, Xu et al., 2017
MRPL19-sfGFP-loxP-pRS416	*MRPL19-sfGFP-loxP-pRS416* with *Ura3* marker	This study
MRPL19-sfGFP-Native 3′UTR-pRS416	*MRPL19-sfGFP-Native 3′UTR-pRS416* with *Ura3* marker	This study
pPDR12-GFP-pRS416	*pPDR12-GFP-pRS416* with *Ura3* marker	Williams, Xu et al., 2017

Appendix II: Supplementary material for Chapter 3 (Positive-feedback, ratiometric biosensor expression improves high-throughput metabolite-producer screening efficiency in yeast)

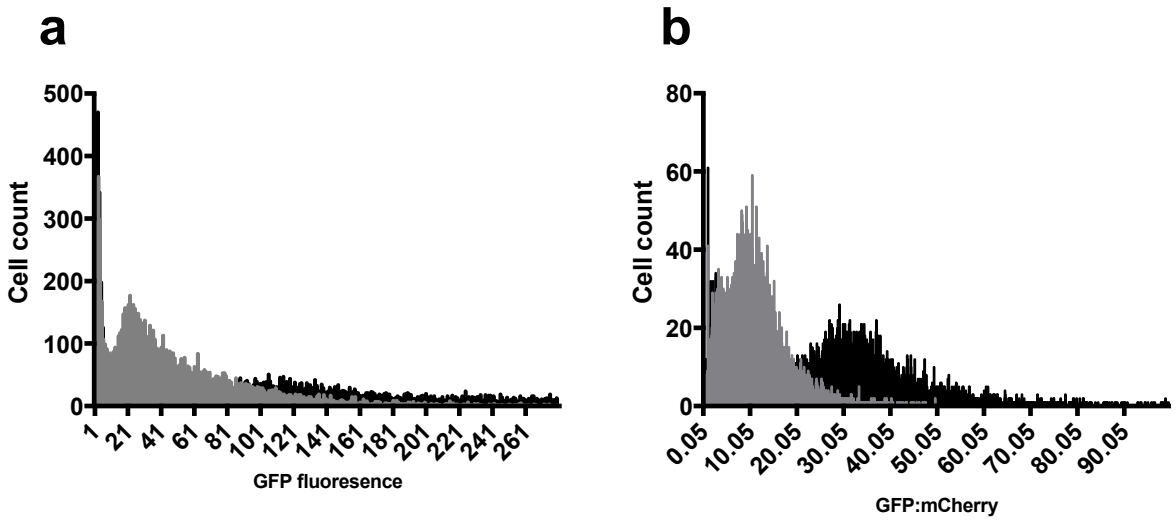


Figure S3-1. PHBA mediated biosensor fluorescence with and without ratiometric normalization. (a) GFP fluorescence distribution of the positive-feedback ratiometric biosensor strain (+FB.GFP.mCherry.415, Table 3) treated with (black bars) and without (grey bars) 50 mM PHBA. (b) The same strain analysed the ratio of GFP to mCherry fluorescence within each cell. Data from single cultures which are representative of repeated experiments are shown here.

Appendix III: Supplementary material for Chapter 5 (Evolutionary engineering in *Saccharomyces cerevisiae* reveals a *TRK1*-dependent potassium influx mechanism for propionic acid tolerance)

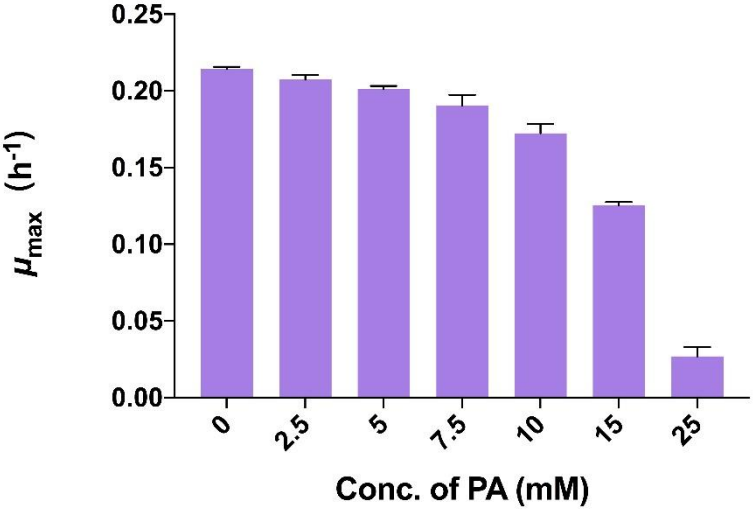


Figure S5-1. PA effect to the growth of *S. cerevisiae*. Bars and error bars represent the mean and standard deviation (SD) of triplicate cultures.

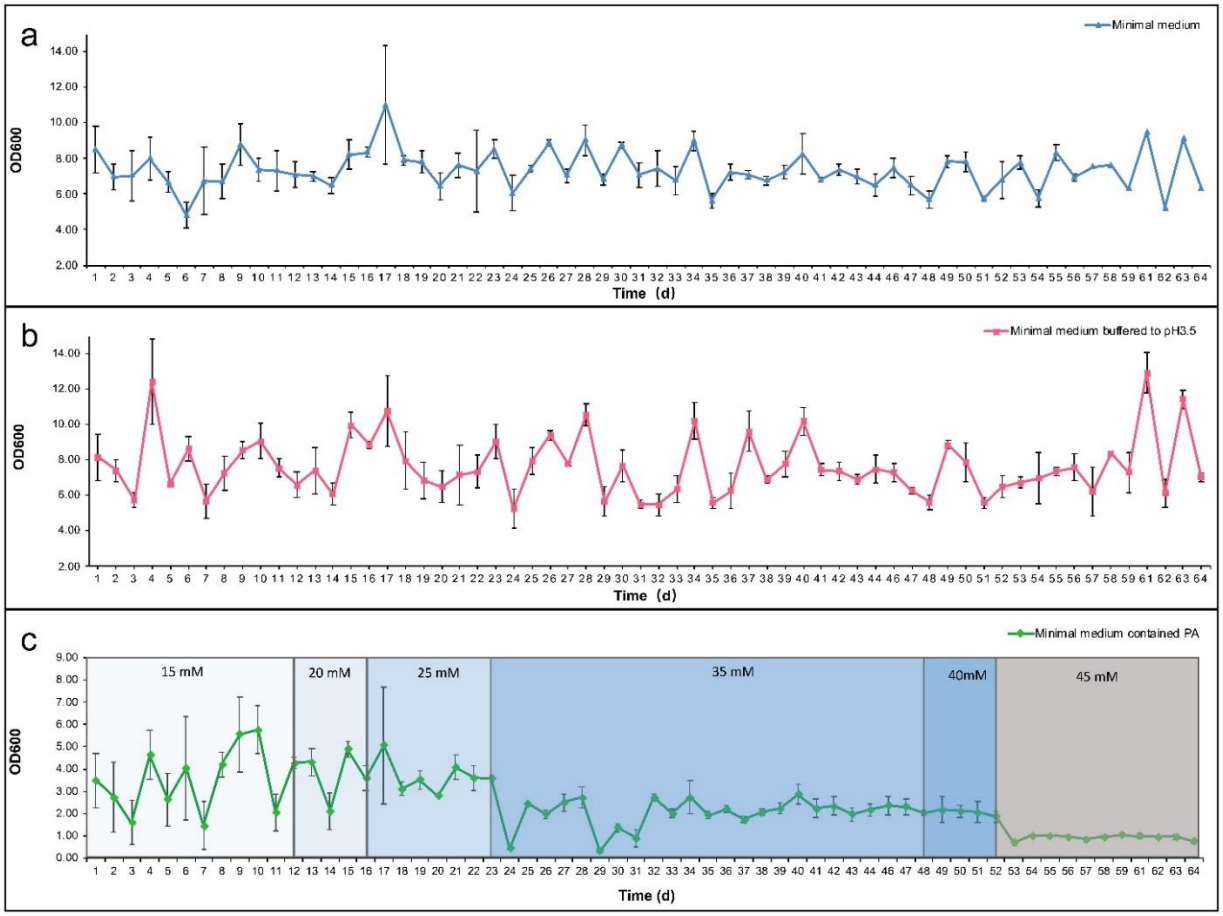


Figure S5-2. The fluctuations of yeast growth through adaptive laboratory evolution. Changes of cell density in minimal medium (pH 5) (a), in buffered minimal medium (pH 3.5) (b), and in buffered minimal medium (pH 3.5) with increasing concentrations of PA from 15 mM to 45 mM (c). Data represents the average of OD_{600} for 3 replicates and SD value is given by error bars.

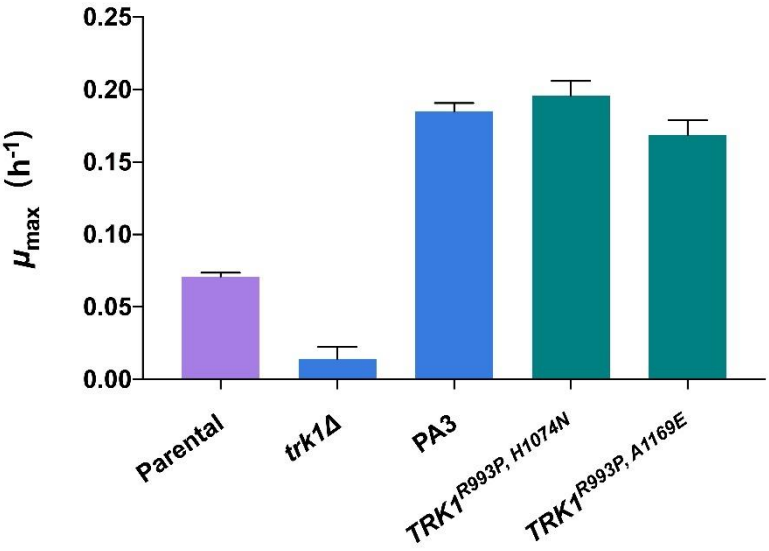


Figure S5-3. Fitness test of *TRK1* mutants containing different combinations of two mutations in 35 mM PA. The evolved isolate PA-3 was engineered with different combinations of two *TRK1* mutations (*TRK1*$^{R993P, H1074N}$ and *TRK1*$^{R993P, A1169E}$). Growth rates of these strains, a *TRK1* deletion strain, and the parental strain were determined in buffered minimal medium containing 35 mM PA. Bars and error bars represent the mean and SD of triplicate cultures.

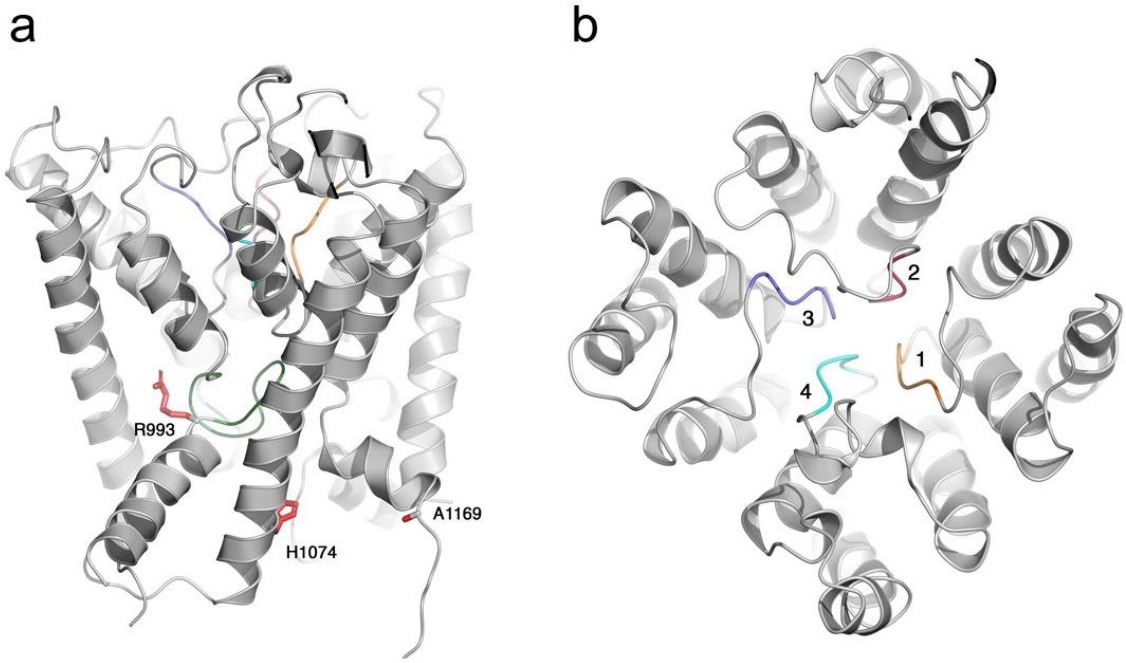


Figure S5-4. Cartoon showing the overall fold of the ScTrk1 channel. Side view, perpendicular to the channel (a), and top view looking down the channel towards the selectivity filter (b). The four segments that constitute the selectivity filter are colored orange, red, blue and cyan, respectively. The amino acids that are subject to mutations (R993, H1074, A1169) are colored red and shown as stick objects. The intramembrane loop (IML, residues 983-992) "below" the selectivity filter is colored green.

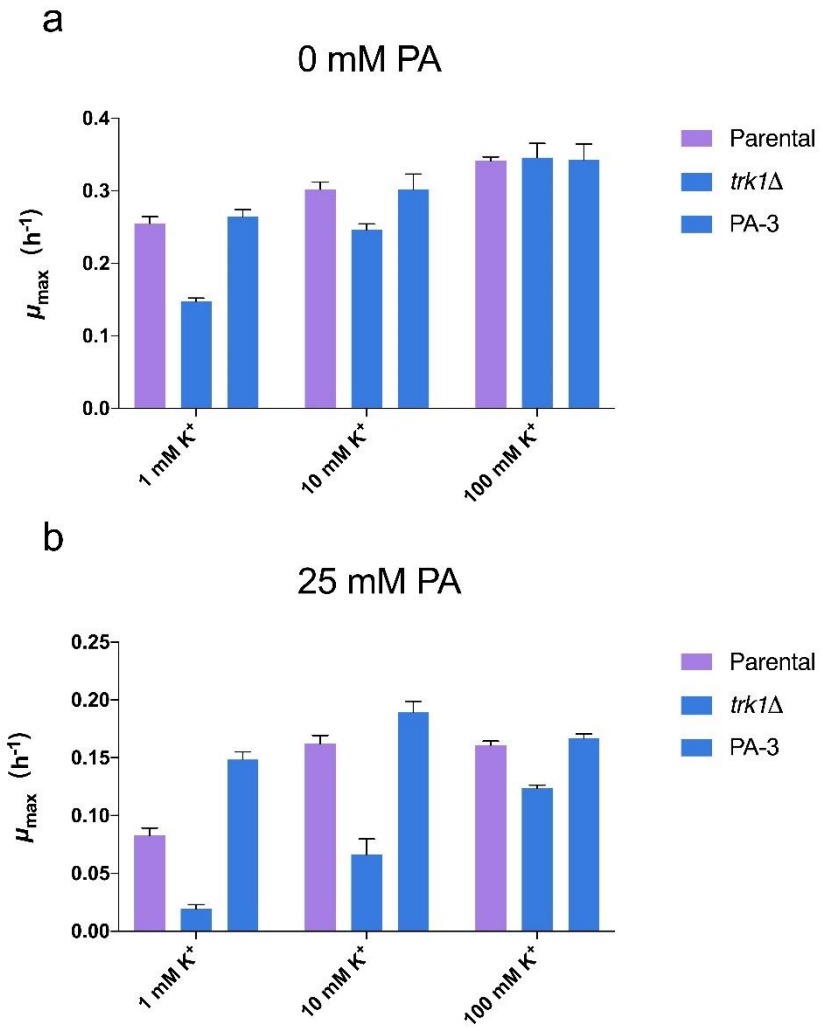


Figure S5-5. The effect of potassium concentrations and *TRK1* on the tolerance of yeast strains to PA in liquid culture. Growth rates of WT, *trk1Δ* and the evolved isolate PA-3 were determined in minimal medium supplemented with increasing concentrations of potassium (1 mM, 10 mM, and 100 mM) with 0 mM (a) and 25 mM PA (b). Bars and error bars represent the mean and SD of triplicate cultures.

Table S5-1. List of plasmids used in this study.

Name	**Details**	**Origin**
pRS413	Yeast centromeric plasmid, *HIS3* marker	Euroscarf
Cas9-gRNA-pRS423	*pTEF1-CAS9-CYC1t-SNR52-CAN1.Y-crRNA-CYC1t* -pRS423, *HphMX* marker	This lab
pPGK1-CYC1t-pRS426	*pPGK1-CYC1t*-pRS426, *URA3* and *HphMX* marker	This lab
pPGK1-TRK1-CYC1t-pRS426	*pPGK1-TRK1-CYC1t*-pRS426, *URA3* and *HphMX* marker	This study
pPGK1-TRK1-CYC1t-pRS413	*pPGK1-TRK1-CYC1t*-pRS413, *HphMX* and *HIS3* marker	This study

Table S5-2. List of primers used in this study.

Primer name	5′ to 3′ sequence
TRK1F	TAGATGCATCATTTGGATAATG
TRK1R	TTAGAGCGTTGTGCTGCT
TRK1-3223F	ACGGATACTGAAGACGATGGT
TRK1-3223R	GGTGTCCGGATAACCTAGCG
TRK1-3509F	TGCATTTGTGAAGGGGACAAG
TRK1-3509R	GCGTTGTGCTGCTCCTTTTA
pCRISPR-TRK1-3223 gRNAF	ATGAAAGCTGTTTCCTTAAAGTTTTAGAGCTAGAAATAGCAAG
pCRISPR-TRK1-3223 gRNAR	TTTAAGGAAACAGCTTTCATGATCATTTATCTTTCACTGCGGA
pCRISPR-TRK1-2981 gRNAF	CATTGGCCATCTCTATTCGAGTTTTAGAGCTAGAAATAGCAAG
pCRISPR-TRK1-2981 gRNAR	TCGAATAGAGATGGCCAATGGATCATTTATCTTTCACTGCGGA
pCRISPR-TRK1-3509 gRNAF	ATTGCACGATCCAGTGAGTAGTTTTAGAGCTAGAAATAGCAAG
pCRISPR-TRK1-3509 gRNAR	TACTCACTGGATCGTGCAATGATCATTTATCTTTCACTGCGGA
TRK1 seq checkF	GTTTCAATCTGTTAGCACAAG
TRK1 seq checkR	CATCAGAAATGTACGTAGGC
Hph-TRK1F	TTTTAGAAGAACGATGAGTAGAGTGCCCACATTGGCATCTCTTGA AATACGATATAAAAAATCTTTCGGCGTTTTTCGACACTGGATGGC
Hph-TRK1R	GTGCTGCTCCTTTTAGGATTTCGGGAATGTGTGGTCTTACGCTTAA GAGCTCCCCAACGATGTTTCACATCACGCGTTTAGCTTGCCTCG
Hph-upstreamF	GAGGGGATGTTATTTCAAGCAC
Hph-upstreamR	GCCATCCAGTGTCGAAAAACGCCGAAAGATTTTTTATATCGTATT
Hph-downstreamF	CGAGGCAAGCTAAACGCGTGATGTGAAACATCGTTGG
Hph-downstreamR	AATCGATGAGTGGGGATT
TRK1-Hph-knock-seqF	CTCTCGCTAAATTCCCCAATG
TRK1-Hph-knock-seqR	GAGGGACAATGTACTAATGGC
TRK1 with pRS426 overhangF	AACAAATTTAATTAAGGTTCCAAGGATGCATTTTAGAAGAACGAT GAG
TRK1 with pRS426 overhangR	AAAGGAAAAGGCGCGCCTTGGAATTAGAGCGTTGTGCTGC
pRS426 with TRK1 overhangF	TAAAAGGAGCAGCACAACGCTCTAATTCCAAGGCGCGCCT
pRS426 with TRK1 overhangR	CTCATCGTTCTTCTAAAATGCATCCTTGGAACCTTAATTAAATTTG TT
TRK1-pPGK1-pRS426-checkF	GCGATCGCTCCCTCCTTCT
TRK1-CYC1t-pRS426-checkR	GGTTAAACGGTTCCAAGGCC
pRS426-pPGK1 with pRS413 overhangF	CGAGGTCGACGGTATCGATTGGGTAACGCCAGGGTTTT
pRS426-Hph with pRS413 overhangR	GGAATTCGATATCAAGCTTATGTGTGGAATTGTGAGCGGA
pRS413-with pRS426 overhangF	TCCGCTCACAATTCCACACATAAGCTTGATATCGAATTCC
pRS413-with pRS426 overhangR	AAAACCCTGGCGTTACCCAATCGATACCGTCGACCTCG
pPGK1-TRK1-CYC1t-pRS413 checkF	ATGACAGAGCAGAAAGCCCTAG
pPGK1-TRK1-CYC1t-pRS413 checkR	GTGTAGAAGTAGTGAACCGCGA

Table S5-3. Genotypic changes in the PA evolved populations.

Gene name	Mutations			Gene function
	Lineage-1	Lineage-2	Lineage-3	
TRK1	H1074N	R993P	A1169E	Component of the Trk1p-Trk2p potassium transport system; 180 kDa high affinity potassium transporter
OCT1	/	A714V	/	Mitochondrial intermediate peptidase; may contribute to mitochondrial iron homeostasis
FKH2	/	M626I	/	Forkhead family transcription factor; rate-limiting activator of replication origins
THP3	/	Q469K	/	Protein that may have a role in transcription elongation; possibly involved in splicing based on pre-mRNA accumulation defect for many intron-containing genes
PUS9	/	/	V404G	Mitochondrial tRNA: pseudouridine synthase
HNM1	/	/	CDS176(AA59) Truncation	Plasma membrane transporter for choline, ethanolamine, and carnitine; involved in the uptake of nitrogen mustard and the uptake of glycine betaine during hypersaline stress

The mutations were identified from end point PA evolved populations after excluding the mutations existed in the parental strain and the strains from control conditions.

Appendix IV: Biosafety approval letter

Office of the Deputy Vice-Chancellor (Research)

Research Office
C5C Research HUB East, Level 3, Room 324
MACQUARIE UNIVERSITY NSW 2109 AUSTRALIA

Phone **+61 (0)2 9850 6848**
Fax +61 (0)2 9850 4465
Email biosafety@mq.edu.au

24th November 2014

Dear Professor Paulsen,

Re: "Yeast 2.0" (Ref: 5201401059)

Thank you for your Biosafety application. Your application has been reviewed by the Institutional Biosafety Committee (IBC) and approval of the above Exempt Dealing has been granted, effective 24th November 2014.

Approval has been granted subject to your compliance with the Office of the Gene Technology Regulator's standard conditions for exempt work listed below:

1. The project must be conducted in accordance with the OGTR Guidance Notes for the Containment of Exempt Dealings (http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/Content/ExemptDealGuideSept11_2-htm)

2. The Guidance Notes are only applicable to exempt dealings conducted under the *Gene Technology Act 2000*. They do not provide guidance for laboratory safety, good laboratory practice or work health and safety issues

For these purposes, refer to AS/NZS 2243.3:2010.

3. You must inform the Institutional Biosafety Committee if you complete or abandon the exempt dealings with GMOs.

Please note the following standard requirements of approval:

1. Approval will be for a period of 5 years *subject to the provision of annual reports*. If, at the end of this period the project has been completed, abandoned, discontinued or not commenced for any reason, you are required to submit a Final Report. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. Please contact the Committee Secretary at biosafety@mq.edu.au in order to obtain a report.

A Progress/Final Report for this project will be due on: 1st November 2015

2. If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of Final Approval to an external organisation as evidence that you have Final Approval, please do not hesitate to contact the Committee Secretary at biosafety@mq.edu.au or by phone 9850 4063.

Please retain a copy of this email as this is your formal notification of final Biosafety approval.

Yours Sincerely
A/Prof Subramanyam Vemulpad
Chair, Macquarie University Institutional Biosafety Committee

Biosafety Secretariat
Research Office
Level 3, Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia

T: +61 2 9850 6848
F: +61 2 9850 4465

http://www.mq.edu.au/research

MACQUARIE UNIVERSITY
SYDNEY ~ AUSTRALIA