

**Single-step in vivo assembly of synthetic chromosomes with a modular design for engineering yeast metabolism**

Postma, E.D.

**DOI**

[10.4233/uuid:a35a96bf-4ca9-4067-b925-d7b0f5aaeabd](https://doi.org/10.4233/uuid:a35a96bf-4ca9-4067-b925-d7b0f5aaeabd)

**Publication date**

2022

**Document Version**

Final published version

**Citation (APA)**

Postma, E. D. (2022). *Single-step in vivo assembly of synthetic chromosomes with a modular design for engineering yeast metabolism*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:a35a96bf-4ca9-4067-b925-d7b0f5aaeabd>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

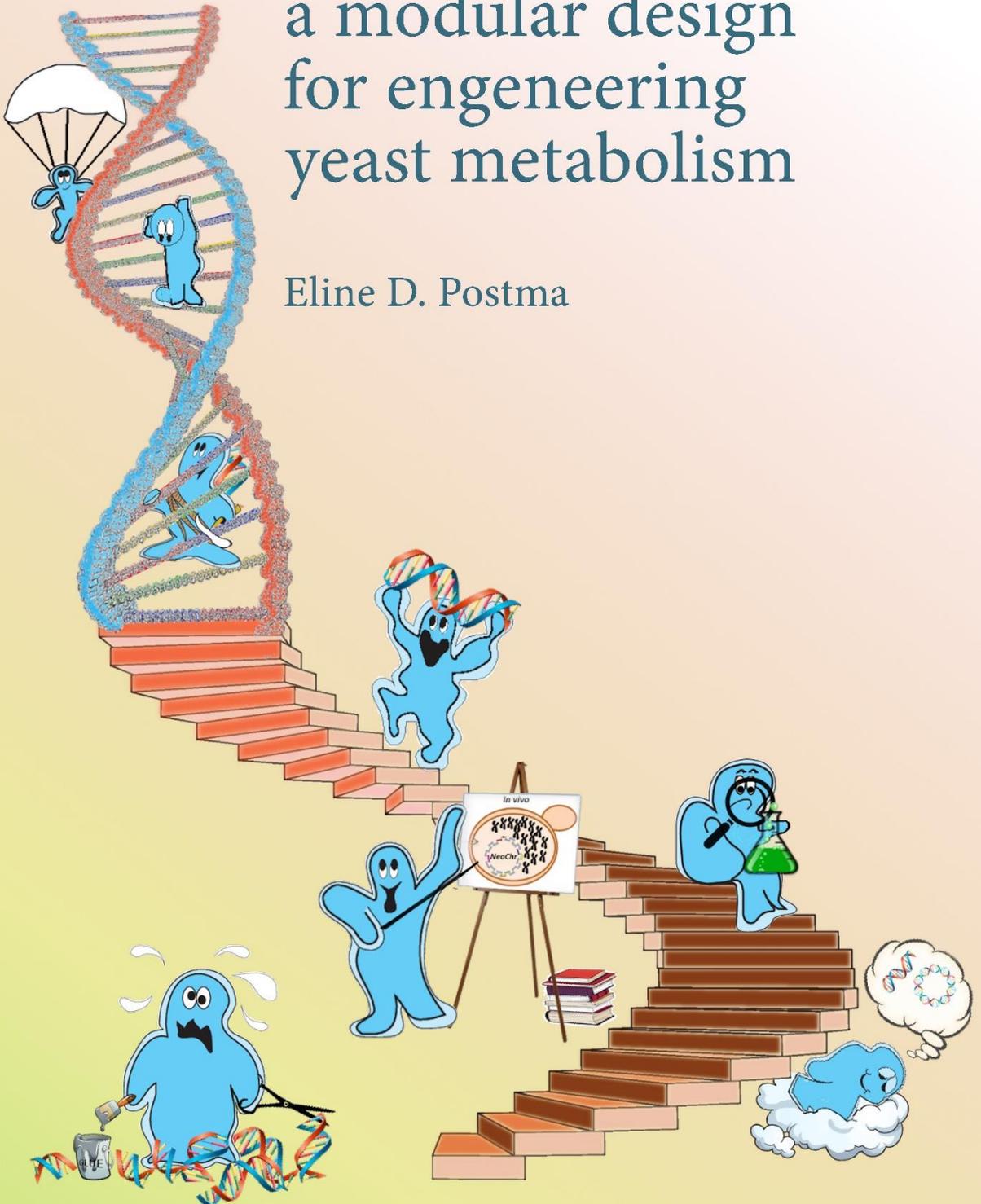
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# Single-step *in vivo* assembly of synthetic chromosomes with a modular design for engineering yeast metabolism

Eline D. Postma



**Single-step *in vivo* assembly of synthetic  
chromosomes with a modular design for  
engineering yeast metabolism**

**Proefschrift**

Ter verkrijging van de graad van doctor aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen,  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op  
Vrijdag 21 januari 2022 om 12:30 uur

door

**Eline Daphne POSTMA**

Master of Science in Biology, Universiteit Leiden  
geboren te Hoorn, Nederland

Dit proefschrift is goedgekeurd door de promotoren  
Prof. dr. P.A.S. Daran-Lapujade en Prof. dr. J.T. Pronk

Samenstelling promotiecommissie:

Rector Magnificus	Voorzitter
Prof. dr. P.A.S. Daran-Lapujade	Technische Universiteit Delft, promotor
Prof. dr. J.T. Pronk	Technische Universiteit Delft, promotor

Onafhankelijke leden:

Prof. dr. A.M. Dogterom	Technische Universiteit Delft
Prof. dr. J. van der Oost	Universiteit Wageningen
Dr. M.M.M. Bisschops	Universiteit Wageningen
Prof. dr. G.P. van Wezel	Universiteit Leiden
Prof. dr. T. Ellis	Imperial College London, VK

Reservelid:

Prof. dr. F. Hollmann	Technische Universiteit Delft
-----------------------	-------------------------------

The research presented in this thesis was performed at the Industrial Microbiology Group, Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, The Netherlands and was funded by a consolidator grant AdLibYeast from the European Research Council (ERC).



European Research Council  
Established by the European Commission

**Cover:** Carla van Leeuwen  
**Layout:** Eline Postma  
**Printed by:** Ipskamp Printing B.V.  
**ISBN:** 978-94-6421-604-2

© 2021 Eline D. Postma

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronically, mechanically, by photo-copying, recording or otherwise, without the prior written permission of the author.

## Contents

<b>Samenvatting</b> .....	5
<b>Summary</b> .....	11
<b>Chapter 1: General introduction</b> .....	15
<b>Chapter 2: Top-down, knowledge-based genetic reduction of yeast central carbon metabolism</b> .....	47
<b>Chapter 3: What's wrong with SwYG? Reaching the limits of <i>Saccharomyces cerevisiae</i> molecular and analytical toolbox</b> .....	97
<b>Chapter 4: A supernumerary designer chromosome for modular <i>in vivo</i> pathway assembly in <i>Saccharomyces cerevisiae</i></b> .....	129
<b>Chapter 5: Modular, synthetic chromosomes as new tools for large scale engineering of metabolism</b> .....	197
<b>Outlook</b> .....	267
<b>Acknowledgements</b> .....	277
<b>Curriculum vitae</b> .....	281
<b>List of publications</b> .....	282



## Samenvatting

Micro-organismen worden al sinds de oudheid gebruikt om gefermenteerd voedsel en dranken te maken, zoals wijn, brood en bier. Met de huidige technieken in de moleculaire biologie kunnen microbiële prestaties worden verbeterd die betrekking hebben op producten die een micro-organisme van nature produceert. Bovendien is het mogelijk om micro-organismen aan te passen om producten te maken die van nature alleen door andere organismen worden gemaakt, of zelfs producten te laten maken die niet in de natuur voorkomen. De op deze wijze gemaakte, zogenaamde microbiële celfabrieken, kunnen brandstoffen en chemicaliën produceren die de verontreinigende en niet duurzame industrie van fossiele brandstoffen kan vervangen. *Saccharomyces cerevisiae*, ook wel bekend als bakkersgist, wordt al decennialang gebruikt als een favoriete celfabriek om een scala aan industrieel relevante producten te produceren. Daarnaast is deze gist een belangrijk modelorganisme in fundamenteel onderzoek. Naast zijn simpele voedingsbehoeften en makkelijke hanteerbaarheid, heeft *S. cerevisiae* zijn populariteit ook te danken aan zijn opmerkelijke vermogen om DNA-moleculen efficiënt en precies aan elkaar te maken via homologe recombinatie (HR). *S. cerevisiae* gebruikt dit natuurlijke mechanisme om potentieel schadelijke breuken in zijn genetische materiaal te repareren. Moleculair genetici gebruiken HR daarentegen om DNA naadloos in te bouwen of te verwijderen. De combinatie van het CRISPR/Cas-systeem voor genoom-modificatie en de HR-precisie-reparatie maakt een hermodellering van het genoom voor de bouw van celfabrieken mogelijk. Het huidige nadeel van deze aanpak is dat slechts een gelimiteerd aantal DNA-veranderingen in één transformatieronde kan worden bewerkstelligd. Dit maakt uitgebreide hermodellering van het genoom nog steeds tot een tijdrovend proces. Het doel van dit proefschrift was om een strategie te ontwikkelen en te evalueren om het metabolisme van *S. cerevisiae* efficiënt en op grote schaal aan te kunnen passen.

Het genoom van *S. cerevisiae* (en van eukaryoten in het algemeen) wordt gekarakteriseerd door een mozaïekstructuur, waarin genen, ongeacht hun cellulaire functie, over de 16 chromosomen verspreid liggen. Bovendien vertoont het gistgenoom een grote mate van genetische redundantie, waarbij meerdere eiwitten dezelfde functie uitvoeren. Sommige van deze eiwitten kunnen misschien belangrijk zijn onder een specifieke groeiomstandigheid, terwijl ze onnodig zijn in een gecontroleerde laboratoriumsituatie of in industriële processen. Verwijdering van de genen die coderen voor deze eiwitten, zou genetische modificatie sterk kunnen vereenvoudigen. Het centrale koolstofmetabolisme (CKM) omvat ongeveer 53 biochemische reacties en vormt een belangrijk deel van het centrale stofwisselingsnetwerk van bakkersgist. Het CKM is de leverancier van de essentiële bouwstenen, co-factoren en energierijke moleculen die nodig zijn voor het maken van

alle industrieel relevante producten die door bakkersgist worden gemaakt. Ook het CKM wordt gekarakteriseerd door een hoge mate van genetische redundantie. In een vorig onderzoek is een eerste stap gezet om deze genetische complexiteit te reduceren, door 13 van de 26 genen betrokken bij de glycolyse en alcoholische fermentatie te verwijderen. De 50% reductie van het aantal genen betrokken bij deze stofwisselingsroutes in de resulterende “minimale glycolyse” (MG) giststam had opmerkelijk weinig invloed op de fysiologie. Dit bleek onder meer uit een vergelijkbare groeisnelheid onder een breed scala van groei- en stresscondities. Gebruikmakend van de MG-stam als uitgangspunt, wordt in **Hoofdstuk 2** de genetische reductie uitgebreid naar het gehele CKM, met inbegrip van de pentosefosfaatroute, citroenzuurcyclus, anaplerotische reacties, mitochondriële transporteiwitten, glyoxylaatcyclus, gluconeogenese, acetyl-CoA-metabolisme en glycerolsynthese. In totaal werden 35 van de betrokken 111 genen in het CKM verwijderd, wat resulteerde in een reductie van genetische complexiteit in het CKM van 32%. Deze reductie leidde tot slechts een marginale reductie in groeisnelheid in gecontroleerde bioreactoren op een synthetisch medium met glucose als koolstofbron. Ook onder de meeste andere onderzochte groei- en stresscondities werd geen of weinig reductie in groeisnelheid waargenomen. Opvallende afwijkingen, waarbij significante verschillen in fysiologie werden gedetecteerd, werden slechts gevonden bij groei op pyruvaat en ethanol als koolstofbron, bij anaerobe groei op glucose en bij groei onder hoge osmotische stress. Gebruikmakend van een grote verscheidenheid aan intermediaire deletie-stammen konden voor een aantal van deze groeicondities de verantwoordelijke snelheids-reducerende gen-deleties worden geïdentificeerd. Voor andere groeicondities leidden deze experimenten tot eliminatie van een aantal hypothesen over de oorzaak van langzamere groei. De minimale CKM-stam is een uitstekende basis voor fundamenteel onderzoek naar het gistmetabolisme en voor intensieve en uitgebreide genommodificatie met de strategieën die worden geïntroduceerd in **Hoofdstuk 4 en 5**.

Zelfs als het CKM versimpeld kan worden door redundante (overbodige) genen te verwijderen, blijft aanpassing omslachtig door de verspreide ligging van de overgebleven genen in het genoom. In een eerdere studie is het concept van “pathway swapping” ontwikkeld. Hierbij worden genen betrokken bij een metabool proces op één locatie van het genoom geclusterd, zodat het gehele proces in twee stappen kan worden verwisseld voor een alternatief ontwerp. Met de combinatie van glycolyse en alcoholische fermentatie als modelsysteem werden als "proof of concept" de 13 overgebleven genen in de MG-stam geclusterd op één enkele chromosomale plek (locus) en de genen van hun natuurlijke plekken (loci) verwijderd. Terwijl de MG-stam geen merkbare verschillen in fysiologie vertoonde, groeide de stam met de geclusterde glycolyse en alcoholische fermentatie, SwYG (Switchable Yeast Glycolysis) genaamd,

substantieel langzamer dan een controlestam die de complete set van verspreide genen voor glycolyse en alcoholische fermentatie bevatte (10-30% langzamer afhankelijk van de meetcondities). In **Hoofdstuk 3** wordt de oorzaak van deze reductie in groeisnelheid onderzocht. Met behulp van een hypothese-gestuurde benadering werden meerdere potentiële oorzaken uitgesloten, waaronder gebruik van “marker genes”, verandering van glycolytische capaciteit en globale veranderingen van genexpressie of DNA-replicatie als gevolg van de organisatie van het integratielocus. “Adaptive laboratory evolution” en “reverse engineering” maakten het mogelijk de groeisnelheid van de SwYG-stam te herstellen naar die van de ouderlijke stam. In de aldus geëvolueerde gist werden twee verantwoordelijke mutaties, in *ATG41* en *SUR2*, geïdentificeerd. Beide genen zijn direct of indirect betrokken bij autofagie. Op basis van deze waarneming werden twee, mogelijk samenhangende, hypothesen opgesteld: Een ongewenste nevenmutatie in *VPS15*, een gen dat ook indirect gelinkt kan worden aan autofagosoomvorming, kan verantwoordelijk zijn voor de verminderde groeisnelheid. De tweede hypothese veronderstelt een mogelijke, maar niet onderzochte “moonlighting”-activiteit van fosfoglyceraatkinase (Pgk1) in autofagie, een goed gekarakteriseerde functionaliteit van het humane ortholoog van Pgk1. Hoewel een in de SwYG-stam waargenomen lagere enzymatische capaciteit van Pgk1 geen negatieve gevolgen leek te hebben voor de glycolytische flux, zou deze wel autofagie kunnen beïnvloeden. Deze hypothese zou daarmee een positieve invloed van de mutaties in *ATG41* en *SUR2* in de geëvolueerde SwYG-stammen kunnen verklaren. De in dit hoofdstuk onderzochte effecten vormen een treffende illustratie van de uitdagingen in onderzoek aan micro-organismen waarin complexe genetische modificaties zijn geïntroduceerd.

Doordat het CKM van belang is voor aanlevering van bouwstenen, energie (ATP) en de redoxbalans tijdens productvorming, zal er bij programma’s voor optimalisering van celfabrieken rekening moeten worden gehouden met het gehele CKM. Alleen op deze wijze kunnen celfabrieken worden aangepast om een maximale productconcentratie (titer), productiesnelheid en productopbrengst te behalen. Intensieve en uitgebreide hermodellering van het CKM kan in principe worden vergemakkelijkt door de uitbreiding van het “pathway swapping” concept van de glycolyse naar het gehele CKM. Dit betekent echter dat een groot aantal genen (76 in het geval van de geminimaliseerde CKM in **Hoofdstuk 2**) zal moeten worden gecentraliseerd op één chromosoom, wat het risico met zich meebrengt van verstoring van de natuurlijke structuur, functie en expressie van gistchromosomen. In **Hoofdstuk 4** wordt de toepasbaarheid van een nieuw, extra, 17<sup>de</sup> (neo)-chromosoom onderzocht als platform om genen voor metabolisme te assembleren, tot expressie te brengen en aan te passen. Bij dit onderzoek werd gebruik gemaakt van de opmerkelijke HR-machinerie die bakkersgist in staat stelt om DNA sequenties nauwkeurig en met hoge efficiëntie te

assembleren. Ontwerp-strategieën zijn belangrijk voor de constructie van dergelijke neochromosomen. Het ideale scenario om op een zo eenvoudig mogelijke manier het metabolisme aan te passen, is een modulaire assemblage, gebaseerd op een *in silico* ontwerp van individuele transcriptie-eenheden met passende promotor- en terminatorsequenties. In **Hoofdstuk 4** wordt aangetoond dat in één stap 100 kb test-neochromosomen, voornamelijk bestaande uit niet-coderend DNA, *in vivo* kunnen worden geassembleerd uit tot wel 44 DNA-fragmenten (met de grootte van transcriptie eenheden), met een hoge efficiëntie en sequentie-getrouwheid. Een belangrijke waarneming hierbij was dat de op deze wijze geassembleerde neochromosomen extreem stabiel waren voor wat betreft hun replicatie en segregatie bij vermeerdering of doorkweken. Tenslotte liet verplaatsing van de minimale set aan genen betrokken bij glycolyse en alcoholische fermentatie naar het neochromosoom zien dat de neochromosomen gebruikt kunnen worden als “landingsplatform” voor de genen betrokken bij complete metabole routes, en dat ze sets aan geclusterde genen kunnen dragen die hoog tot expressie komen en die betrokken zijn bij essentiële cellulaire processen.

In **Hoofdstuk 5** worden de neochromosomen verder onderzocht en geïmplementeerd als middel voor het aanpassen van de stofwisseling van bakkersgist. De in **Hoofdstuk 4** beschreven neochromosomen waren circulair terwijl in **Hoofdstuk 5** voor het eerst wordt aangetoond dat ook 100 kb lineaire chromosomen in één stap *in vivo* geassembleerd kunnen worden uit 43 fragmenten, waarbij tegelijkertijd stabiele einden van chromosomen (telomeren) worden gevormd. Ondanks dit succes bleek de assemblage van de lineaire neochromosomen minder efficiënt en sequentiegetrouw plaats te vinden dan die van de circulaire neochromosomen. Eenmaal geassembleerd vertoonden de lineaire en circulaire chromosomen echter een identieke stabiliteit en hadden ze een vergelijkbaar effect op de fysiologie van de giststam. Als bewijs van bruikbaarheid van neochromosomen voor grootschalige aanpassingen aan metabole processen werd de complete set van genen die benodigd is voor het maken van het anthocyaan pelargonidine 3-O-glucoside (P3G; een kleurstof en nutraceutical), van substraat tot product, ge(her)lokaliseerd op het synthetische chromosoom. De biosynthese van P3G begint bij glucose, en loopt via de glycolyse, de pentosefosfaatroute, de shikimaatroute naar een vanuit planten afkomstige route naar anthocyaanverbindingen. Eerst werden 42-43 fragmenten, waaronder een minimale set aan glycolyse- en pentosefosfaatroute genen, als mede de genen die in *Escherichia coli* coderen voor de enzymen van de shikimaatroute, in één stap *in vivo* geassembleerd in circulaire en lineaire neochromosomen. Deze chromosomen werden vervolgens gebruikt als landingsplatforms voor integratie van de genen van de anthocyaanroute uit planten. Het uiteindelijke neochromosoom bevatte 41, volledig functionele genen, waardoor P3G kon worden geproduceerd in bakkersgist. Dit

geslaagde voorbeeld van de implementatie van neochromosomen voor het aanpassen van het metabolisme, opent mogelijkheden voor grootschalige genoom-modificatie door middel van neochromosomen. Het is goed mogelijk dat wanneer we de minimale set aan genen van het CKM gedefinieerd in **Hoofdstuk 2** combineren met het “pathway swapping” concept en de neochromosomen in **Hoofdstuk 4 en 5**, we toekomstige celfabrieken kunnen maken met modulaire genomen waarin genen voor het metabole kernnetwerk en productvormingsroutes zijn gelokaliseerd op gespecialiseerde neochromosomen die vervolgens kunnen worden verwisseld voor alternatieve ontwerpen. Daarnaast zouden de neochromosomen kunnen functioneren als landingsplatform om er andere heterologe functionaliteiten dan stofwisseling op te plaatsen.



## Summary

Microorganisms have been used by humanity since ancient times to produce fermented food and beverages such as wine, bread and beer. With the current molecular biology techniques, it is not only possible to increase microbial performance for the production of native products, but also to adapt microorganism to make heterologous or even novel-to-nature compounds. Hereby, microbial cell factories can produce fuels and chemicals to replace the polluting and unsustainable fossil fuel industry. *Saccharomyces cerevisiae*, also known as baker's yeast has been used since decades as a preferred host to produce a range of industrial relevant compounds and as model organism in fundamental research. Next to its simple nutritional requirements and tractability, the popularity of *S. cerevisiae* also stems from its remarkable ability to efficiently and precisely stitch DNA molecules together via homologous recombination (HR). While *S. cerevisiae* uses this native mechanism to repair deleterious breaks in its genetic material while remaining faithful to the original sequence, HR has been harnessed by genetic engineers to seamlessly introduce or delete specific DNA sequences. The combination of CRISPR/Cas genome editing ability with HR precision repair enables genome remodelling for the construction of cell factories. A current drawback of this approach is that only a limited number of DNA modifications can be achieved in one transformation round, making extensive remodelling a time-consuming process. The goal of this thesis was to design and evaluate a strategy for extensive and efficient engineering of metabolism in *S. cerevisiae*.

The genome of *S. cerevisiae* (and eukaryotes in general) is characterized by a mosaic structure, in which genes, irrespective of their cellular function, are scattered over the 16 chromosomes, and by genetic redundancy, the fact that multiple proteins can have the same function. While these proteins might be important under specific environmental conditions, they could be dispensable in a controlled laboratory environment or a controlled industrial process. Removal of redundant genes, coding for these proteins could greatly simplify genetic engineering. The central carbon metabolism (CCM) encompassing approximately 53 reactions, is an important part of the metabolic network that delivers precursors, cofactors and energy-rich molecules that are required for all industrially relevant products made by yeast. CCM is characterized by a high degree of genetic redundancy. In a previous study, a first step was made in the reduction of this genetic complexity by removing 13 from the 26 genes from glycolysis and the ethanolic fermentation. In this minimal glycolysis (MG) strain, the 50% gene reduction was remarkably harmless, resulting in no measurable phenotypic effect under a wide range of growth and stress conditions. Using the MG strain as starting point, **Chapter 2** expands this genetic reduction to the entire CCM,

including the pentose phosphate pathway, the TCA cycle, anaplerotic reactions, mitochondrial carriers, the glyoxylate cycle, gluconeogenesis, acetyl-CoA metabolism, and glycerol synthesis. In total 35 genes of the 111 genes present in the CCM were deleted resulting in a reduction of genetic complexity of 32%. Only a marginal decrease in growth rate could be observed in controlled bioreactors with synthetic medium and glucose as carbon source. No or little decrease in growth rate were observed under a broader range of stress and growth conditions. Remarkable exceptions, where more significant changes in physiology were detected, were growth on pyruvate and ethanol as carbon source, anaerobic growth on glucose and growth under high osmotic stress conditions. Using a library of intermediate deletion strains, for some of these conditions genes responsible for the decreases in growth rate could be identified, while for other conditions, hypothetical targets could be excluded. This minimal CCM strain presents an excellent platform for fundamental research on yeast metabolism and for intensive and extensive genome engineering by strategies introduced in **Chapter 4 and 5**.

Even though metabolism can be simplified by deleting redundant genes, engineering metabolism remains cumbersome due to the scattered arrangement of the remaining genes across the genome. In a previous study the pathway swapping concept was developed, i.e. the clustering of all genes from a (metabolic) pathway at one genetic location, such that it can easily be swapped in two steps by an alternative design. Using glycolysis and ethanolic fermentation as proof of concept, the 13 major paralogs remaining in the MG strain were clustered at one genomic location and the genes from the native loci were deleted. However, while the MG strain showed no noticeable change in physiology, the strain with the clustered glycolysis, called SwYG (Switchable Yeast Glycolysis), grew substantially slower than the control strain with a full set of scattered glycolysis and fermentation genes (10-30% slower depending on the measuring conditions). In **Chapter 3** the cause of this decrease in growth rate is investigated. Following a hypothesis-driven approach several potential causes such as: marker usage, glycolytic capacity and change in expression or replication at the cassette integration locus, could be ruled out. Adaptive laboratory evolution and reverse engineering enabled to increase the growth rate of the SwYG strain to the level of the unedited parental strain and discover the causal mutations: *ATG41* and *SUR2*. Both of these genes are directly or indirectly linked to autophagy. Two, possibly intertwined, hypotheses could be formulated. An unwanted side mutation in *VPS15*, a gene also indirectly linked to autophagosome formation, could be responsible for the decreased growth rate. The second scenario entails a potential, yet unexplored moonlighting activity of phosphoglycerate kinase (encoded by *PGK1*) in autophagy, a well characterized function of its human orthologue. While the lower enzymatic capacity of P<sub>gk1</sub> observed in the SwYG strain is probably not negatively affecting the

glycolytic flux, it might affect autophagy, a mechanism in line with the mutation observed in *ATG41* and *SUR2* in the strains evolved from SwYG. This chapter highlights the challenges faced in intensive strain engineering programs.

Considering the importance of CCM for precursor and ATP supply and redox balancing for product formation, cell factories improvement programs have to take into account the whole CCM and to tailor it to reach maximal product titer, rate and yield (TRY). Intensive and extensive remodelling of CCM could be facilitated by the extension of the pathway swapping concept from glycolysis to the entire CCM. However, it means that a large number of genes (76 in case of a minimised CCM set as determined in **Chapter 2**) would have to be centralised at one chromosomal location, with the risk of interfering with the native chromosomal structure, function and expression. In **Chapter 4** the suitability of a supernumerary 17<sup>th</sup> neochromosome is investigated as orthogonal platform to assemble, express and remodel metabolism. This achievement was made possible by the remarkable HR machinery that enables yeast to stitch together DNA molecules with high efficiency and fidelity. Design considerations are important for the construction of such neochromosomes for metabolic engineering approaches. To easily remodel metabolism, the ideal scenario entails the modular assembly, based on an *in silico* design of individual transcriptional units with customized promoters and terminators. In **Chapter 4**, 100 kb test neochromosomes, mostly consisting of non-coding DNA, could be assembled *in vivo* from up to 44 transcriptional-unit-sized DNA fragments with a high efficiency and fidelity. Importantly, neochromosomes were extremely stable in terms of replication and segregation as well as genetic stability upon propagation and sub-culturing. Finally, relocalization of the minimized set of glycolytic and fermentative genes to the neochromosomes demonstrated that they can be used as landing pads for metabolic pathways, and that they can carry sets of heavily transcribed, clustered genes coding for essential cellular processes.

In **Chapter 5**, neochromosomes are further investigated and implemented as tools for metabolic engineering. While the neochromosomes built in **Chapter 4** were circular, in **Chapter 5** it is demonstrated for the first time that 100 kb linear chromosomes can be assembled in a single step *in vivo* from 43 fragments, while simultaneously forming stable telomeres. Despite this success, the assembly of linear chromosomes seemed to proceed less efficiently and faithfully than circular chromosomes. Once assembled the linear and circular chromosomes displayed identical stability and had a similar impact on the host physiology. As proof of principle of the applicability of neochromosomes for extensive engineering of metabolic pathways, the complete set of genes for the formation of the anthocyanin pelargonidin 3-O-glucoside (P3G) from substrate to product were (re)localised to a synthetic chromosome. The biosynthesis of P3G (colorant and nutraceutical) proceeds from glucose, via glycolysis, the pentose

phosphate pathway, the shikimate pathway and the anthocyanin plant pathway. First a total of 42-43 fragments, including a minimized set of the glycolytic and pentose phosphate pathway genes as well as the *Escherichia coli* shikimate pathway, were assembled *in vivo* in one step into circular or linear neochromosomes. These chromosomes served as landing pad for integration of the anthocyanin plant pathway. The final neochromosomes harbouring 41 metabolic genes, were fully functional and enabled the production of P3G in yeast. This successful example of implementation of neochromosomes for metabolic engineering opens the door to neochromosome-assisted large scale genome remodelling. Combining the minimal set of genes for CCM defined in **Chapter 2**, the pathway swapping concept and neochromosomes established in **Chapter 4 and 5**, we can envisage future microbial cell factories with modular genomes in which core metabolic network and processes, localized on satellite, specialized neochromosomes can be swapped for alternative configurations, next to landing pads neochromosomes serving for the addition of (heterologous) functionalities.

# Chapter 1

## General introduction

The section “Synthetic genomics from a yeast perspective “ in the general introduction, is essentially as a manuscript in preparation for submission

## **History of microbial biotechnology**

Biotechnology can be described as “the utilization of biological processes, organisms or systems to produce compounds that are anticipated to improve human lives” (1). Besides the domestication of plants and animals to serve human needs, humans already unknowingly started making use of microorganisms very early in history. Foraging humans in Israel fermented a variety of plants already 13,700 years ago (2), followed by Chinese fermenting a mixture of rice, honey and fruits into wine (7000 BC) (3), Sumerians and Babylonians brewing beer (6000 BC) (4) and finally wine fermentation by ancient Egyptians (3500 BC) (5). Later also a range of non-alcoholic fermentations emerged, such as pickling to preserve crops and the manufacture of cheeses (6). van Leeuwenhoek already observed micro-organisms with his microscope in 1680 and Schwann (7) and Cagniard-Latour (8) observed that alcoholic fermentation was linked to living microorganisms. However, it was not until Pasteur’s work ( $\approx$  1876) on yeast (9) that alcoholic fermentation by microorganisms was unequivocally proven (10). Armed with this knowledge, the period until 1890 was marked by improvement of fermentation processes, such as the use of monocultures and prevention of infection by closed fermentation vessels. Between 1890 and 1940, the microbial industry greatly widened the range of fermentation products, such as the Weizmann process to produce butanol for making synthetic rubber (11), glycerol production to manufacture explosives (12) and sorbose as an intermediate for vitamin C production (10). The discovery of antibiotics and venture to increase antibiotic productivity gave birth to the classical non-targeted microbial strain improvement techniques including strain screening, laboratory evolution and random mutagenesis techniques (13,14). By these methods biomass-specific penicillin production was increased over 1000-fold (10). The real breakthrough in biotechnology was however in the last few decades with the discovery of molecular tools (gel electrophoresis, restriction-ligation procedures, PCR, sanger sequencing, *etc.*) for targeted genetic engineering (1). Finally, the time of modern biotechnology started with the approval of human insulin produced by recombinant DNA technology by the US Food and drug administration, marketed by Genentech and Ely Lilly for treating diabetes (6,10).

## **Metabolic engineering for the development of microbial cell factories**

The development of microbes for biotechnological processes follows one of two general strategies (15). The first strategy is to select a natural producer of the molecule of interest. Using natural producers offer the advantage of the presence of the

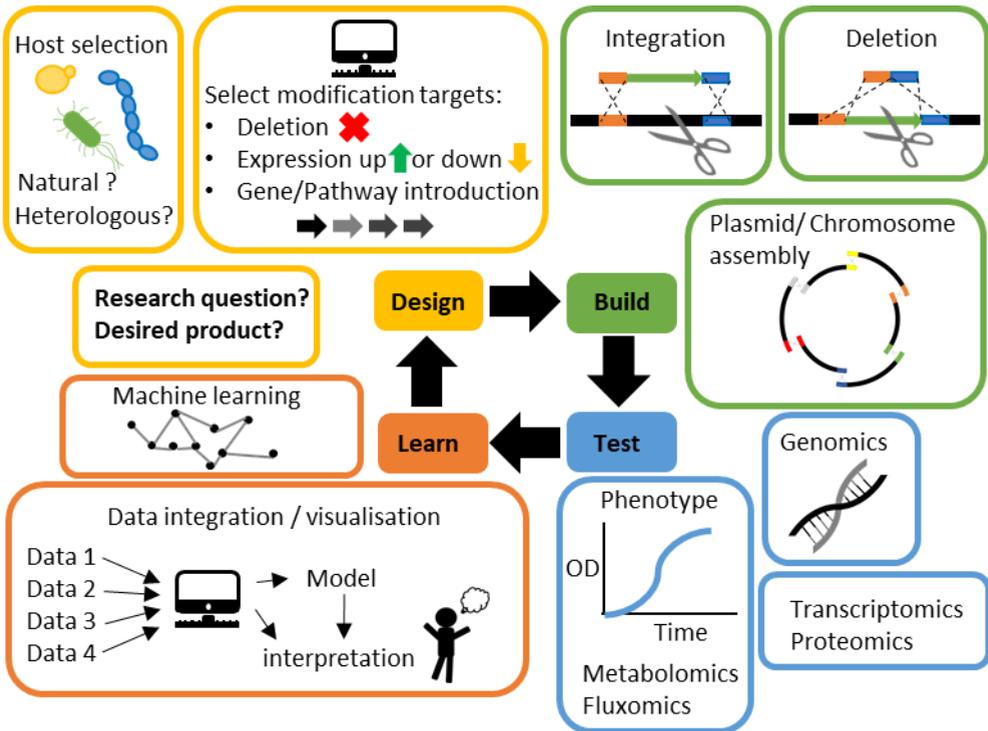
metabolic and secretion pathway of interest, and of innate tolerance to the potential toxicity of the product. Several current biotechnological workhorses such as *S. cerevisiae* for the production of ethanol, *Corynebacteria* for amino acids, *Penicillium chrysogenum* and *Streptomyces* for antibiotics, and *Aspergilli* for enzymes are improved natural producers (16). However, obtaining microbes with economically relevant product yields, titres and robustness presents a major challenge, and the chances of success largely depend on the nature of the microbe. Natural producers might not be tractable and not, or poorly, accessible for performance improvement by targeted genetic engineering. Process improvement is then limited to empiric approaches such as random mutagenesis and laboratory evolution. As a powerful alternative to natural producers, the second strategy involves targeted genetic engineering of well-known, tractable and genetically accessible organisms by non-native production of bio-chemicals. This strategy is part of a branch of biology called metabolic engineering. In metabolic engineering, knowledge of the host genome and biological processes combined with their genetic accessibility enables to improve productivity and yields by eliminating side products, enhancing robustness, and increasing product export or flux in the production pathway. Together with this, the acquisition of new functionalities via heterologous expression in these hosts allows to broaden product and substrate range (17). Heterologous production in microorganisms is key for the industrial production of many relevant compounds of interest that originate from plants (*e.g.* anti-cancer drugs) or mammalian cells (*e.g.* hormones).

Due to their economic interest, several natural producers (such as the ones mentioned above) have undergone decades of intensive research, reaching a level of knowledge and tractability suitable for metabolic engineering and for the development of heterologous production. Other current industrial hosts for heterologous production, such as the gram-negative bacterium *Escherichia coli* (production of diols such as 1,3-propanediol and 1,4-butanediol) or the gram-positive bacterium *Bacillus subtilis* (proteins, fine chemicals), were well-studied for fundamental research and easy to engineer and were thus propelled from laboratory to industrial production (17,18).

*S. cerevisiae* produces native (ethanol, succinate) as well as heterologous products and is one of the best-known industrial cell factories and model organism in fundamental research (15,18). The popularity of *S. cerevisiae* as model organism and cell factory stems from its easy manipulation by genetic engineering, the vast knowledge on its physiology and genomics, and the availability of large scale industrial cultivation processes (18). *S. cerevisiae* is well suited for the production of chemicals and proteins which are natively predominantly produced by eukaryotes since it shares multiple biochemical processes with higher eukaryotes (16,18). For instance, the plant-derived drug precursor artemisinic acid, requires cytochrome P450 enzymes for its synthesis.

Due to the membrane bound nature of these enzymes and the necessary post-translational modifications, they are difficult to express in bacterial hosts, but artemisinic acid has been produced using *S. cerevisiae* (19). Therefore, as cell factory *S. cerevisiae* now produces a wide range of recombinant proteins (e.g. insulin against diabetes, human platelet-derived growth factor against ulcers, and several vaccine proteins to prevent hepatitis A and B) (20), natural plant compounds such as resveratrol, organic acids for utilization as food additives or precursors to polymers (e.g. lactic acid, succinic acid), and fuel additive like bioethanol (21).

Despite several successful examples of microbial production processes, constructing and optimizing microbes for biochemical production remains extremely challenging, particularly as the economic feasibility of the process requires extremely high product yields and productivity. Microbes have been shaped by eons of evolution to survive in ever-changing natural environments, to replicate themselves and not to make products for human convenience. Due to the still incomplete understanding of biological processes, predicting the impact of targeted genetic modifications on product titer/rate/yield (TRY) remains very difficult. To mitigate this lack of predictability, metabolic engineering goes through iterative design, build, test and learn cycles (Figure 1) (22). In the design phase, heterologous genes are selected to be introduced



**Figure 1 - The Design-Build-Test-Learn cycle employed in metabolic engineering.** For each step of the cycle some representative examples are given.

in the host organism, while native genes are chosen for their increase or decrease in expression. The following build step involves a number of genetic interventions on the host cell: integration of heterologous genes under control of chosen promoters and terminators; deletion of native genes and replacement of native promoters with alternative promoters. With the test phase, a battery of (omics) analytical techniques is implemented to evaluate the impact of the genetic changes on the host physiology and performance. Finally, these acquired data are critically analyzed in the learn phase leading again to a new design, thereby stepping into a new cycle (22).

For each step of the metabolic engineering cycle a broad range of powerful tools is available and is rapidly increasing in number. For the design phase, the choice of the best host-product combination is key but remains largely empiric, as can be seen by the multiplicity of studies attempting to produce the same chemical using a variety of prokaryotic or eukaryotic hosts. A good example is the production of succinic acid that has been attempted with over 15 different prokaryotic and eukaryotic organisms. Each host has its specific advantages and disadvantages. For succinic acid production, some organisms such as *S. cerevisiae* have the advantage to grow at low pH but do not naturally produce succinic acid. Other organisms such as *Mannheimia succiniciproducens* are natural succinic acid producers but are pH sensitive and harbor several auxotrophies (23). Some studies prefer to start with several hosts, focusing on the most promising as the research progresses. For example, the production of artemisinic acid was simultaneously attempted in both *E.coli* and *S. cerevisiae*, where the latter was eventually chosen as a more suitable host (19). The field would greatly benefit from the availability of databases with characteristics of starting material, products and organisms accompanied by corresponding software (*i.e.* BioCAD) such that the most suitable organism can be chosen for a particular conversion (22). Genome-scale metabolic models (GEMs) are becoming available for an increasing number of industrial microbes, and several tools have been developed to aid in selecting the most promising engineering targets. With flux balance analysis, GEMs can be used to model all reactions of the cell and link them to the enzymes catalyzing these reactions and their corresponding genes (24,25). There are now even models which integrate enzyme constraints and protein 3D structures such as ecYeast8 and proYeast8<sup>DB</sup> for *S. cerevisiae* (18,26).

In the build phase, heterologous product formation requires introduction of one to several heterologous genes. For example, introduction of 31 (heterologous) genes were necessary for production of noscapine in *S. cerevisiae* (27). In addition, a screening for the highest product titers is often performed comparing heterologous enzymes from different organisms catalyzing the same enzymatic step (28). Introduction or screening of this many (heterologous) genes requires large scale and automated synthesis and assembly of DNA. Moreover, extensive engineering of the

host's native pathways and cellular processes is required to redirect flux towards the product or production and secretion of protein. With the development of different CRISPR-Cas systems (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins) it has become possible to simultaneously, rapidly and efficiently edit multiple targets (*e.g.* up until nine simultaneous deletions at five individual loci in yeast) in an organism (29,30). Due to the lack of predictability of genetic modification, frequently multiple engineering strategies are tested in parallel (22). Therefore, CRISPR-Cas has to be optimized for efficient multiplex strain construction and made compatible with high-throughput strain construction using liquid handling robots.

For many metabolic engineering efforts, constructed strains are only evaluated on product titers during the test phase. While product titers provide the most industrial relevant information, it overlooks the underlying mechanism of why a strain is performing better, especially if multiple simultaneous modifications have been applied. Deep characterization of a strain can be performed using several -omics techniques. This in-depth knowledge can aid in the subsequent design phase and might thus decrease the number of metabolic engineering cycles. Using genomics technologies such as short read sequencing and long read nanopore sequencing, it can be verified if the genetic modification of the build phase were successful. A combination of transcriptomics, proteomics, metabolomics and fluxomics together with growth characteristics of the strain should give an elaborate view on how the genetic modifications have affected the strain (31).

In the learn cycle, it is important to have software that can integrate all data gathered in the test phase and process it in a format, such that comprehensible conclusions on the effect of the genetic modifications can be drawn. Moreover, this software should be able to suggest new engineering targets for the next design-build-test-learn cycle, effectively uniting the learn and design phases. Although used scarcely up-to-date, machine-learning algorithms will probably have an important role to play (22).

## **Synthetic Biology**

Engineering disciplines are characterized by concepts as standardization, modularity, hierarchy and predictability (32). For example, electronic engineering relies on several standardized parts such as resistors, capacitors, wires, switches and outputs like lamps. These can be modularly assembled into a range of possible circuits, each with a predictable and predefined output. Synthetic biology has been defined as “the design and construction of new biological systems that do not exist in nature through the assembly of well-characterized, standardized and reusable components” (33). Effectively, this sub-branch of metabolic engineering, aims to apply engineering

principles, like abstraction and standardization of parts, to the fundamental components of biology (34,35). Hereby desired biological systems possibly producing interesting biochemicals are built in a hierarchical fashion starting from parts (promoters, ribosome binding sites (RBS), open reading frames (ORFS) and terminators), which are clustered into circuits executing a certain function and which could then be assembled into entire genomes (35,36).

Most extensive research has focused on development of parts and their use in metabolic engineering (36). An important aspect in metabolic engineering is the tuning of expression level of genes to obtain an optimum flux. In *S. cerevisiae*, over 900 promoters have been cloned and characterized on expression level under various conditions (37). In addition, yeast synthetic promoters have been engineered by modifying elements in native promoters (38), adapting nucleosome architecture (39), constructing hybrid promoters (40) or using genetic logic gate controllers (responding to environmental or cellular stimuli) (36). Moreover, CRISPR technologies have been used to tune gene expression by fusing activator or repressor proteins to a catalytically inactive Cas9 (dCas9) targeted to a native promoter (41). Similarly, gene expression can be tuned by fusing dCas9 to methyltransferases or dioxygenases to change the methylation profile of DNA and thereby expression, as has been demonstrated in mammalian cells (42). Expression levels can also be tuned by choice of terminator, for which several synthetic variants have been studied in *S. cerevisiae* (36,43). At the mRNA level, a heterologous RNA interference (RNAi) machinery has been introduced in *S. cerevisiae* which does not natively carry this system (35,44). Finally, parts for translational control have been only scarcely studied (36). Most notable are the riboswitches which are mRNA structures in the 5'UTR region of an mRNA which can form a particular conformation (aptamer) able to bind ligands. Upon formation of the ligand-aptamer complex the ribosome is unable to bind or progress (36).

Functional parts such as promoters, ORFs and terminators can be further assembled into devices or circuits to achieve more complex functions. One of the first examples applying this on a large scale is the Biobricks system in *E.coli* (45). A Biobrick part is "a basic unit with an indivisible biological function" (46). In the Biobricks system, these parts are flanked by DNA sequences containing enzymatic restriction sites. All of these parts are stored in a publicly accessible database (<http://partsregistry.org>). Using classical restriction-ligation techniques the scientific community can use each other's Biobrick parts to assemble pathways and devices that perform a specific function. In addition to the Biobricks system there is now a large range of *in vitro* but also *in vivo* techniques, which can be used to assemble not only devices and circuits but even whole chromosomes and genomes.

# Synthetic genomics from a yeast perspective

## A general introduction to synthetic genomics

Synthetic Genomics (SG) is a recent Synthetic Biology discipline that focuses on the construction of rationally designed chromosomes and genomes. SG offers a novel approach to address fundamental biological questions by restructuring, recoding, and minimizing (parts of) genomes (as recently reviewed by (47)). SG, purely academic until recently, is now spurring technological developments in both academia and industry (e.g. (48)). Humankind's best microbial friend, the baker's yeast *S. cerevisiae*, has played, and continues to play a key role in SG advances, both by enabling the construction of chromosomes for other hosts, and in the refactoring of its own genome. Here the reasons for this strategic positioning of *S. cerevisiae* in SG is explored, the main achievements enabled by this yeast are surveyed and future developments are reflected upon.

## Construction of chromosomes and genomes

While small-sized viral chromosomes were the first to be chemically synthesized, the breakthrough in the field of SG came with the synthesis and assembly of the 592 kilobase (kb) chromosome of *Mycoplasma genitalium* (49,50). The unicellular eukaryote *S. cerevisiae* has made a key contribution to this famous milestone. To understand how this microbe, commonly used in food and beverages, contributes to the assembly of synthetic genomes, let us recapitulate how synthetic chromosomes can be constructed (Figure 2).

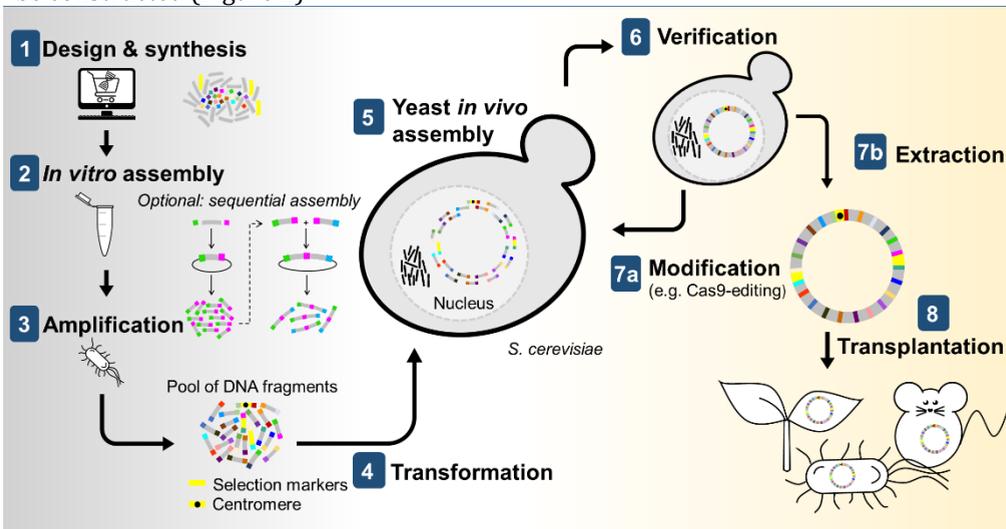


Figure 2 - Simplified overview of chromosome construction using *S. cerevisiae* for genome assembly and production.

It starts with the customized synthesis of short DNA molecules called oligonucleotides. Oligonucleotides are mostly synthesized via phosphoramidite chemistry, a 40 year-old method (51) that, despite decades of technological developments, struggles to deliver error-free oligonucleotides longer than 200 base pairs (bp). While the implementation of microarrays has substantially decreased the synthesis cost, it has not increased oligo length, an achievement that requires new synthesis methods (52). Enzymatic alternatives for DNA synthesis are under development (53,54), but still have considerable shortcomings regarding automation and scalability that must be overcome before commercial scale can be considered (reviewed in (55-58)). Considering that a theoretical minimal genome would be around 113 kb long (59) and that the first fully synthesized genome of *M. genitalium* contains about 583 kb (49), thousands of oligos must be stitched together to construct a complete synthetic genome. These DNA oligos can be assembled into longer DNA fragments owing to a plethora of *in vitro* methods (reviewed in (56,60,61)). A method that has gained tremendous popularity since its development is the homology-based Gibson isothermal assembly (62), devised to assemble the *M. genitalium* genome. As all *in vitro* methods, Gibson assembly is limited by the number of fragments that can reliably be stitched together in one reaction, usually around a dozen, requiring a stepwise assembly procedure of increasingly large genomic DNA constructs (63). DNA must be recovered from the reaction, amplified and verified in each round, to allow further processing. Selection and amplification of correctly cloned DNA is routinely performed in *E. coli*, however, maintenance of large constructs of exogenous DNA, especially from prokaryotic origins, in this bacterium is often limited by expression and toxicity of gene products (64). *In vitro* alternatives for efficient and faithful selection and amplification of correctly assembled DNA are under development, but these are currently limited in length of amplified DNA and scalability (65-68). While in principle stepwise *in vitro* assembly can lead to a DNA molecule of any size, and selection and amplification in *E. coli* worked well for DNA constructs up to 72 kb, *E. coli* had great difficulties maintaining quarter *M. genitalium* genomes, causing Gibson and colleagues to turn to baker's yeast (49,50).

*S. cerevisiae* seems a logical host for SG as it naturally maintains a 12 Mb genome consisting of 16 chromosomes ranging from 230 to 1500 kb in its haploid version, lives as polyploid in natural environments, and is extremely robust to changes in genome content and architecture (69). The extreme robustness of *S. cerevisiae* to supernumerary, chimeric chromosomes, a key feature for SG, was already demonstrated in the late '80s (70,71). A second key feature of *S. cerevisiae* is its preference for homologous recombination (HR) to repair double-strand DNA breaks, a rare trait among eukaryotes (72). *S. cerevisiae* ability to efficiently and with high fidelity stitch together linear DNA molecules that present homologous regions as short

as 40 bp (73) at their ends, was rapidly valorized for genetic manipulations and assembly of heterologous DNA. Recently renamed *in vivo* (or *in yeasto*) assembly, this cloning technique (Figure 2) contributes to the remarkable genetic tractability and popularity of *S. cerevisiae* as model and industrial microbe (62,74). The combination of *S. cerevisiae*'s HR efficiency and fidelity, chromosome maintenance and propagation enabled the construction of the full *Mycoplasma* genome. Reflecting that “*in the future, it may be advantageous to make greater use of yeast recombination to assemble chromosomes*”, this study propelled *S. cerevisiae* as powerful ‘genome foundry’ (49). In the challenge to synthesize genomes, Ostrov and colleagues rightfully identified assembly and introduction of these long DNA constructs as ‘the most critical hurdle’ (55). To date, *S. cerevisiae* has been key to assembling entire or partial genomes in most synthetic genome projects (Table 1).

**Table 1 - Overview of the contribution of *S. cerevisiae* in the assembly of microbial and viral genomes**

Chromosome/ Genome	Number of transformed fragments <sup>1</sup>	Size of transformed fragments <sup>1</sup>	Size of final construct	DNA isolated / transplanted	Refer- ence
<b>Herpes simplex virus</b>	11	14 kb	152 kb	Isolated, virus reconstituted	(75)
<b>SARS-Cov-2</b>	20	Up to 12 kb	30.5 kb	Isolated, virus reconstituted	(76)
<b><i>Mycoplasma genitalium</i></b>	25	17-35 kb	592 kb	N/A	(50)
<b><i>Mycoplasma genitalium</i></b>	6	Up to 144 kb	592 kb	N/A	(49)
<b><i>Mycoplasma mycoides</i></b>	11	100 kb	1 Mb	Isolated, transplanted to <i>M. capricolum</i>	(77)
<b><i>Prechlorococcus marinus</i></b>	2	580-675 kb	1.66 Mb	Isolated	(78)
<b><i>Synechococcus elongatus</i></b>	4	100 - 200 kb	454 kb	Isolated	(79)
<b><i>Acholeplasma Laidlawii</i></b>	3 <sup>2</sup>	121-897 kb	1.38 Mb	N/A	(80)

<b><i>Escherichia coli</i></b>	7-14	6-13 kb	100 kb	Isolated, transplanted in <i>E. coli</i>	(81)
<b><i>Caulobacter crescentus</i></b>	16	38-65 kb	785 kb	N/A	(82)
<b><i>Phaeodactylum tricornerutum</i></b>	5	106-128 kb	497 kb	N/A	(83)
<b><i>Chlamydomonas reinhardtii</i> chloroplast genome</b>	6	34-129 kb	230 kb	Isolated, transplanted to <i>C. reinhardtii</i>	(84)
<b>Yeast chromosome XII</b>	33 <sup>3</sup>	26-39 kb	976 kb	N/A	(85)
<b>Single-chromosome yeast</b>	15 <sup>3</sup>	230-1500 kb	11 Mb	N/A	(69)
<b>Human <i>HPRT1</i> gene</b>	13	3-83 kb	125 kb	Isolated, transplanted to mouse embryonic stem cells	(86)
<b>Human <i>HPRT1</i> gene</b>	29	3 kb	82 kb	N/A	(86)

<sup>1</sup> In case of a sequential assembly, the fragment number and size of the last assembly is used.

<sup>2</sup> Initial assembly of the entire genome failed due to gene toxicity.

<sup>3</sup> Assembly was performed by stepwise integration in multiple rounds.

N/A – Not applicable

For instance, the entire 785 kb refactored *Caulobacter crescentus* (renamed *C. ethensis*) genome was assembled *in vivo* from 16 fragments (82), while the recoded *E. coli* genome was split over 10 fragments of 91 to 136 kb individually assembled *in vivo*, and then sequentially integrated in the *E. coli* chromosome to replace native segments (Table 1) (81). In *yeasto* assembly is not limited to prokaryotic genomes; the rapid and faithful HR-based assembly (Figure 2) of *S. cerevisiae* recently enabled the reconstruction of a synthetic SARS-CoV-2 genome in a single week (76). More recently, it was used to assemble and transplant a 101 kb human gene into mouse embryonic cells (86). Additionally, *S. cerevisiae* was logically selected for the construction of the first synthetic eukaryotic genome. The international Sc2.0 consortium, spearheaded by Jef Boeke, undertook less than ten years ago the daunting task of synthesizing recoded versions of the 16 yeast chromosomes. Via stepwise, systematic replacement of 30 to 40 kb (using ca. 12 DNA fragments of 2 to 4 kb) of the native yeast sequence, the

consortium is close to the completion of the largest synthetic genome to date (87,88), with the ambition to reshape and minimize the *S. cerevisiae* genome (89).

While *S. cerevisiae* is not the only microbial host available for the construction of (neo)chromosomes (see Table 2), several key features make it superior to its bacterial alternatives *B. subtilis* and *E. coli* as genome foundry: i) *S. cerevisiae* has the natural ability to carry large amounts of DNA and therefore to host multiple exogenous bacterial genomes (90); ii) *E. coli* frequently struggles with toxicity caused by the expression of exogenous bacterial sequences (50,64,91), while *S. cerevisiae* is very resistant to the presence of heterologous DNA from prokaryotic or eukaryotic origin (78); iii) *S. cerevisiae* can, in a single transformation, assemble many DNA oligonucleotides into (partial)genomes. *B. subtilis* can also maintain large exogenous DNA constructs, but requires a stepwise method for DNA assembly, in which each DNA part is integrated sequentially into *B. subtilis* genome (92). This approach is intrinsically more labor-intensive and time-consuming than *S. cerevisiae* single transformation assembly.

Surprised by *S. cerevisiae*'s genetic tractability, Gibson and colleagues wondered “*how many pieces can be assembled in yeast in a single step?*” (49). In **Chapter 4** of this thesis, this limit was probed by assembling large, modular synthetic chromosomes from many transcriptional-unit-sized DNA fragments. Once assembled, synthetic chromosomes can be easily adapted in *S. cerevisiae* thanks to the powerful combination of its HR-capacity and toolbox rich in efficient (CRISPR-based) DNA-editing tools.

While *S. cerevisiae* is natively proficient for SG, several aspects of ‘*in yeasto*’ assembly are still far from optimal. Firstly, compared to bacterial genome foundries, *S. cerevisiae* cells grow slowly with a maximum specific growth rate around 0.4 - 0.5 h<sup>-1</sup> and are hard to disrupt due to their sturdy cell wall. Considering that large DNA constructs above a few hundred kilobases are sensitive to shear stress, chromosome extraction and purification from *S. cerevisiae* is possible, but remains tenuous and inefficient, leading to low DNA yields and potentially damaged chromosomes (93). Secondly, the strength of *S. cerevisiae* can become its weakness, as the HR machinery can be overzealous and recombine any (short) DNA sequence with homology within or between the (neo)chromosomes, which may lead to misassemblies. Lastly, non-homologous end joining and microhomology-mediated end joining, DNA repair mechanisms that assemble pieces of DNA with no or minimal homology, are present in *S. cerevisiae* with low activity (94,95), and can also cause misassemblies. Similar to how *E. coli* was engineered to become a lab tool for DNA amplification, these shortcomings could be alleviated by engineering *S. cerevisiae* into a more powerful genome foundry.

Are there future alternatives to *S. cerevisiae*? Naturally, *B. subtilis* and *E. coli* could also be engineered. However, considering the minute fraction of the vast microbial biodiversity that has been tested for genetic accessibility and DNA assembly, it is likely that microbes yet to be discovered are even better genome foundries. Environments causing extreme DNA damage (high radiation, toxic chemicals, *etc.*) might be a source of HR-proficient organisms (*e.g.* (96,97)) better suited for SG. The choice of host can also depend on the (industrial) applicability of the organism. For example, *Yarrowia lipolytica*, a promising industrial workhorse due to its wide substrate- and product range, was recently used to assemble a synthetic chromosome, enabling pathway engineering in this genetically inaccessible yeast (98).

**Table 2 - Strengths and weaknesses of *in vitro* and *in vivo* assembly methods**

	<i>In vitro</i>	<i>E. coli</i>	<i>B. subtilis</i> <sup>1</sup>	<i>S. cerevisiae</i>
				
<b>One-pot assembly</b>				
Number of fragments	Up to 20 (99)	Up to 6 (100)	2 (101)	Up to 29 (86)
Construct size	Up to 144 kb (49)	Up to ca. 180 kb (100)	Up to 200 kb 39	Up to 1.66 Mb (78)
<b>Sequential assembly</b>				
Number of assembled fragments	600 (63) <sup>2</sup>	-	31 (101)	1078 (77)
Number of sequential HR events	4	-	18	3
Final construct size	16.3 kb	-	134.5 kb	1.08 Mb
<b>Other factors</b>				
Compatibility with heterologous genes	***	*	***	***
DNA yield	*	***	Unknown	**
<b>Overall proficiency to assemble synthetic genomes</b>	- <sup>3</sup>	*	**	***

<sup>1</sup> Assembly of fragments in *B. subtilis* is performed by integration into the host genome.

<sup>2</sup> Between rounds of sequential assembly, transformation into *E. coli* is conventional for selection and amplification of constructs.

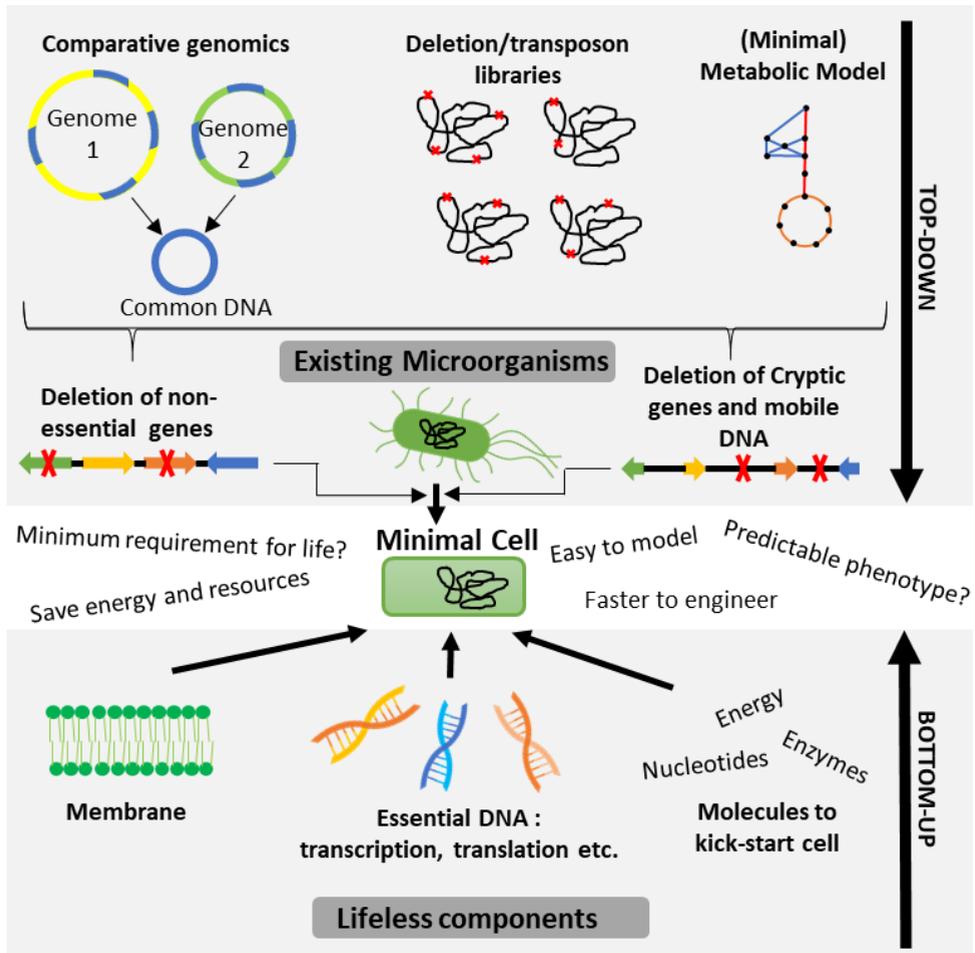
<sup>3</sup> Requires *in vivo* amplification and selection in a microbial host

In a more distant future, *in vitro* alternatives might replace the need for live DNA foundries altogether, thereby accelerating and simplifying genome construction. However, this will require major technological advances in *in vitro* DNA assembly and amplification. Already substantial efforts have led to the development of methods for DNA amplification, such as rolling circle amplification by the phage  $\phi$ 29 DNA polymerase (102), recently implemented for the amplification of a 116 kb multipartite genome (65) and the *in vitro* amplification of synthetic genomes using the *E. coli* replisome, which already demonstrated to be capable of amplification of 1Mb synthetic genomes (67). Targets for improvement of these methods are the maximal length of amplified DNA fragments, the yield of amplification, the need for restriction of the amplified, concatenated molecules or the formation of non-specifically amplified products. The development of an *in vitro* approach that can parallel *S. cerevisiae in vivo* assembly capability seems even more challenging. While an interesting avenue might be to transplant *S. cerevisiae* HR DNA repair *in vitro*, it presents a daunting task considering that all players and their respective role have not been fully elucidated yet (95,103). Still, considering that highly complex systems such as the transcription and translation machineries have been successfully implemented *in vitro* and are commercially available (104), *in vitro S. cerevisiae* HR might become a reality in the coming years.

### **Outlook for synthetic genomics: building of bottom-up synthetic cells**

Since the first genome synthesis in 2008, relatively few genomes have been synthesized, despite remarkable technological advances. Low-cost, customizable construction of designer genomes, accessible for small viral, organellar or bacterial constructs, is still out of reach for large (eukaryotic) genomes. Looking further ahead, SG is anticipated to assist in the development of new cell therapies and other medical applications, which is the ambition of the Genome Project-Write (105). Another future application of SG is the construction of bottom-up synthetic cells to address both fundamental questions, as well as applied challenges (Figure 3). There are several research consortia studying bottom-up synthetic cell biology including BaSyC (<http://www.basyc.nl>), Build-a-cell (<http://buildacell.io>), MaxSynBio (<https://www.maxsynbio.mpg.de>) and the Synthetic cell initiative (<http://www.syntheticcell.eu>) (106). Although a bottom-up synthetic cell has not been made yet, several achievements have been made in the three main features necessary for performing the essential functions of life: i) the engineering of membranes/microenvironments able to enclose and protect cell constituents and allow exchange of molecules and energy; ii) the restructuring of macromolecules such as DNA or RNA that contain genetic information used for dictating the function and dynamics of the cell; iii) the remodeling of metabolic pathways to provide energy and constituents of the cell to self-maintain, self-renew and self-process information

(107,108). These synthetic cells can answer questions such as: where did life come from? How do lifeless components interact to form life? Moreover, gained understanding in biology, chemistry and physics necessary for building a minimal synthetic cell that is alive, could help in building applied synthetic cells that make human-desired products. We can envision a future in which modularly assembled synthetic cells, are specifically built to execute a desired biotechnological function.



**Figure 3 - Comparison between a top-down and a bottom-up approach for constructing minimal cells.** In a top-down approach a maximum of DNA is removed from an existing cell until no more can be removed, thus resulting in a minimal cell. Sometimes instead of deleting DNA in the native organism a synthetic minimal genome is assembled based on the native genome and transplanted in a living recipient cell. In a bottom-up approach life-less components are (step-by-step) assembled into ever more complex functioning systems until a viable minimal cell is obtained.

## Top-down approaches to minimal genomes

There are still numerous technical, financial, and computational limitations, as well as a lack in fundamental knowledge on the absolute requirements of life, that must be overcome towards the future prospect of bottom-up synthetic cells, tailored to applications in biobased industry. Therefore, for now, study and production of relevant chemicals and pharmaceuticals is still based on existing microorganisms. However, these microorganisms possess complex, mosaic genomes that have evolved to survive in natural environments. The mosaic architecture of genomes limits genetic accessibility, since multiple loci have to be targeted to alter a functionality. In addition, the complexity of microbial genomes limits our fundamental understanding and thus makes it more difficult to engineer microorganisms for biotechnological purposes. Moreover, micro-organisms can undergo both smaller-scale duplications (SSD) or whole genome duplications (WGD), thus introducing paralogous genes. Although these gene duplicates might confer a selective advantage by neofunctionalization, sub-functionalization or gene dosage effect, this might be only under very specific growth conditions (109). These acquired functions might not be relevant for specialized cell factories, cultured under specific conditions. In microbial cell factories, it is desired to have cellular energy and resources optimally allocated to perform a specific biotechnological process. Reducing genetic content by eliminating redundant genes and non-essential DNA, can improve understanding of the platform organism as well as cause enhanced genetic accessibility and reduce cellular expenses as illustrated by **Chapter 2** of this thesis (110).

### Benefits and applications of minimal genomes

Genome reduction can be defined as: “the repeated deletion of irrelevant genes...” or other DNA “...by methods of genetic engineering with the purpose of constructing a functionalized cell for a selected application” (110). One application of genome reduction is to gain fundamental understanding on the minimal number of genes that could sustain (self-replicating) life. On the other hand, genome reduction could also have benefits for biotechnological applications. First, genome reduction could save energy and carbon, invested in RNA and protein synthesis, unnecessary for a particular biotechnological process (110-112). Many platform microorganisms are generalists, which can adapt to changing environments by expression of genes that are not actually used in the laboratory environment. A study in *E.coli* suggests that up to 13% of proteins are only expressed to adapt to fluctuations in growth conditions and are thus described as the “stand-by proteome” (113). For biotechnological processes, specialized organisms are required to grow in a controlled environment in order to produce a specific product. Even though large reactors fluctuate in dissolved oxygen, pH and nutrient availability (114), thus necessitating some of this “stand-by

proteome”, still a large number of genes could be saved upon. In this way, energy and resources can be better allocated to the desired biotechnological product.

Removal of genetic content can also facilitate genetic engineering. For instance, the effort to remove the rarest stop codon in *E. coli* by Lajoie *et al.* (115) would have taken much less time, if there would have been less genes to edit. In addition, the complexity of natural microorganisms necessitates the design-build-test-learn cycle adopted in metabolic engineering. Minimal genomes are easier to model (116,117) and might consequently result in more predictable phenotypes. The use of minimal genomes could hypothetically therefore result in less metabolic engineering cycles.

Finally, several theoretical advantages of reducing genome content can be envisioned (110). On a DNA level, an estimated 1.5% of total ATP is invested in DNA synthesis (110-112) and reduction of DNA content could decrease ATP cost, which could be potentially invested in other processes. Furthermore, less DNA to replicate could possibly result in faster doubling times of the organism. The real impact that can be achieved by genetic reduction on ATP cost and growth rate might be small and, until now there have not been any experimental evidence to support either of these hypotheses.

### **Top-down approaches with reduction of native genomes**

Genome reductions efforts have been made in different microorganisms, including *E.coli*, *B. subtilis*, *C. glutamicum*, *Pseudomonas* species, *Lactococcus lactis*, *Streptomyces* species, *Schizosaccharomyces pombe*, and *Saccharomyces cerevisiae* (see for instance (110), (118), (119), (120), (121), (122), (123) (124), (125)). In these studies, two types of DNA sequences were targeted: cryptic genes and mobile DNA such as prophages and insertion elements (119-121), and irrelevant/non-essential genes (118,124,125) such as the flagellar biosynthesis machinery (121). The removal of cryptic genes and mobile elements showed clear advantages, such as an increased genome stability (121). Genome reduction efforts by deletion of irrelevant/non-essential genes often led to changes in phenotype like decreased growth rate (118,119,124,125). The selection of deletion targets can be executed by a variety of methods including comparative genomics, deletion libraries, transposon mutagenesis libraries and GEMs (Figure 3) (110). One example of a genome reduction resulting in quantifiable benefits was the deletion of 4.3% of genomic content from *Pseudomonas putida* KT2440 (121). This genome reduction caused an increase in maximum specific growth rate and biomass yield in minimal medium, a reduced maintenance coefficient, an increased plasmid stability and an increase in specific heterologous protein production.

A single study has so far been dedicated to specifically reduce genome content in *S. cerevisiae* on large scale. Marakami *et al.* (125) deleted 15 terminal chromosomal

regions of *S. cerevisiae* SH5209, which were predicted to be non-essential for growth and accounted up to 5% of the genome content (531.5 kb containing 247 ORFs). Despite many drawbacks (*i.e.* reduced mitochondrial function, reduced growth rate and a differential gene expression pattern), the authors managed to obtain a 1.8-fold increase in ethanol production. In addition, the study of Shao *et al.* (69), in which they fused all 16 chromosomes to create one circular chromosome, resulted in a genome reduction of 9% by removal of telomeres, long repeats and centromeres. Although under standard laboratory conditions the growth rate was only slightly reduced and minimal changes in transcriptome were observed, testing under a wider variety of growth conditions revealed more significant reduction in growth rate.

Interesting, but little used, targets for genome reduction in *S. cerevisiae* are its paralogous genes resulting from the WGD and SSD events. *S. cerevisiae* underwent a whole genome duplication (WGD) approximately 100 million years ago (109). Although approximately 90% of these duplicates were lost during evolution, 551 gene pairs remain. In an effort to reduce the genetic complexity and thereby study sugar transport Wieczorke *et al.* (126) engineered a *S. cerevisiae* strain (Hxt0) in which all 21 hexose transporters were deleted. Unfortunately, the use of the *LoxP/cre* system resulted in large chromosomal rearrangement. Wijsman *et al.* (30), making use of the CRISPR technology and taking advantage of the homologies between *HXT* genes, reconstructed the Hxt0 strain with no chromosomal rearrangement and only a few SNPs. In order to make strain engineering more efficient as well as to ameliorate the understanding of the glycolytic pathway, Solis-Escalante *et al.* (127) deleted the 13 minor paralogs of glycolysis and alcoholic fermentation. Surprisingly these deletions did not lead to any change in phenotype on a wide range of growth conditions including different carbon sources and stress conditions. In **Chapter 2** of this thesis, redundant (paralogous) genes were studied further as a source for genome reduction and to ameliorate genetic engineering.

There are several drawbacks to reducing genetic complexity of native genomes. The more DNA or genes that needs to be deleted the more advanced the genome editing tools need to be in terms of accuracy and multiplexing. Otherwise, a higher number of transformation rounds are required to reach the desired number of correct alterations. Each transformation round and inherent propagation of the strain can lead to undesired mutations as investigated in **Chapter 3** of this thesis. An alternative to reduction of native genomes is the chemical synthesis of reduced synthetic genomes, discussed in the following subsections.

### **Top-down approaches with synthetic genomes**

Craig Venter and coworkers demonstrated that synthetic genomes, almost identical to *M. genitalium* (580 kb) and *Mycoplasma mycoides* JCVI-Syn1.0 (1.08 mb) could be

assembled by *in vitro* techniques and *in vivo* assembly in yeast (49,50,77). Transplantation of the synthetic genome JCVI-Syn1.0 in a *M. capricolum* recipient cell that was forced to lose its own genome, resulted in a synthetic cell (77). In pursuit of a minimal genome, the same group (128) used several approaches to design an *in silico* minimal genome. First, by using transposon mutagenesis of *M. genitalium*, essential, quasi-essential and non-essential genes were identified. In addition, comparative genomics of *Haemophilus influenzae* and *M. genitalium* together with biological knowledge on proteins necessary to sustain life, revealed a possible core set of genes. This *in silico* minimal synthetic genome, referred to as JCVI-syn3.0, was synthesized, assembled and transplanted similarly to JCVI-syn1.0 efforts. Interestingly, out of the 473 genes in JCVI-syn3.0, the cellular function of 149 was unknown, highlighting the gaps in fundamental biological knowledge. This smallest (synthetic) self-replicating organism had a three times increased doubling time as compared to its *M. genitalium* 'parent' (128).

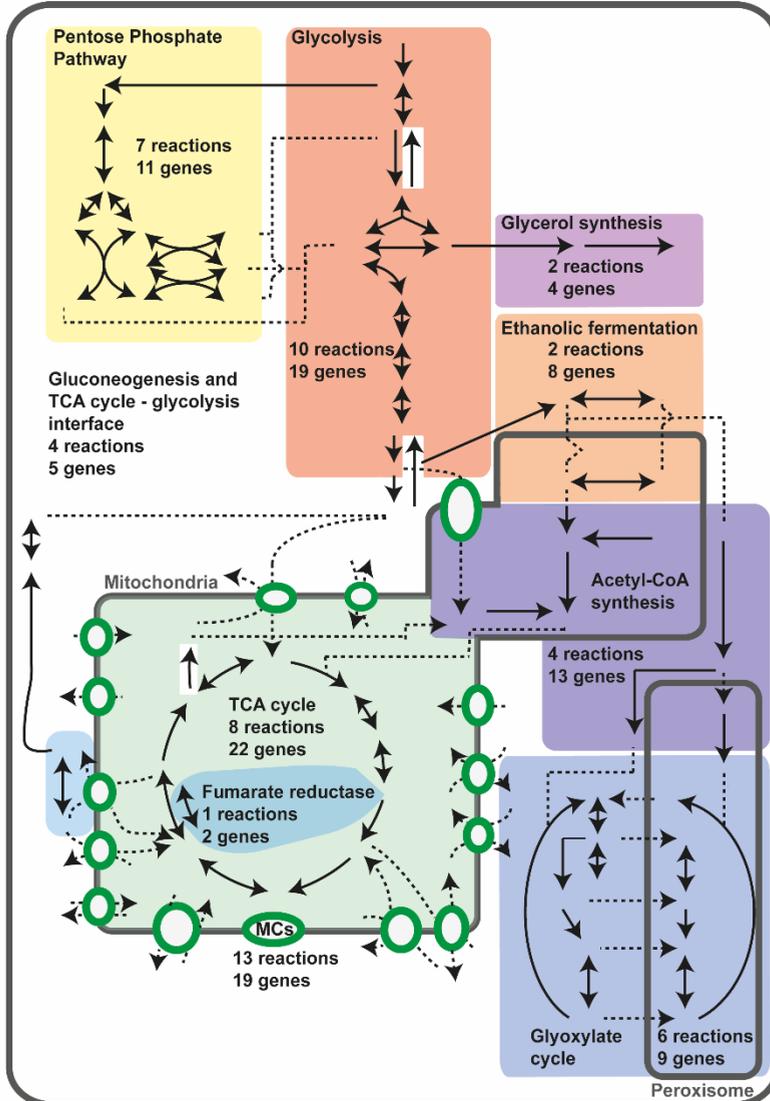
The largest scale effort of making a synthetic eukaryotic genome is the Sc2.0 project, discussed earlier (129), in which multiple research groups worldwide collectively synthesize and assemble a synthetic *S. cerevisiae* genome. The synthetic genome is nearly identical to the *S. cerevisiae* template. However, some changes were introduced to promote a more stable genome including removal of Ty elements, Long Terminal Repeats (LTR) and introns. Although the aim of this project was not genome reduction, these changes resulted in a total genome reduction of approximately 8%. While this type of genetic reduction of non-coding regions may increase genome stability and possibly decrease ATP expenses, these alterations by themselves do not really facilitate metabolic engineering. However, other changes that could facilitate metabolic engineering, such as addition of LoxP sites for inducible evolution (called SCRAMBLE) and the removal of the TAG stop codon for utilization in an orthogonal system, were also introduced (129).

## **A radically new approach to pathway engineering in *S. cerevisiae***

The Sc2.0 project offers a unique possibility at studying yeast genome architecture and the SCRAMBLE system allows for selecting interesting phenotypes by inducible evolution. Contrary to the discovery-driven approach of engineering *S. cerevisiae* for biotech products by the Sc2.0 SCRAMBLE system, this section discusses alternative approaches based on efficient modular pathway engineering.

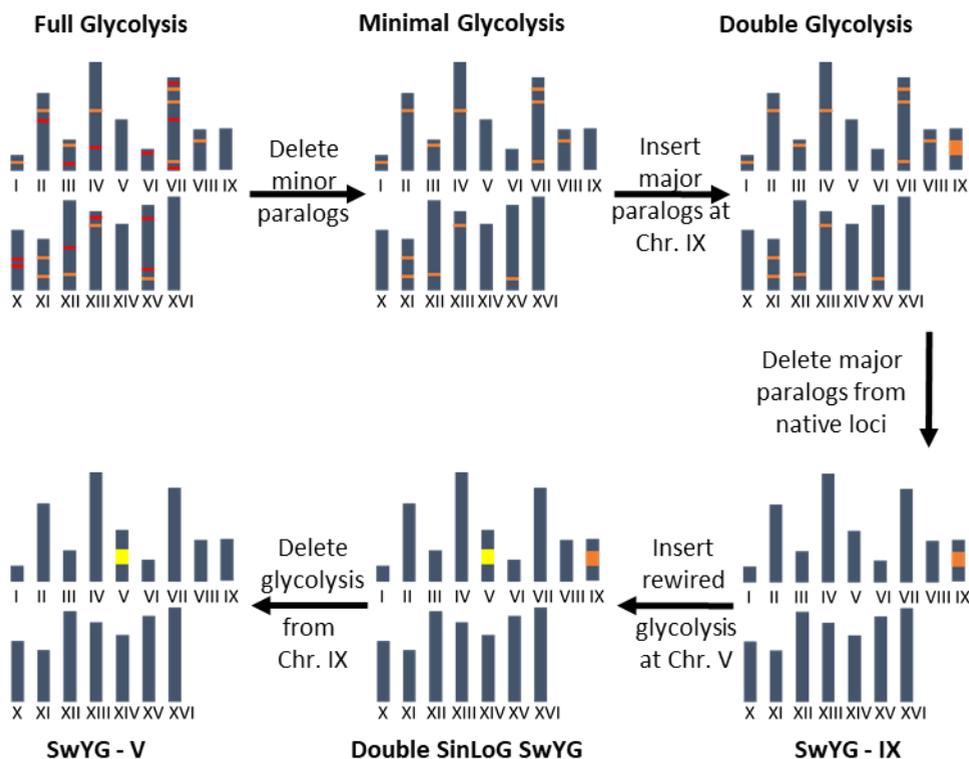
Any product will depend on the host core machinery for energy conservation, co-factor balancing and supply of relevant precursors. Construction of microbial cell factories with economically relevant TRYs therefore entails intensive and extensive engineering of these core processes. The central carbon metabolism (CCM), as considered in

**Chapter 2** (excluding reserve carbohydrates and plasma membrane transporters), encompasses circa 53 enzymatic reactions (glycolysis, gluconeogenesis, ethanolic fermentation, pentose phosphate pathway, tricarboxylic acid cycle, anaplerotic reactions mitochondrial carriers, fumarate reductases, glyoxylate cycle, acetyl-CoA synthesis, glycerol synthesis, gluconeogenesis), encoded by approximately 111 genes (Figure 4). The CCM provides most of the relevant precursors, cofactors and energy for



**Figure 4 - Schematic overview of the Central Carbon Metabolism.** MCs are mitochondrial carriers and TCA cycle is tricarboxylic acid cycle. Four reactions are present in the both the TCA cycle and the glyoxylate cycle, namely: malate dehydrogenase, citrate synthase, aconitase and isocitrate dehydrogenase. One gene, namely *ACO1* has been counted for both the TCA cycle and the glyoxylate cycle. An elaborate figure which includes genes and metabolites can be found in **Chapter 2**.

high-value chemical building blocks and is therefore an important target for genetic engineering of cell factories (130). The production of farnesene in *S. cerevisiae* is an excellent example of the need to rewire the CCM to obtain a relevant TRY for industrial application. By introducing four non-native reactions, the generation of acetyl-CoA was altered resulting in higher farnesene yield and titer, while requiring less oxygen than the control strain (131). Despite this and some other examples (132,133), only limited knowledge-based engineering of central carbon metabolism has been performed for three main reasons: (i) the central role of the CCM enzymes in metabolism implies that the enzymes are functionally entwined with each other and with peripheral pathways, (ii) the CCM genes/enzymes are under tight and intricate regulatory control mechanisms and (iii) the CCM genes have only limited genetic accessibility. The latter point implies that genes cannot simply be deleted and replaced, because most are essential for survival of the cell. Furthermore, the approximately 111 genes of the CCM are scattered across yeast's sixteen chromosomes, making large scale reprogramming even more tedious. Focusing on the Embden-Meyerhoff-Parnas pathway of glycolysis as paradigm, Solis-Escalante *et al.* (127) and Kuijpers *et al.* (134), proposed a radically



**Figure 5 - Construction of the Switchable Yeast Glycolysis (SwYG) strain and representation of pathway swapping.** Red is minor glycolytic paralogs; orange is major glycolytic paralogs and yellow is rewired glycolysis. SinLoG is Single Locus Glycolysis.

new approach to rapidly and efficiently edit a key pathway of the CCM in yeast. The Minimal Glycolysis strain described in the previous section is a powerful platform to study genetic redundancy and to easily perform single complementation of the glycolytic and alcoholic fermentation genes. However, the remaining major paralogs are still scattered across different chromosomes, making a full replacement of glycolysis laborious and time consuming (127). Starting from the MG strain, Kuijpers *et al.* (134) developed a model for pathway swapping (Figure 5). The concept of pathway swapping is based on relocalising all genes from a particular pathway to one chromosomal location, such that the entire pathway can be easily replaced by an alternative design. In the case of glycolysis and fermentation, the 13 remaining major paralogs of the MG strain were clustered at a single locus on chromosome IX, while their native copies were removed from their original loci. Remodelling of this single locus glycolytic pathway could be achieved within 2 weeks, by integrating a second copy of glycolysis in chromosome V and removing the first copy from chromosome IX. The SwYG strain has been used to study the effect of watermarking on gene expression and yeast physiology and could be applied to study the expression, regulation and product formation of heterologous glycolysis designs (135).

## Scope of this thesis

The aim of this PhD thesis is to design, construct and use a versatile yeast synthetic platform in which the central carbon metabolism (CCM) can be remodeled at will.

The evolutionary forces responsible for the maintenance of genetic redundancy in all eukaryotic and some prokaryotic genomes are still largely unresolved. This redundancy presents an obstacle for fundamental studies and for metabolic engineering programs. The construction of the MG and SwYG strains demonstrates that genetic redundancy can be reduced without visible impact on physiology and that strains with reduced genomes can be used to remodel essential pathways in *S. cerevisiae* CCM. Inspired by these earlier studies, **Chapter 2** aims at extending these achievements to the whole CCM of *S. cerevisiae*, with the pathways specified in the previous section (Figure 4). Based on literature, an attempt was made at reducing CCM genetic complexity by deleting genes considered as redundant for growth. Redundancy was defined as null or minor effect (less than 25% decrease) on specific growth rate in the presence or absence of the genes in, chemically defined medium with glucose or ethanol as sole carbon source. This top-down approach represents a reduction in genetic complexity of 32 % in the CCM.

In an era rapidly moving towards designer genomes, the lack of predictability of biological systems presents a considerable hurdle. Considering the growing complexity of pathways genetic design, the ability to efficiently design large genetic

constructs from DNA parts and to predict the outcome of this genetic design becomes increasingly important. While the MG strain did not show a measurable phenotypic response under a wide array of growth and stress conditions, the SwYG strain in which the same set of glycolytic genes is clustered instead of scattered, did show a substantial decrease in growth rate. This unexpected response could not be explained by a decreased activity of glycolytic enzymes, but several other mechanisms could be involved. For instance, integration of this 35 kb glycolytic DNA cassette, irrespective of the genes carried, could alter structure, transcription or replication at the integration locus. Alternatively, as glycolytic genes are among the most intensively transcribed in the cells, a glycolytic DNA cassette might present a transcriptional hotspot, potentially clashing with replication or interfering with the expression of genes framing the integration site. To explore the causes of SwYG's slower growth rate, in **Chapter 3** an in-depth study of its physiology is combined with adaptive laboratory evolution and reverse engineering. Although the exact mechanism underlying the reduced growth rate was not unveiled, two realistic hypotheses could be formulated and important guiding principles for the genetic design of synthetic pathways were identified.

Expanding the pathway swapping approach from glycolysis to the entire CCM would require the clustering of the 76 major paralogs identified in **Chapter 2** at one genomic locus. Stitched altogether these genes represent a DNA construct of approximately 200 kb, the size range of *S. cerevisiae* smallest chromosome. Introduction or removal of a DNA cassette of this size into a native chromosome presents a risk to disrupt or alter the activity of neighboring genes, or even distort the integrity and stability of chromosomes. **Chapter 4** explores the potential of supernumerary synthetic neochromosomes to serve as modular platform to engineering pathways in yeast. This chapter evaluates the assembly efficiency of many overlapping transcriptional-unit sized DNA fragments into large test-chromosomes mostly consisting of non-coding DNA. The genomic stability of different test chromosome designs is probed and the impact on yeast physiology is evaluated. Moreover, a first step towards pathway swapping using neochromosomes is set in this chapter by relocalising the essential glycolytic pathway on this synthetic chromosome.

In **Chapter 5** the knowledge obtained from the previous chapters is combined in order to construct a neochromosome which serves as expression platform for native as well as heterologous pathways. The neochromosome is constructed in a strain in which part of the CCM (glycolysis, alcoholic fermentation and pentose phosphate pathway) is reduced in genetic complexity, as demonstrated in **Chapter 2**. With the knowledge of neochromosome design and construction from **Chapter 3** and **Chapter 4**, chromosomes carrying the major paralogs of glycolysis and the pentose phosphate pathway as well as the genes from the *E. coli* shikimate pathway were constructed. Subsequently, the glycolytic and pentose phosphate pathway genes were deleted from

their native location, resulting in a strain in which these pathways can be easily swapped. Finally, the constructed neochromosomes were used as a landing pad for the plant flavonoid biosynthetic pathway, such that genes encoding all metabolic enzymes from substrate to product required for anthocyanin production were located on the neochromosome. This work illustrates the potential of modular genomes in which core metabolic networks, localized on satellite, specialized neochromosomes can be swapped for alternative configurations and serve as landing pads for addition of functionalities.

## References

1. Bhatia, S. and Goli, D. (2018) *Introduction to pharmaceutical biotechnology, Volume 1 Basic techniques and concepts*. 1st ed. IOP Publishing, Bristol, UK.
2. Liu, L., Wang, J., Rosenberg, D., Zhao, H., Lengyel, G. and Nadel, D. (2018) Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. *Journal of Archaeological Science: Reports*, **21**, 783-793.
3. McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., Nunez, A., Butrym, E.D., Richards, M.P., Wang, C.S. *et al.* (2004) Fermented beverages of pre- and proto-historic China. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 17593-17598.
4. Damerow, P. (2012) Sumerian beer: the origins of brewing technology in ancient Mesopotamia. *Cuneiform Digital Library Journal*, **2**, 1-20.
5. Cavalieri, D., McGovern, P.E., Hartl, D.L., Mortimer, R. and Polsinelli, M. (2003) Evidence for *S. cerevisiae* fermentation in ancient wine. *J. Mol. Evol.*, **57 Suppl 1**, S226-232.
6. Zhang, Y.P., Sun, J. and Ma, Y. (2017) Biomanufacturing: history and perspective. *J. Ind. Microbiol. Biotechnol.*, **44**, 773-784.
7. Schwann, T. (1837) Vorläufige mittheilung, betreffend versuche über die weingährung und fäulniss. *Annalen der Physik*, **117**, 184-193.
8. Cagniard-Latour, C. (1838) Mémoire sur la fermentation vineuse. *Ann. Chim. Phys.*, **68**, 206-222.
9. Pasteur, L. (1876) *Études sur la bière: ses maladies, causes qui les provoquent, procédé pour la rendre inaltérable; avec une théorie nouvelle de la fermentation*. Gauthier-Villars.
10. Buchholz, K. and Collins, J. (2013) The roots - a short history of industrial microbiology and biotechnology. *Appl. Microbiol. Biotechnol.*, **97**, 3747-3762.
11. Weizmann, C. (1918) Fermentation process for the production of acetone and butyl alcohol. US1437697A.
12. Connstein, W. and Lüdecke, K. (1919) Preparation of glycerol by fermentation. *J. Soc. Chem. Ind.*, **38**, 691A-692A.
13. Katz, L. and Baltz, R.H. (2016) Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.*, **43**, 155-176.
14. Adrio, J.L. and Demain, A.L. (2006) Genetic improvement of processes yielding microbial products. *FEMS Microbiol. Rev.*, **30**, 187-214.
15. Kavcssek, M., Strazar, M., Curk, T., Natter, K. and Petrovic, U. (2015) Yeast as a cell factory: current state and perspectives. *Microb. Cell Fact.*, **14**, 94.
16. López-Gómez, J.P. and Pérez-Rivero, C. (2019) Cellular Systems. In: Moo-Young, M. (ed.), *Comprehensive Biotechnology (Third Edition)*. Pergamon, Oxford, pp. 9-21.
17. Keasling, J.D. (2010) Manufacturing molecules through metabolic engineering. *Science*, **330**, 1355-1358.
18. Nielsen, J. (2019) Yeast systems biology: model organism and cell factory. *Biotechnol. J.*, **14**, e1800421.
19. Paddon, C.J. and Keasling, J.D. (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.*, **12**, 355-367.
20. Pham, J.V., Yilma, M.A., Feliz, A., Majid, M.T., Maffetone, N., Walker, J.R., Kim, E., Cho, H.J., Reynolds, J.M., Song, M.C. *et al.* (2019) A review of the microbial production of bioactive natural products and biologics. *Front. Microbiol.*, **10**.
21. Nandy, S.K. and Srivastava, R. (2018) A review on sustainable yeast biotechnological processes and applications. *Microbiol. Res.*, **207**, 83-90.
22. Nielsen, J. and Keasling, Jay D. (2016) Engineering cellular metabolism. *Cell*, **164**, 1185-1197.
23. Ahn, J.H., Jang, Y.-S. and Lee, S.Y. (2016) Production of succinic acid by metabolically engineered microorganisms. *Curr. Opin. Biotechnol.*, **42**, 54-66.

24. Palsson, B.Ø. (2015) Network Reconstruction: The Process. In: Palsson, B. Ø. (ed.), *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press, Cambridge, pp. 33-49.
25. Thiele, I. and Palsson, B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, **5**, 93-121.
26. Lu, H., Li, F., Sánchez, B.J., Zhu, Z., Li, G., Domenzain, I., Marcišauskas, S., Anton, P.M., Lappa, D., Lieven, C. *et al.* (2019) A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.*, **10**, 3586.
27. Li, Y., Li, S., Thodey, K., Trenchard, I., Cravens, A. and Smolke, C.D. (2018) Complete biosynthesis of noscaphine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E3922-E3931.
28. Zha, J., Wu, X. and Koffas, M.A. (2020) Making brilliant colors by microorganisms. *Curr. Opin. Biotechnol.*, **61**, 135-141.
29. Adiego-Pérez, B., Randazzo, P., Daran, J.M., Verwaal, R., Roubos, J.A., Daran-Lapujade, P. and van der Oost, J. (2019) Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiol. Lett.*, **366**, fnz086.
30. Wijisman, M., Swiat, M.A., Marques, W.L., Hettinga, J.K., van den Broek, M., Torre Cortés, P., Mans, R., Pronk, J.T., Daran, J.M. and Daran-Lapujade, P. (2019) A toolkit for rapid CRISPR-SpCas9 assisted construction of hexose-transport-deficient *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.*, **19**, foy107.
31. Nielsen, J. (2017) Systems biology of metabolism. *Annu. Rev. Biochem.*, **86**, 245-275.
32. Andrianantoandro, E., Basu, S., Karig, D.K. and Weiss, R. (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, **2**, 2006.0028.
33. Nielsen, J., Fussenegger, M., Keasling, J., Lee, S.Y., Liao, J.C., Prather, K. and Palsson, B. (2014) Engineering synergy in biotechnology. *Nat. Chem. Biol.*, **10**, 319-322.
34. Qian, Y., McBride, C. and Vecchio, D.D. (2018) Programming cells to work for us. *Annual Review of Control, Robotics, and Autonomous Systems*, **1**, 411-440.
35. Krivoruchko, A., Siewers, V. and Nielsen, J. (2011) Opportunities for yeast metabolic engineering: lessons from synthetic biology. *Biotechnol. J.*, **6**, 262-276.
36. Chen, B., Lee, H.L., Heng, Y.C., Chua, N., Teo, W.S., Choi, W.J., Leong, S.S.J., Foo, J.L. and Chang, M.W. (2018) Synthetic biology toolkits and applications in *Saccharomyces cerevisiae*. *Biotechnol. Adv.*, **36**, 1870-1881.
37. Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A. *et al.* (2013) Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.*, **9**, 701.
38. Decoene, T., De Maeseneire, S.L. and De Mey, M. (2019) Modulating transcription through development of semi-synthetic yeast core promoters. *PLoS one*, **14**, e0224476.
39. Curran, K.A., Crook, N.C., Karim, A.S., Gupta, A., Wagman, A.M. and Alper, H.S. (2014) Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat. Commun.*, **5**, 4002.
40. Wang, J., Zhai, H., Rexida, R., Shen, Y., Hou, J. and Bao, X. (2018) Developing synthetic hybrid promoters to increase constitutive or diauxic shift-induced expression in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **18**, foy098.
41. Deaner, M. and Alper, H.S. (2017) Systematic testing of enzyme perturbation sensitivities via graded dCas9 modulation in *Saccharomyces cerevisiae*. *Metab. Eng.*, **40**, 14-22.
42. Choi, K.R., Jang, W.D., Yang, D., Cho, J.S., Park, D. and Lee, S.Y. (2019) Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol.*, **37**, 817-837.
43. Wang, Z., Wei, L., Sheng, Y. and Zhang, G. (2019) Yeast synthetic terminators: fine regulation of strength through linker sequences. *Chembiochem*, **20**, 2383-2389.

44. Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R. and Bartel, D.P. (2009) RNAi in budding yeast. *Science*, **326**, 544-550.
45. Cameron, D.E., Bashor, C.J. and Collins, J.J. (2014) A brief history of synthetic biology. *Nat. Rev. Microbiol.*, **12**, 381-390.
46. Ho-Shing, O., Lau, K.H., Vernon, W., Eckdahl, T.T. and Campbell, A.M. (2012) Assembly of Standardized DNA Parts Using BioBrick Ends in *E. coli*. In: Peccoud, J. (ed.), *Gene Synthesis: Methods and Protocols*. Humana Press, Totowa, NJ, pp. 61-76.
47. Coradini, A.L.V., Hull, C.B. and Ehrenreich, I.M. (2020) Building genomes to understand biology. *Nat. Commun.*, **11**, 6177.
48. Zhang, W., Mitchell, L.A., Bader, J.S. and Boeke, J.D. (2020) Synthetic genomes. *Annu. Rev. Biochem.*, **89**, 77-101.
49. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215-1220.
50. Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O. and Hutchison, C.A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 20404-20409.
51. Beaucage, S.L. and Caruthers, M.H. (1981) Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.*, **22**, 1859-1862.
52. Hughes, R.A. and Ellington, A.D. (2017) Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.*, **9**, a023812.
53. Lee, H., Wiegand, D.J., Griswold, K., Punthambaker, S., Chun, H., Kohman, R.E. and Church, G.M. (2020) Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage. *Nat. Commun.*, **11**, 5246.
54. Lee, H.H., Kalhor, R., Goela, N., Bolot, J. and Church, G.M. (2019) Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat. Commun.*, **10**, 2383.
55. Ostrov, N., Beal, J., Ellis, T., Gordon, D.B., Karas, B.J., Lee, H.H., Lenaghan, S.C., Schloss, J.A., Stracquandano, G., Trefzer, A. *et al.* (2019) Technological challenges and milestones for writing genomes. *Science*, **366**, 310-312.
56. Paul, S.S., Trabelsi, H., Yaseen, Y., Basu, U., Altaïi, H.A. and Dhali, D. (2021) Chapter 2 - Advances in long DNA synthesis. In: Singh, V. (ed.), *Microbial cell factories engineering for production of biomolecules*. Academic Press, pp. 21-36.
57. Hao, M., Qiao, J. and Qi, H. (2020) Current and emerging methods for the synthesis of single-stranded DNA. *Genes*, **11**, 116.
58. Eisenstein, M. (2020) Enzymatic DNA synthesis enters new phase. *Nat. Biol.*, **38**, 1113-1115.
59. Forster, A.C. and Church, G.M. (2006) Towards synthesis of a minimal cell. *Mol. Syst. Biol.*, **2**, 45.
60. Chao, R., Yuan, Y. and Zhao, H. (2015) Recent advances in DNA assembly technologies. *FEMS Yeast Res.*, **15**, 1-9.
61. Casini, A., Storch, M., Baldwin, G.S. and Ellis, T. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.*, **16**, 568-576.
62. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343-345.
63. Gibson, D.G., Smith, H.O., Hutchison, C.A., 3rd, Venter, J.C. and Merryman, C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods*, **7**, 901-903.
64. Karas, B.J., Suzuki, Y. and Weyman, P.D. (2015) Strategies for cloning and manipulating natural and synthetic chromosomes. *Chromosome Res.*, **23**, 57-68.

65. Libicher, K., Hornberger, R., Heymann, M. and Mutschler, H. (2020) *In vitro* self-replication and multicistronic expression of large synthetic genomes. *Nat. Commun.*, **11**, 904.
66. van Nies, P., Westerlaken, I., Blanken, D., Salas, M., Mencía, M. and Danelon, C. (2018) Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nat. Commun.*, **9**, 1583.
67. Mukai, T., Yoneji, T., Yamada, K., Fujita, H., Nara, S. and Su'etsugu, M. (2020) Overcoming the challenges of megabase-sized plasmid construction in *Escherichia coli*. *ACS Synth. Biol.*, **9**, 1315-1327.
68. Su'etsugu, M., Takada, H., Katayama, T. and Tsujimoto, H. (2017) Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. *Nucleic Acids Res.*, **45**, 11525-11534.
69. Shao, Y., Lu, N., Wu, Z., Cai, C., Wang, S., Zhang, L.-L., Zhou, F., Xiao, S., Liu, L., Zeng, X. *et al.* (2018) Creating a functional single-chromosome yeast. *Nature*, **560**, 331-335.
70. Burke, D.T., Carle, G.F. and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, **236**, 806-812.
71. Larionov, V., Kouprina, N., Graves, J., Chen, X.N., Korenberg, J.R. and Resnick, M.A. (1996) Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 491-496.
72. Kunes, S., Botstein, D. and Fox, M.S. (1985) Transformation of yeast with linearized plasmid DNA. Formation of inverted dimers and recombinant plasmid products. *J. Mol. Biol.*, **184**, 375-387.
73. Noskov, V.N., Koriabine, M., Solomon, G., Randolph, M., Barrett, J.C., Leem, S.H., Stubbs, L., Kouprina, N. and Larionov, V. (2001) Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. *Nucleic Acids Res.*, **29**, E32.
74. Larionov, V., Kouprina, N., Eldarov, M., Perkins, E., Porter, G. and Resnick, M.A. (1994) Transformation-associated recombination between diverged and homologous DNA repeats is induced by strand breaks. *Yeast*, **10**, 93-104.
75. Oldfield, L.M., Grzesik, P., Voorhies, A.A., Alperovich, N., MacMath, D., Najera, C.D., Chandra, D.S., Prasad, S., Noskov, V.N., Montague, M.G. *et al.* (2017) Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E8885-E8894.
76. Thi Nhu Thao, T., Labroussaa, F., Ebert, N., V'Kovski, P., Stalder, H., Portmann, J., Kelly, J., Steiner, S., Holwerda, M., Kratzel, A. *et al.* (2020) Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature*, **582**, 561-565.
77. Gibson, D.G., Glass, J.L., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M. *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**, 52-56.
78. Tagwerker, C., Dupont, C.L., Karas, B.J., Ma, L., Chuang, R.-Y., Benders, G.A., Ramon, A., Novotny, M., Montague, M.G., Venepally, P. *et al.* (2012) Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res.*, **40**, 10375-10383.
79. Noskov, V.N., Karas, B.J., Young, L., Chuang, R.Y., Gibson, D.G., Lin, Y.C., Stam, J., Yonemoto, I.T., Suzuki, Y., Andrews-Pfannkoch, C. *et al.* (2012) Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synth. Biol.*, **1**, 267-273.
80. Karas, B.J., Tagwerker, C., Yonemoto, I.T., Hutchison, C.A., 3rd and Smith, H.O. (2012) Cloning the *Acholeplasma laidlawii* PG-8A genome in *Saccharomyces cerevisiae* as a yeast centromeric plasmid. *ACS Synth. Biol.*, **1**, 22-28.
81. Fredens, J., Wang, K., de la Torre, D., Funke, L.F.H., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., Beránek, V. *et al.* (2019) Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, **569**, 514-518.

82. Venetz, J.E., Del Medico, L., Wölfle, A., Schächle, P., Bucher, Y., Appert, D., Tschan, F., Flores-Tinoco, C.E., van Kooten, M., Guennoun, R. *et al.* (2019) Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8070-8079.
83. Karas, B.J., Molparia, B., Jablanovic, J., Hermann, W.J., Lin, Y.C., Dupont, C.L., Tagwerker, C., Yonemoto, I.T., Noskov, V.N., Chuang, R.Y. *et al.* (2013) Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng.*, **7**, 30.
84. O'Neill, B.M., Mikkelsen, K.L., Gutierrez, N.M., Cunningham, J.L., Wolff, K.L., Szyjka, S.J., Yohn, C.B., Redding, K.E. and Mendez, M.J. (2012) An exogenous chloroplast genome for complex sequence manipulation in algae. *Nucleic Acids Res.*, **40**, 2782-2792.
85. Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., Wang, A., Jiang, S., Jiang, Q., Gong, J. *et al.* (2017) Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science*, **355**, eaaf3981.
86. Mitchell, L.A., McCulloch, L.H., Pingley, S., Berger, H., Bosco, N., Brosh, R., Bulajić, M., Huang, E., Hogan, M.S., Martin, J.A. *et al.* (2021) *De novo* assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics*, **218**.
87. Pretorius, I.S. and Boeke, J.D. (2018) Yeast 2.0-connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Res.*, **18**, foy032.
88. Eisenstein, M. (2020) How to build a genome. *Nature*, **578**, 633-635.
89. Dai, J., Boeke, J.D., Luo, Z., Jiang, S. and Cai, Y. (2020) Sc3.0: revamping and minimizing the yeast genome. *Genome Biol.*, **21**, 205.
90. Benders, G.A., Noskov, V.N., Denisova, E.A., Lartigue, C., Gibson, D.G., Assad-Garcia, N., Chuang, R.-Y., Carrera, W., Moodie, M., Algire, M.A. *et al.* (2010) Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.*, **38**, 2558-2569.
91. Sorek, R., Zhu, Y., Creevey, C.J., Francino, M.P., Bork, P. and Rubin, E.M. (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, **318**, 1449-1452.
92. Itaya, M., Sato, M., Hasegawa, M., Kono, N., Tomita, M. and Kaneko, S. (2018) Far rapid synthesis of giant DNA in the *Bacillus subtilis* genome by a conjugation transfer system. *Sci. Rep.*, **8**, 1-6.
93. Blount, B.A., Driessen, M.R. and Ellis, T. (2016) GC preps: fast and easy extraction of stable yeast genomic DNA. *Sci. Rep.*, **6**, 1-4.
94. Lee, K., Ji, J.H., Yoon, K., Che, J., Seol, J.H., Lee, S.E. and Shim, E.Y. (2019) Microhomology selection for microhomology mediated end joining in *Saccharomyces cerevisiae*. *Genes*, **10**, 284.
95. Ranjha, L., Howard, S.M. and Cejka, P. (2018) Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma*, **127**, 187-214.
96. Albarracín, V.H., Pathak, G.P., Douki, T., Cadet, J., Borsarelli, C.D., Gärtner, W. and Farias, M.E. (2012) Extremophilic *Acinetobacter* strains from high-altitude lakes in Argentinean Puna: remarkable UV-B resistance and efficient DNA damage repair. *Orig. Life Evol. Biosph.*, **42**, 201-221.
97. Sato, T., Takada, D., Itoh, T., Ohkuma, M. and Atomi, H. (2020) Integration of large heterologous DNA fragments into the genome of *Thermococcus kodakarensis*. *Extremophiles*, **24**, 339-353.
98. Guo, Z.P., Borsenberger, V., Croux, C., Duquesne, S., Truan, G., Marty, A. and Bordes, F. (2020) An artificial chromosome  $\gamma$ AC enables efficient assembly of multiple genes in *Yarrowia lipolytica* for biomanufacturing. *Commun. Biol.*, **3**, 199.
99. Chandran, S. (2017) Rapid assembly of DNA via ligase cycling reaction (LCR). *Methods Mol. Biol.*, **1472**, 105-110.
100. Kostylev, M., Otwell, A.E., Richardson, R.E. and Suzuki, Y. (2015) Cloning should be simple: *Escherichia coli* DH5 $\alpha$ -mediated assembly of multiple DNA fragments with short end homologies. *PLoS one*, **10**, e0137466.

101. Itaya, M., Fujita, K., Kuroki, A. and Tsuge, K. (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat. Methods*, **5**, 41-43.
102. Dean, F.B., Nelson, J.R., Giesler, T.L. and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.*, **11**, 1095-1099.
103. Kwon, Y., Daley, J.M. and Sung, P. (2017) Reconstituted system for the examination of repair DNA synthesis in homologous recombination. *Methods Enzymol.*, **591**, 307-325.
104. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biol.*, **19**, 751-755.
105. Boeke, J.D., Church, G., Hessel, A., Kelley, N.J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V.W., Holt, L. *et al.* (2016) Genome engineering. The genome project-write. *Science*, **353**, 126-127.
106. Mutschler, H., Robinson, T., Tang, T.D. and Wegner, S. (2019) Special issue on bottom-up synthetic biology. *ChemBiochem*, **20**, 2533-2534.
107. Xu, C., Hu, S. and Chen, X. (2016) Artificial cells: from basic science to applications. *Mater. Today*, **19**, 516-532.
108. Jia, H. and Schwille, P. (2019) Bottom-up synthetic biology: reconstitution in space and time. *Curr. Opin. Biotechnol.*, **60**, 179-187.
109. Escalera-Fanjul, X., Quezada, H., Riego-Ruiz, L. and Gonzalez, A. (2019) Whole-genome duplication and yeast's fruitful way of life. *Trends Genet.*, **35**, 42-54.
110. Lara, A.R. and Gosset, G. (2019) *Minimal cells: design, construction, biotechnological applications*. Springer International Publishing.
111. Stouthamer, A.H. (1973) A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek*, **39**, 545-565.
112. Farmer, I.S. and Jones, C.W. (1976) The energetics of *Escherichia coli* during aerobic growth in continuous culture. *Eur. J. Biochem.*, **67**, 115-122.
113. Price, M.N., Wetmore, K.M., Deutschbauer, A.M. and Arkin, A.P. (2016) A comparison of the costs and benefits of bacterial gene expression. *PLoS one*, **11**, e0164314.
114. Lara, A.R., Galindo, E., Ramirez, O.T. and Palomares, L.A. (2006) Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. *Mol. Biotechnol.*, **34**, 355-381.
115. Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., Mosberg, J.A. *et al.* (2013) Genomically recoded organisms expand biological functions. *Science*, **342**, 357-360.
116. Smallbone, K., Messiha, H.L., Carroll, K.M., Winder, C.L., Malys, N., Dunn, W.B., Murabito, E., Swainston, N., Dada, J.O., Khan, F. *et al.* (2013) A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. *FEBS Lett.*, **587**, 2832-2841.
117. Totis, N., Tagherloni, A., Beccuti, M., Cazzaniga, P., Nobile, M.S., Besozzi, D., Pennisi, M. and Pappalardo, F. (2020) Efficient and settings-free calibration of detailed kinetic metabolic models with enzyme isoforms characterization. Springer International Publishing, Cham, pp. 187-202.
118. Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., Ote, T., Yamakawa, T., Yamazaki, Y., Mori, H. *et al.* (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.*, **55**, 137-149.
119. Morimoto, T., Kadoya, R., Endo, K., Tohata, M., Sawada, K., Liu, S., Ozawa, T., Kodama, T., Kakeshita, H., Kageyama, Y. *et al.* (2008) Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res.*, **15**, 73-81.
120. Suzuki, N., Nonaka, H., Tsuge, Y., Inui, M. and Yukawa, H. (2005) New multiple-deletion method for the *Corynebacterium glutamicum* genome, using a mutant *lox* sequence. *Appl. Environ. Microbiol.*, **71**, 8472-8480.

121. Lieder, S., Nikel, P.I., de Lorenzo, V. and Takors, R. (2015) Genome reduction boosts heterologous gene expression in *Pseudomonas putida*. *Microb. Cell Fact.*, **14**, 23.
122. Zhu, D., Fu, Y., Liu, F., Xu, H., Saris, P.E.J. and Qiao, M. (2017) Enhanced heterologous protein productivity by genome reduction in *Lactococcus lactis* NZ9000. *Microb. Cell Fact.*, **16**, 1.
123. Komatsu, M., Uchiyama, T., Omura, S., Cane, D.E. and Ikeda, H. (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 2646-2651.
124. Sasaki, M., Kumagai, H., Takegawa, K. and Tohda, H. (2013) Characterization of genome-reduced fission yeast strains. *Nucleic Acids Res.*, **41**, 5382-5399.
125. Murakami, K., Tao, E., Ito, Y., Sugiyama, M., Kaneko, Y., Harashima, S., Sumiya, T., Nakamura, A. and Nishizawa, M. (2007) Large scale deletions in the *Saccharomyces cerevisiae* genome create strains with altered regulation of carbon metabolism. *Appl. Microbiol. Biotechnol.*, **75**, 589-597.
126. Wieczorke, R., Krampe, S., Weierstall, T., Freidel, K., Hollenberg, C.P. and Boles, E. (1999) Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.*, **464**, 123-128.
127. Solis-Escalante, D., Kuijpers, N.G., Barrajon-Simancas, N., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2015) A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell*, **14**, 804-816.
128. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L. *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253.
129. Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y. *et al.* (2017) Design of a synthetic yeast genome. *Science*, **355**, 1040-1044.
130. Nielsen, J. (2019) Cell factory engineering for improved production of natural products. *Nat. Prod. Rep.*, **36**, 1233-1236.
131. Meadows, A.L., Hawkins, K.M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., Dahl, R.H., Tai, A., Mahatdejkul-Meadows, T., Xu, L. *et al.* (2016) Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature*, **537**, 694-697.
132. Sánchez-Pascuala, A., Fernández-Cabezón, L., de Lorenzo, V. and Nikel, P.I. (2019) Functional implementation of a linear glycolysis for sugar catabolism in *Pseudomonas putida*. *Metab. Eng.*, **54**, 200-211.
133. Liu, Q., Yu, T., Li, X., Chen, Y., Campbell, K., Nielsen, J. and Chen, Y. (2019) Rewiring carbon metabolism in yeast for high level production of aromatic chemicals. *Nat. Commun.*, **10**, 4976.
134. Kuijpers, N.G., Solis-Escalante, D., Lutтик, M.A., Bisschops, M.M., Boonekamp, F.J., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2016) Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 15060-15065.
135. Boonekamp, F.J., Dashko, S., Duiker, D., Gehrman, T., van den Broek, M., den Ridder, M., Pabst, M., Robert, V., Abeel, T., Postma, E.D. *et al.* (2020) Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences. *ACS Synth. Biol.*, **9**, 1361-1375.



# Chapter 2

## Top-down, knowledge-based genetic reduction of yeast central carbon metabolism

Eline D. Postma  
Lucas G.F. Couwenberg  
Roderick N. van Roosmalen  
Jordi Geelhoed  
Philip A. de Groot  
Pascale Daran-Lapujade

## Abbreviations

CCM	Central Carbon Metabolism
WGD	Whole Genome Duplication
SSD	Smaller-Scale Duplication
MG	Minimal Glycolysis
PPP	Pentose Phosphate Pathway
MC	Mitochondrial Carrier
TCA cycle	Tricarboxylic Acid cycle
glyc <sup>min</sup>	minimized glycolysis
fer <sup>min</sup>	minimized ethanolic fermentation
ppp <sup>min</sup>	minimized pentose phosphate pathway
tca <sup>min</sup>	minimized tricarboxylic acid cycle
mc <sup>min</sup>	minimized mitochondrial carriers
fum <sup>min</sup>	minimized fumarate reductases
glyox <sup>min</sup>	minimized glyoxylate cycle
ace <sup>min</sup>	minimized Acetyl-CoA synthesis
glycerol <sup>min</sup>	minimized glycerol synthesis
Glc	Glucose
Glc-6P	Glucose-6-phosphate
Fru-6P	Fructose-6-phosphate
Fru-1,6-BP	Fructose-1,6-bisphosphate
GAP	Glyceraldehyde-3-phosphate
DHAP	Dihydroxyacetone phosphate
1,3-BPG	1,3-bisphosphoglycerate
3-PG	3-phosphoglycerate
2-PG	2-phosphoglycerate
PEP	Phosphoenolpyruvate
Pyr	Pyruvate
AcAl	Acetaldehyde
EtOH	Ethanol
6P-GLCN-lac	6-Phosphogluconolactone
6P-GLCN	6-Phosphoglucononate
RL5P	Ribulose 5-phosphate
R5P	Ribose 5-phosphate
XUL-5P	Xylulose 5-phosphate
S7P	Sedoheptulose 7-phosphate
Ery-4P	Erythrose 4-phosphate
OAA	Oxaloacetate
Cit	Citrate

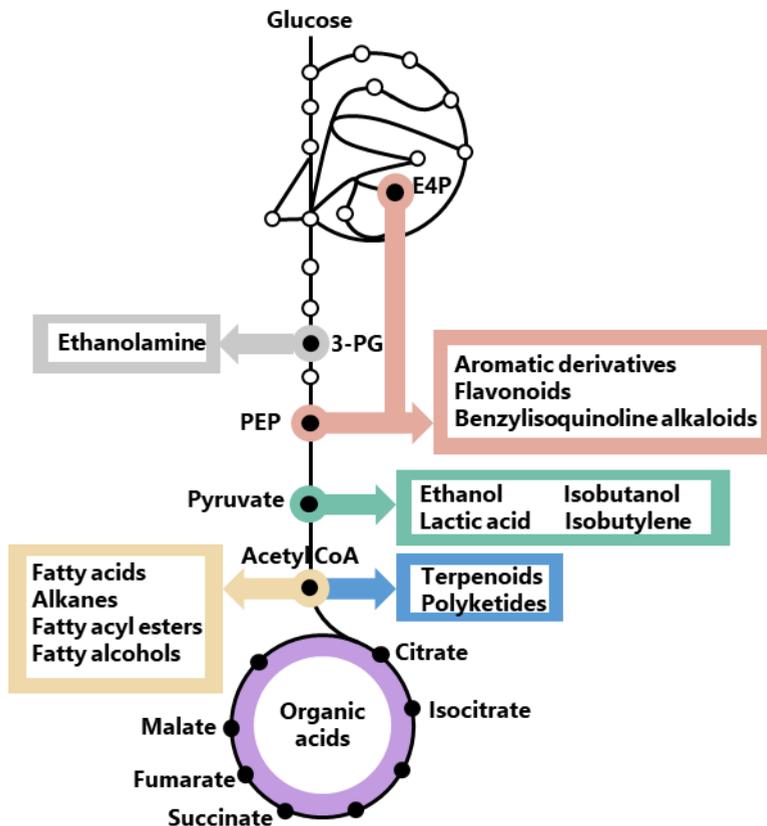
Cis-Aco	Cis-aconitate
Isocit	Isocitrate
$\alpha$ -KG	$\alpha$ -ketoglutarate
Suc-CoA	Succinyl-CoA
Suc	Succinate
Fum	Fumarate
Mal	Malate
Suc	Succinate
Glyox	Glyoxylate
Ace	Acetate
Ac-CoA	Acetyl-CoA
Glyc-3P	Glycerol-3-phosphate
YPD	Yeast Peptone Dextrose
SMD	Synthetic Medium Dextrose
SME	Synthetic Medium Ethanol
SMGal	Synthetic Medium Galactose
SMMal	Synthetic Medium Maltose
SMSuc	Synthetic Medium Sucrose
SMFruc	Synthetic Medium Fructose
SMPyr	Synthetic Medium Pyruvate

## Abstract

*Saccharomyces cerevisiae*, whose evolutionary past includes a whole-genome duplication event, is characterized by a mosaic genome configuration with substantial apparent genetic redundancy. This apparent redundancy raises questions about the evolutionary driving force for genomic fixation of 'minor' paralogs and complicates modular and combinatorial metabolic engineering strategies. While isoenzymes might be important in specific environments, they could be dispensable in controlled laboratory or industrial contexts. The present study explores the extent to which the genetic complexity of the central carbon metabolism (CCM) in *S. cerevisiae*, here defined as the combination of glycolysis, pentose phosphate pathway, tricarboxylic acid cycle and a limited number of related pathways and reactions, can be reduced by elimination of (iso)enzymes without major negative impacts on strain physiology. Cas9-mediated, groupwise deletion of 35 from the 111 genes yielded a 'minimal CCM' strain, which despite the elimination of 32 % of CCM-related proteins, showed only a minimal change in phenotype on glucose-containing synthetic medium in controlled bioreactor cultures relative to a congenic reference strain. Analysis under a wide range of other growth and stress conditions revealed remarkably few phenotypic changes of the reduction of genetic complexity. Still, a well-documented context-dependent role of *GPD1* in osmotolerance was confirmed. The minimal CCM strain provides a model system for further research into genetic redundancy of yeast genes and a platform for strategies aimed at large-scale, combinatorial remodeling of yeast CCM.

## Introduction

The fundamental challenge of defining the minimum complement of genes required for life has been addressed by theoretical as well as experimental approaches. Bottom-up and top-down strategies mainly focused on bacteria with small genomes (1-9). Their larger genome sizes might appear to make eukaryotic microorganisms less relevant for this type of research. However, they do offer attractive models to explore the biological significance of (apparent) genetic redundancy. Different evolutionary advantages have been proposed for the fixation of duplicated genes in genomes, including provision of a molecular landscape for functional (minor or major) innovation (e.g., neo- and subfunctionalization), a functional backup, gene dosage effects or increased buffering to respond to environmental cues (10,11). Systematically identifying the physiological significance underlying gene fixation presents a daunting challenge.



**Figure 1 - CCM precursors of industrially relevant chemicals.** Abbreviations: 3-PG is 3-Phosphoglycerate and PEP is phosphoenolpyruvate.

With its relatively small genome (12 Mb), tractability and high genetic accessibility, the yeast *Saccharomyces cerevisiae* is a valuable model for fundamental research on minimal genetic requirements. *S. cerevisiae* underwent a whole genome duplication (WGD) approximately 100 million years ago as well as Smaller-Scale Duplication (SSD) events. While 90 % of the WGD genes were lost during evolution, some duplicates remain (11). As in humans, a substantial fraction of the total gene duplicates in *S. cerevisiae* originates from the WGD (approximately 63 % of duplicates in *S. cerevisiae* genome and 62 % in the human genome), while a smaller fraction originates from SSD (approximately 37 % of duplicates in *S. cerevisiae* genome and 38 % in the human genome) (12,13). While systematic, large scale studies like the construction of the yeast deletion collections (14-17), the synthetic genetic array projects (18-20) or the recent SCRaMBLE-based genome compaction (21), have provided valuable information on the dispensability of (a combination of) genes, the physiological role of many of these paralogous genes remains poorly defined.

In addition to the fundamental scientific questions raised by genetic redundancy, it also complicates genome engineering of *S. cerevisiae*. The conversion of substrate into product via native or engineered pathways, relies on the microbial host core pathways for the supply of metabolic precursors, energy-rich molecules and redox equivalents. These biochemical reactions are catalyzed by sets of 'metabolic' genes that are characterized by a high genetic redundancy in eukaryotes (11). Not only has the physiological role of many paralogous genes not been fully elucidated, but the manipulation of specific biochemical reactions is hindered by the presence of multiple paralogous genes with poorly known functions that are scattered over the 12 Mb, mosaic yeast genome and its 16 chromosomes. Additionally, expression of these redundant genes dissipates cellular resources (e.g., carbon, energy) that might be better invested in industrially-relevant properties such as high product yield or cellular robustness to the stressful environment of large-scale fermentation.

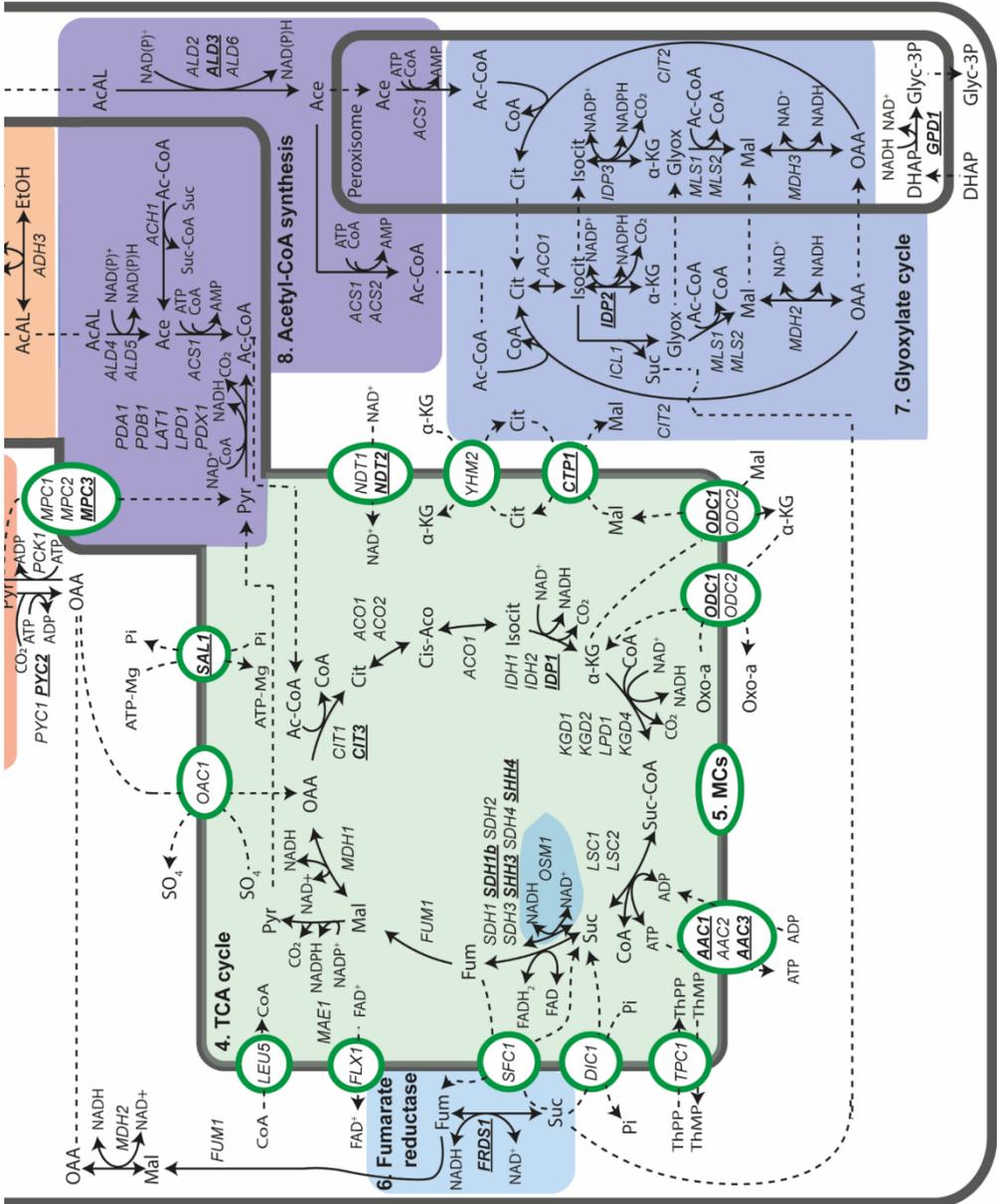
To tackle these fundamental and applied challenges, taking glycolysis and ethanolic fermentation as starting point, Solis-Escalante and colleagues pioneered the genetic reduction of central carbon metabolism in *S. cerevisiae*. The set of 26 genes encoding the (iso)enzymes catalyzing 12 reactions was reduced to 13 genes (22). Remarkably, this 50 % genetic reduction did not result in any visible phenotypic effect, although a wide range of growth conditions were tested. These observations argued against gene dosage being a strong driving force in the evolution of Crabtree positive yeasts (11,22) and raised questions on the mechanisms involved the fixation of these gene duplicates in the *S. cerevisiae* genome. A recent study suggests that the role of the redundant paralogs might be highly context-dependent and that some relevant conditions were not tested by Solis-Escalante *et al.* (22) (e.g. the role of Pyruvate kinase 2 in the utilization of three-carbon substrates such as dihydroxyacetone (23)). The surprising

lack of phenotype of the “Minimal Glycolysis” yeast strain (called MG) enabled the construction of a genetically simplified version of the glycolytic pathway, which was subsequently relocalized to a single chromosomal locus (24). The resulting yeast strain with a single-locus glycolysis presents a powerful tool to remodel the glycolytic pathway in two single steps into any redesigned (heterologous) version.

Glycolysis is an important but small part of Central Carbon Metabolism (CCM), a set of reactions required for the conversion of carbon feedstocks into any industrially-relevant product (Figure 1). For cells, CCM is primarily the set of reactions that convert carbon sources into the 12 building blocks required for the synthesis of cellular components (25). CCM encompasses ca. 111 genes, with 66 % of duplicates (Figure 2). Reducing its genetic complexity would be the first step in an attempt to construct a modular, designer yeast genome, with a single-locus CCM, as previously achieved for glycolysis and fermentation. Modular, specialized synthetic chromosomes could be ideal platforms for the centralization of the CCM genes (26).

The main goal of this study was to explore the extent to which the number of genes encoding CCM enzymes in *S. cerevisiae* can be reduced without substantially affecting fitness under a set of chosen growth conditions. To this end, redundancies were first predicted based on literature data on gene expression, enzyme activities and phenotype of (single) deletion mutants. Subsequently, phenotypes of mutants with mutations in sets of genes encoding CCM enzymes were tested under a wider range of growth conditions. In this first attempt of genetic reduction of yeast CCM at this scale, special attention is given to possible synergistic effects of mutations that were previously studied in separate strains.





**Figure 2 - Reactions of the central carbon metabolism (CCM) of *S. cerevisiae* considered for genetic reduction.** Schematic representation of sub-processes and pathways in CCM metabolic pathways considered for genetic reduction in this study. Enzyme-catalysed conversions (black lines) and transport processes (dotted lines) are shown between intermediates and through mitochondrial transporters (circles and ovals), respectively. Directionality and reversibility of reactions was based on <https://pathway.yeastgenome.org/>. Enzyme localization was based on literature information. Genes retained in the genetic reduction strategy are shown in black, genes selected for deletion in the minimal CCM strain are indicated in bold and underlined. Occurrence of pathways in different cellular compartments is shown by grey borders. Simplifications have been made for visualization reasons, for example H<sub>2</sub>O and inorganic phosphate are not shown. Abbreviations can be found in the abbreviation list.

## Results

### Genetic reduction strategy

In this study the CCM of *S. cerevisiae* was defined as the set of biochemical reactions encompassed by glycolysis, ethanolic fermentation, pentose-phosphate pathway, acetyl-CoA synthesis, tricarboxylic acid cycle, anaplerosis, gluconeogenesis, glyoxylate cycle and glycerol metabolism. As CCM reactions occur in multiple compartments, mitochondrial transporters were also considered (Figure 2). Transport over the peroxisomal membrane was not considered as this phenomenon is poorly studied (27). For the construction of a minimal CCM strain, decisions to remove or retain genes were based on (i) transcript levels from an expression compendium encompassing 170 different cultivation conditions (28), (ii) enzyme activities in cell extracts of mutant strains when data were available, and (iii) on reported phenotypes of null mutants. Genes encoding proteins with reported secondary ('moonlighting') functions or proteins known to cause auxotrophy upon deletion were retained (29).

Genes were classified as functionally redundant when at least 75 % of the specific growth rate of the congeneric reference strains CEN.PK113-7D (Ura<sup>+</sup>) or IMX581 (Ura<sup>-</sup>) was retained during aerobic batch cultivation on synthetic medium supplied with either glucose or ethanol. Ethanol-grown cultures were included as, in contrast to glucose, ethanol can only be dissimilated by respiration and because its metabolism involves different sets of CCM enzymes and transporters. In addition, testing for growth on ethanol ensured that the intensive engineering undergone by the strains, including removal of several mitochondrial proteins, did not cause respiratory deficiency.

The previously constructed MG strain, in which 13 out of the 26 existing paralogs of genes encoding glycolytic enzymes and fermentation enzymes were deleted without the detection of major phenotypes, was used as starting point of the present CCM reduction endeavor. To identify any synergistic effects between the newly introduced deletions and the 13 deletions already present in the MG strain, a congeneric naïve reference strain (IMX581) with a full complement of glycolytic and fermentation genes was also used in parallel to MG for serial deletions. To accelerate the deletion workflow, genes involved in individual pathways or processes were deleted in sets of two to four (Figure 3). When substantial loss of fitness was observed, the contribution of individual deletions was dissected by constructing additional strains with various combinations and numbers of deletions.

## **Deletion of 35 CCM genes has minimal impact on specific growth rate on chemically defined glucose medium**

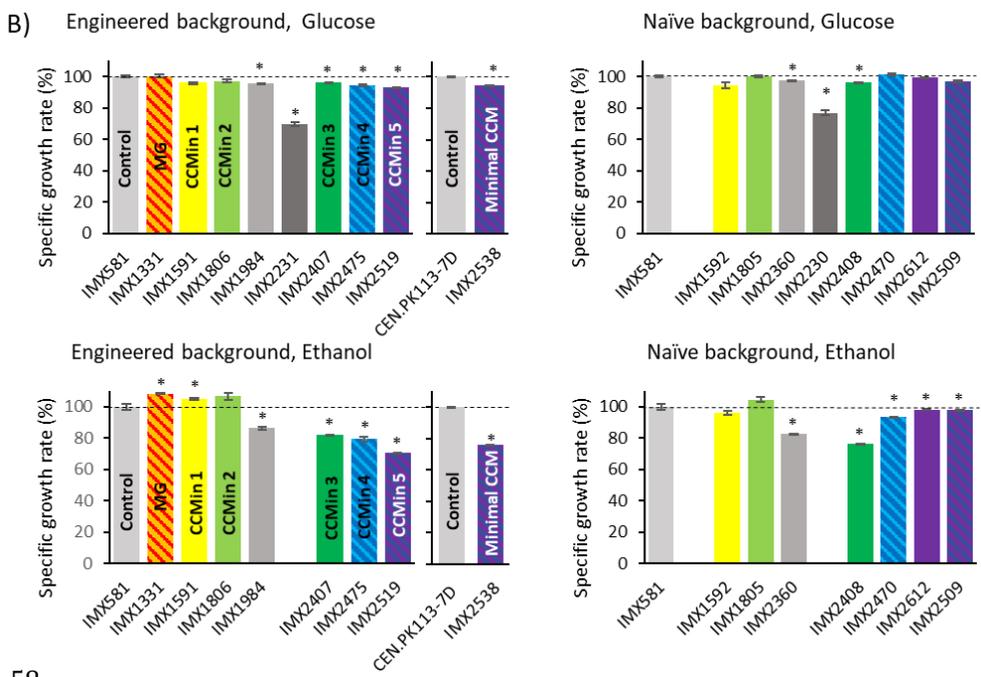
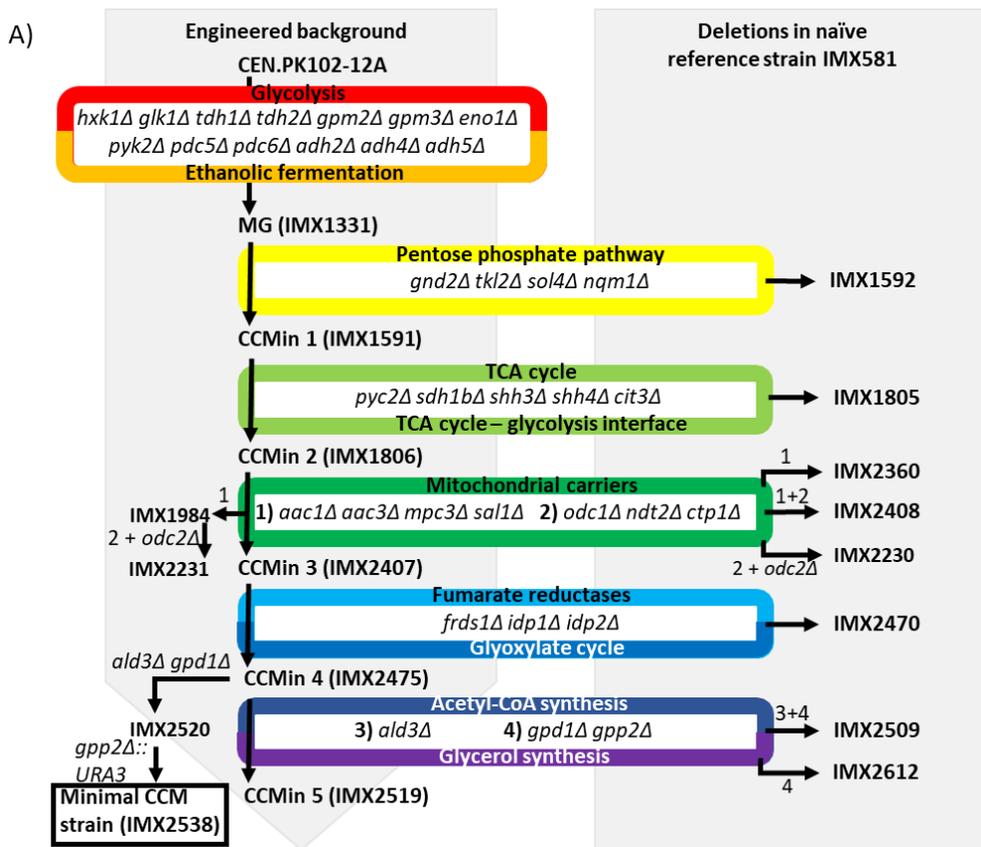
### Pentose-phosphate pathway

The pentose-phosphate pathway (PPP) reduces cellular NADP<sup>+</sup>, generates ribose-5-phosphate and erythrose-4-phosphate for nucleic acid and amino-acid synthesis and, in strains engineered for pentose fermentation, acts as dissimilatory pathway (30).

Four of the seven reactions in the PPP are catalyzed by pairs of isoenzymes encoded by WGD paralogs, for which sequence similarities range from 47 % (*SOL3* and *SOL4*) to 87 % (*GND1* and *GND2*) (Supplementary Table S1). Based on transcript levels across a wide ranges of cultivation conditions (28), *SOL4*, *GND2*, *TKL2* and *NQM1* were considered as minor paralogs. Moreover, deletion of *TKL2* and *NQM1* was previously reported not to affect growth on glucose synthetic medium (31-33). While similar *in vitro* enzyme activities were reported for Sol3 and Sol4 in cell extracts (34), *SOL4* was deleted based on its consistently lower transcript level (28). Simultaneous deletion of *GND2*, *TKL2*, *SOL4* and *NQM1* in the naïve reference strain or in the MG strain, while retaining *SOL3*, *GND1*, *TKL1* and *TAL1*, did not significantly affect growth rate on either synthetic medium with 2 % (w/v) glucose (SMD) or 2 % (v/v) ethanol (SME) as carbon sources (strains IMX1592 (ppp<sup>min</sup>) and IMX1591 (called CCMin1: glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup>), Figure 3). A previously reported extended lag phase and slower growth on ethanol of *sol4* null mutants (35) was not observed. This difference may be related to the use of different *S. cerevisiae* strain backgrounds.

### Tricarboxylic acid cycle, anaplerotic reactions and gluconeogenesis

In addition to its dissimilatory role in oxidizing acetyl-CoA units to CO<sub>2</sub>, the tricarboxylic acid (TCA) cycle supplies precursors, NADH, FADH<sub>2</sub> and ATP (36). During growth on fermentable sugars, the TCA cycle is a mitochondrial pathway, with acetyl-CoA resulting from oxidative decarboxylation of pyruvate by the pyruvate dehydrogenase (PDH) complex. To replenish use of TCA-cycle intermediates for biosynthesis, the cycle's acceptor molecule oxaloacetate can be imported from the cytosol, where it is produced by carboxylation of pyruvate. The nine biochemical reactions of the TCA cycle involve 22 mitochondrial enzymes, which show little genetic redundancy. Two reactions are catalyzed by single enzymes, Mdh1 and Fum1, while three steps are catalyzed by complexes of two to five proteins. Deletion of genes encoding individual subunits of the  $\alpha$ -ketoglutarate and succinyl-CoA synthetase complexes renders the complexes dysfunctional (37-42). In contrast, the succinate dehydrogenase complex, in which four functions are performed by seven proteins, does show some redundancy. *SDH1*, *SDH3* and *SDH4* have *SDH1b*, *SHH3*, and *SHH4*, respectively, as homologs originating from the WGD, while *SDH2* is a unique gene



**Figure 3 - Deletion strategy and specific growth rates of resulting strains.** Construction workflow for relevant *S. cerevisiae* strains (A) and their respective specific growth rates measured in shake flask rate on synthetic medium with glucose (SMD) or ethanol (SME) as carbon source, supplemented with uracil (B). Specific growth rates represent average and standard deviation of measurements on independent duplicate cultures for each strain and are expressed as a percentage of specific growth rate of the naïve uracil auxotrophic reference strain *S. cerevisiae* IMX581 or the naïve uracil prototrophic reference strain *S. cerevisiae* CEN.PK113-7D. Significant differences in specific growth rate relative to the control strain are indicated with a \* (two-tailed paired homoscedastic t-test  $p < 0.05$ ).

(43-48). Deletion of *SDH1b*, *SHH3* and *SHH4* has minor or no effect on complex integrity and yeast physiology and these genes are considered as functionally redundant (45,48,49). Citrate synthase (Cit1 and Cit3) and isocitrate dehydrogenase (Idh1, Idh2 and Idp1) have functionally redundant mitochondrial enzymes. Based on expression data and lack of a phenotypic difference during fermentative and respiratory growth, Cit3 and the NADP<sup>+</sup>-dependent Idp1 were considered to be redundant (28,50-52). Single deletion of *ACO1* or *ACO2*, which encode aconitase isoenzymes, causes amino acid auxotrophies (53,54). Idh1 and Idh2 are part of a complex and are both required for isocitrate dehydrogenase activity (37,38). Based on this information, only five of the 22 TCA cycle mitochondrial proteins were considered functionally redundant and, therefore, selected as candidates for elimination: Cit3, Idp1, Sdh1b, Shh3 and Shh4. *IDP1* was targeted in a later deletion round, along with extra-mitochondrial paralogs of the TCA-cycle that are part of the glyoxylate cycle.

There are several enzymes that form an interface between the TCA cycle and glycolysis. The WGD paralog pair *PYC1* and *PYC2* encode isoenzymes of the anaplerotic enzyme pyruvate carboxylase. Transcript levels of these two highly similar genes (92 %, Supplementary Table S1) are condition-dependent, and despite some conflicting reports on the physiological impact of *PYC1* and *PYC2* deletion (28,55-57), one study shows that only deletion of *PYC1* leads to aspartate auxotrophy (57). *PYC2* was therefore deleted. Deletion of *MAE1*, which encodes a mitochondrial malic enzyme catalyzing the oxidative decarboxylation of malate to pyruvate, does not show a clear phenotype. However, double deletion of *MAE1* and *PYK2* reduces the specific growth rate on ethanol by 62 % (58). As *PYK2* was deleted in the MG strain, *MAE1* was retained. The gluconeogenic enzymes PEP carboxykinase (Pck1) and fructose-1,6-bisphosphatase (Fbp1) are essential for bypassing the irreversible pyruvate-kinase and phosphofructokinase reactions, respectively, during growth on non-fermentable carbon sources (59,60).

*CIT3*, *SDH1b*, *SHH3*, *SHH4* and *PYC2* were deleted in two consecutive transformations rounds in the naïve reference strain and the CCMin1 strain (glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup>), resulting in IMX1805 (tca<sup>min</sup>) and IMX1806 (CCMin2: glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup>), respectively. Both strains grew as well as their parental strains in chemically defined medium supplemented with glucose or ethanol (Figure 3).

### Mitochondrial Carriers

The 35 nuclearly encoded mitochondrial carriers (MCs) mediate transport of numerous metabolites, nucleotides, cofactors and inorganic anions between mitochondrial matrix and cytosol (61). Based on extensive functional analysis studies (62,63), 19 MCs involved in transport of pyruvate, TCA-cycle intermediates, CoA, ADP, ATP, Pi, NAD<sup>+</sup>, FAD and thiamine pyrophosphate (cofactor of pyruvate dehydrogenase and  $\alpha$ -ketoglutarate dehydrogenase) were considered to be part of CCM (Supplementary Table S2, Figure 2). Potential genetic redundancy was identified for 10 of these MCs, with protein sequence similarity varying between 51 and 87 % (Supplementary Table S1). In addition to genetic redundancy, functional redundancy has to be considered since several genetically distinct transporters can transport the same solutes, as exemplified by the antiport of ADP and ATP across the mitochondrial membrane by three Aac homologs as well as by Sal1. Aac2 and Aac3 originate from WGD while Aac1 does not. Sal1 shares no homology with the Aac carriers and harbors an additional Ca<sup>2+</sup>-binding domain (64,65). Several studies indicate Aac2 as major paralog, whose presence suffices to sustain adenine nucleotide transport during respiratory growth (28,64-69). *AAC1*, *AAC3* and *SAL1* were therefore all candidates for deletion. NAD<sup>+</sup>, synthesized in the cytosol and required for the NAD<sup>+</sup>-dependent mitochondrial dehydrogenases in CCM, is imported by two MCs encoded by *NDT1* and *NDT2*, paralogs with 51 % similarity at the protein level (Supplementary Table S1). *NDT1* and *NDT2* are individually dispensable for growth on glucose or ethanol, but deletion of both precludes growth on non-fermentable carbon sources (70). Therefore, only one of the paralogs, *NDT2* was chosen for deletion. Since import of FAD, CoA and thiamine pyrophosphate is crucial for mitochondrial activity, the corresponding unique genes (*FLX1*, *LEU5* and *TPC1*) were retained (71-74).

Pyruvate is located at the interface of glycolysis and the TCA cycle and, in addition, mitochondrial pyruvate is required for synthesis of branched-chain amino acids (BCAA). Pyruvate import into mitochondria is mediated by three isoenzymes: Mpc1, Mpc2 and Mpc3. Mpc1 is constitutively expressed and forms complexes with either of the highly homologous Mpc2 or Mpc3 isoenzymes (75). *MPC2* is expressed during fermentative growth, while *MPC3* is expressed during respiratory growth. Deletion of *MPC2* leads to a severe growth defect, even in glucose-containing medium supplied with BCAA, while *MPC3* deletion leads to a modest (20 %) decrease of specific growth rates on non-fermentable carbon sources (76,77). Based on these literature data, it was decided to delete *MPC3*.

Sfc1 and Dic1 employ different mechanisms to import succinate into mitochondria and, since both are essential for growth on ethanol (78-80), neither were eliminated. Oxaloacetate is mainly transported by Oac1, whose removal only has a minor impact on specific growth rate on glucose medium, which is linked to its secondary function

as an exporter of  $\alpha$ -isopropylmalate for leucine biosynthesis (81,82). Since we observed a 26 % reduction of the specific growth rate on glucose upon deletion of *OAC1* in the CEN.PK genetic background used in this study (Supplementary Figure S1), it was retained in strain construction.

Four additional and partially functionally redundant MCs with different transport mechanisms and affinities mediate organic acid transport. Ctp1 is a citrate/malate antiporter, the paralogous carriers Odc1 and Odc2 are  $\alpha$ -ketoglutarate/oxodicarboxylate antiporters and Yhm2 exchanges  $\alpha$ -ketoglutarate and citrate, thereby enabling NADPH shuttling between cytosol and mitochondria (involving isocitrate dehydrogenase and aconitase) (83-89). Deletion of *CTP1* or double deletion of *ODC1* and *ODC2* does not affect growth, while triple deletion of *YHM2*, *ODC1* and *ODC2* does (84,85). Based on these literature data, *CTP1*, *ODC1* and *ODC2* were selected for deletion, with the realization that their combined deletion might affect di- and tri-carboxylic acids trafficking.

In total, eight MCs were targeted for elimination. First, *AAC1*, *AAC3*, *SAL1* and *MPC3* were simultaneously deleted, followed by simultaneous deletion of *NDT2*, *CTP1*, *ODC1* and *ODC2*. Deletion of *AAC1*, *AAC3*, *SAL1* and *MPC3* in the naïve reference strain, resulting in strain IMX2360 only marginally affected specific growth rate on SMD (3-5 % decrease), but had a stronger impact on growth on SME (14-18 % slower growth, Figure 3). These results are in agreement with the reported roles of these MCs in respiratory growth. When introduced in CCMin2 (*glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup>*), resulting in strain IMX1984, the same set of deletions did not affect specific growth rate on either SMD or SME.

Combined deletion of *NDT2*, *CTP1*, *ODC1* and *ODC2* reduced specific growth rate on SMD by 23 % in the naïve reference strain (resulting in strain IMX2230) and 30 % in IMX1984 (*glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup> aac1 $\Delta$  aac3 $\Delta$  sal1 $\Delta$  mpc3 $\Delta$* ; resulting in strain IMX2231) (Figure 3). When, instead, only *NDT2*, *CTP1* and *ODC1* were deleted in the naïve reference strain (resulting in strain IMX2404) or in engineered background strain IMX1984 (resulting in strain IMX2407, called CCMin3: *glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup> mc<sup>min</sup>*), specific growth rate on SMD was not affected and only a small (3-7 %) reduction of growth rate was observed on SME (Supplementary Figure S2 and Figure 3). CCMin3 (*glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup> mc<sup>min</sup>*), retained 96 % of the specific growth rate of the reference strain on SMD and 82 % of its specific growth rate on SME.

#### Fumarate reductases, acetyl-CoA synthesis and glyoxylate cycle

Cytosolic (Frds1) and mitochondrial (Osm1) fumarate reductases re-oxidize FADH<sub>2</sub> which has been proposed to be important for protein folding under anaerobic conditions (90-93). Double deletion of *FRDS1* and *OSM1* has no phenotypic effect on

complex glucose medium under aerobic conditions (94). However, Osm1 has a moonlighting function outside CCM, as it contains two translation sites, leading to the targeting to the ER of an Osm1 variant. Therefore, only *FRDS1* was considered for deletion in the design of a minimal CCM strain.

The glyoxylate cycle, which is essential for providing biosynthetic precursors with more than 2 carbon atoms during growth on fatty acids and two-carbon compounds, encompasses reactions in the peroxisome and cytosol (95,96) and uses acetyl-CoA as substrate made by the acetyl-CoA synthesis pathway. Ethanol is converted into acetyl-CoA via alcohol dehydrogenase (already reduced in the MG strain), acetaldehyde dehydrogenases and acetyl-CoA synthetases. Five acetaldehyde dehydrogenase isoenzymes, Ald2 to Ald6, oxidize acetaldehyde to acetate with either NADP<sup>+</sup> or NAD<sup>+</sup> as cofactor. The mitochondrial isoenzymes Ald4 and Ald5, required for growth on ethanol (97,98) and for maintenance of a functional respiratory chain (98), were both retained. Ald6 is the major cytosolic isoenzyme, whose elimination strongly affects growth on fermentable and non-fermentable carbon sources (99). The other two cytosolic acetaldehyde dehydrogenases, Ald2 and Ald3 are involved in conversion of 3-aminopropanal to  $\beta$ -alanine for pantothenic acid biosynthesis (99,100). As single deletion of *ALD2* or *ALD3* does affect growth on ethanol or glucose and *ALD2* is the major paralog in pantothenic acid production (100,101), *ALD3* was considered for deletion.

Acetate is then converted to acetyl-CoA via Acs1, whose localization is under debate. Acs1 has been reported to occur in the cytosol, the nucleus and in peroxisomes, depending on growth conditions (102,103). Acs1 and its isoenzyme Acs2 are essential for growth on non-fermentable and fermentable carbon sources, respectively (103). The mitochondrial acetyl-CoA hydrolase Ach1, is also able to convert acetate into acetyl-CoA, but uses succinyl-CoA as CoA donor. Deletion of *ACH1* leads to reduced chronological lifespan, severe mitochondrial damage and accumulation of reactive oxygen species (104,105). *ACS1*, *ACS2* and *ACH1* were therefore retained.

The glyoxylate cycle is initiated by Cit2, an extramitochondrial isoenzyme of the mitochondrial Cit1 and Cit3 citrate synthases, whose localization is under debate and has been reported in the cytosol and peroxisome (103,106). Citrate is then converted into isocitrate in the cytosol by the dually localized enzyme Aco1 in cytosol and mitochondria (107). Via a series of cytosolic and peroxisomal reactions (some localizations under debate), including: the isocitrate lyase Icl1 (cytosol) (108,109), the malate synthase (Mls1/Mls2 cytosol and peroxisome) (103,110,111) and malate dehydrogenase (Mdh2 in cytosol and Mdh3 occurs in the peroxisome) (111), the net synthesis of TCA-cycle intermediates is enabled from acetyl-CoA. Possible redundancies of glyoxylate enzymes also involved in TCA cycle were discussed above,

with only Cit3 selected for elimination in a minimal CCM strain. Since the glyoxylate cycle enzymes Cit2, Mls1, Icl1 and Mdh2 (cytosolic isoenzyme of the mitochondrial Mdh1) are either essential for growth on C<sub>2</sub>-compounds or their elimination leads to strong reductions in growth rate, they were retained in the minimal CCM design (95,103,112-116). The proteins Icl2, Mls2 and Mdh3 are homologous to Icl1, Mls1 and Mdh1/Mdh2, respectively, but have (additional) functions outside CCM (113,117,118) and were therefore also retained.

The peroxisomes harbor the NADP<sup>+</sup>-dependent isocitrate dehydrogenase Idp3. Deletion of its mitochondrial homologs Idp1 and Idp2 does not affect growth on ethanol or glucose (51,52). Idp1 and Idp2 were therefore the only genes considered for elimination in the minimal CCM design.

Triple deletion of *FRDS1*, *IDP1* and *IDP2* in the naïve reference strain (resulting in strain IMX2470: fum<sup>min</sup> glyox<sup>min</sup>) did not affect specific growth rate on SMD and caused a 7 % lower growth rate on SME (Figure 3 and Supplementary Figure S2). Deletion in CCMin3 (IMX2407: glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup> mc<sup>min</sup>) did not affect specific growth rate on either SMD or SME (Figure 3). The resulting strain CCMin4 (IMX2475: glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup> mc<sup>min</sup> fum<sup>min</sup> glyox<sup>min</sup>) retained 94 % and 79 % of the specific growth rate of the naïve reference strain IMX581, on SMD and SME, respectively. For reasons of experimental efficiency, *ALD3* was removed in the final deletion round (see below).

### Glycerol synthesis

Glycerol production is essential for redox balancing in anaerobic *S. cerevisiae* cultures (119). In addition, glycerol plays a key role in osmotolerance and maintenance of cellular volume and turgor pressure during growth under hypertonic conditions (120,121). The conversion of dihydroxyacetone phosphate to glycerol-3-phosphate is catalyzed by the isoenzymes Gpd1 and Gpd2. *gpd1* deletion mutants are osmosensitive, but show no growth defects in the absence of stress (122-124). In contrast, *gpd2* null mutants show a mild reduction of aerobic growth rates and strongly decreased growth rates under anaerobic conditions (125,126). Glycerol-3-phosphate is converted into glycerol by the redundant Gpp1 and Gpp2 isoenzymes. Single deletion of either enzyme neither affects osmotolerance nor growth on glucose or ethanol, while *gpp1* mutants have been reported to show extended lag phases in anaerobic cultures (127,128). Therefore, *GPD1* and *GPP2* were chosen for deletion.

Triple deletion of *ALD3*, *GPD1* and *GPP2* did not significantly affect specific growth rate on SMD, while a small growth rate reduction was observed on SME in both the naïve reference strain and the engineered background (IMX2509: Ace<sup>min</sup> Glycerol<sup>min</sup> and IMX2519/CCMin5: glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup> mc<sup>min</sup> fum<sup>min</sup> glyox<sup>min</sup> Ace<sup>min</sup> Glycerol<sup>min</sup>, respectively Figure 3). The lower specific growth rate on SME could be attributed to

the double deletion of *GPD1* and *GPP2* (strain IMX2612, Figure 3 and Supplementary Figure S2).

The auxotrophic 35-deletion strain IMX2519 (*glyc*<sup>min</sup> *fer*<sup>min</sup> *ppp*<sup>min</sup> *tca*<sup>min</sup> *mc*<sup>min</sup> *fum*<sup>min</sup> *glyox*<sup>min</sup> *Ace*<sup>min</sup> *Glycerol*<sup>min</sup>) grew at 93 % and 71 % of the specific growth rate of control strain IMX581 on uracil-supplemented SMD and SME, respectively. Integration of a *URA3* cassette yielded the uracil-prototrophic 35-deletion strain IMX2538, which was labelled ‘minimal CCM strain’. This prototrophic strain grew at 94 % of the prototrophic control strain with a full set of CCM genes CEN.PK113-7D on SMD and at 76 % on SME (Figure 3). These values were within the 25 % boundary that were initially set, and the physiology of the minimal CCM strain was further explored.

### **A *S. cerevisiae* strain with minimalized CCM shows only mild growth defects on synthetic media**

The genome sequence of the minimal CCM strain was analyzed by short-read and long-read techniques. Long-read sequencing revealed that 9 transformation rounds and deletion of 22 genes from the MG strain had not led to chromosomal rearrangements or deletions. Previously reported duplicated regions on chromosomes 3 and 5 of the MG strain based on karyotyping and short read sequencing (22) were also observed in this study with long-read sequencing. Sequence analysis confirmed that all 22 targeted CCM genes were correctly deleted from the MG strain. The genome of the minimal CCM strain showed 45 Single Nucleotide Polymorphisms (SNPs) relative to the MG strain, of which eight were located in genes and only four led to an amino acid change (Table 1), none of which affected proteins involved in CCM.

**Table 1 - Single-nucleotide mutations identified in coding regions of the prototrophic minimal CCM strain IMX2538.** Single-nucleotide changes in *S. cerevisiae* IMX2538 (prototrophic minimal CCM) relative to the genome sequence of *S. cerevisiae* IMX372 (prototrophic minimal glycolysis (MG)) (22).

<b>Systematic name</b>	<b>Name</b>	<b>Type*</b>	<b>Amino acid change</b>
YBR114W	<i>RAD16</i>	NS	Ile-202-Thr
YDL035C	<i>GPR1</i>	S	Asn-523-Asn
YDR098C	<i>GRX3</i>	NS	Glu-239-Asp
YFL062W	<i>COS4</i>	S	Cys-151-Cys
YLR002C	<i>NOC3</i>	NS	Asp-526-Glu
YML058W	<i>SML1</i>	NS	Gly-52-Ser
YMR154C	<i>RIM13</i>	S	Lys-265-Lys
YNL273W	<i>TOF1</i>	S	Asn-117-Asn

\*S: synonymous, NS: non-synonymous

**Table 2 - Physiological characterization of a 35-deletion, minimal CCM prototrophic *S. cerevisiae* strain in aerobic bioreactor batch cultures.** *S. cerevisiae* strains were grown at pH 5.0 and at 30 °C in aerobic bioreactors on synthetic medium with glucose as sole carbon source. Data are presented as average and standard deviation of 3 biological replicates for *S. cerevisiae* strains CEN.PK113-7D (naïve reference) and IMX2538 (minimal CCM). Data for *S. cerevisiae* IMX372 (minimal glycolysis) were recalculated from the raw data of Solis-Escalante *et al.* (22) obtained with two biological replicates. Statistical significance with respect to CEN.PK113-7D is represented in bold and with an \* (two-tailed t-test, equal variances,  $P < 0.05$ ).

	CEN.PK113-7D (Naïve reference)	IMX372 (Minimal Glycolysis)	IMX2538 (Minimal CCM)
<b>Glucose phase</b>			
$\mu_{\max}$ (h <sup>-1</sup> )	0.37 ± 0.00	<b>0.38 ± 0.01*</b>	<b>0.34 ± 0.00*</b>
$q_s$ (mmol.g <sub>DW</sub> <sup>-1</sup> .h <sup>-1</sup> )	-16.2 ± 0.2	-15.7 ± 0.7	-15.4 ± 0.5
$q_{\text{Ethanol}}$ (mmol.g <sub>DW</sub> <sup>-1</sup> .h <sup>-1</sup> )	23.5 ± 1.5	23.1 ± 1.1	23.2 ± 2.1
$q_{\text{Glycerol}}$ (mmol.g <sub>DW</sub> <sup>-1</sup> .h <sup>-1</sup> )	1.52 ± 0.05	1.40 ± 0.02	<b>1.11 ± 0.05*</b>
$q_{\text{Acetate}}$ (mmol.g <sub>DW</sub> <sup>-1</sup> .h <sup>-1</sup> )	0.44 ± 0.03	<b>0.82 ± 0.04*</b>	<b>0.71 ± 0.02*</b>
$q_{\text{CO}_2}$ (mmol.g <sub>DW</sub> <sup>-1</sup> .h <sup>-1</sup> )	23.4 ± 0.2	-	22.6 ± 0.5
$q_{\text{O}_2}$ (mmol.g <sub>DW</sub> <sup>-1</sup> .h <sup>-1</sup> )	-6.8 ± 0.4	-	-7.0 ± 0.2
$Y_{\text{biomass/glucose}}$ (g <sub>DW</sub> .g <sub>glucose</sub> <sup>-1</sup> )	0.13 ± 0.00	0.13 ± 0.00	0.12 ± 0.00
$Y_{\text{ethanol/glucose}}$ (mol.mol <sup>-1</sup> )	1.45 ± 0.09	1.48 ± 0.01	1.51 ± 0.09
$Y_{\text{glycerol/glucose}}$ (mol.mol <sup>-1</sup> )	0.09 ± 0.00	0.09 ± 0.01	<b>0.07 ± 0.00*</b>
$Y_{\text{acetate/glucose}}$ (mol.mol <sup>-1</sup> )	0.03 ± 0.00	0.05 ± 0.00	<b>0.05 ± 0.00*</b>
<b>Post-diauxic phase</b>			
$\mu_{\max}$ (h <sup>-1</sup> )	0.10 ± 0.00	<b>0.12 ± 0.00*</b>	<b>0.08 ± 0.01*</b>
$q_{\text{Ethanol}}$ (mmol.g <sub>DW</sub> <sup>-1</sup> .h <sup>-1</sup> )	-3.10 ± 0.19	<b>-3.93 ± 0.04*</b>	-3.07 ± 0.31

The physiology of the minimal CCM strain was compared to that of the congenic reference strain CEN.PK113-7D, which has a full complement of CCM genes, in pH-controlled aerobic bioreactor cultures on SMD. Consistent with the analyses in shake flasks, the specific growth rate of the minimal CCM strain in these cultures was 8 % lower than the rate of the reference strain CEN.PK113-7D (Table 2). During the glucose consumption phase, biomass-specific glucose and oxygen consumption rates of the two strains, as well as their ethanol and CO<sub>2</sub> production rates and their biomass and ethanol yields on glucose were also similar. The minimal CCM strain did exhibit a higher acetate production rate and yield (63 % and 71 % higher, respectively) than the reference strain, a difference already observed for the MG strain (22). Similarly, a lower glycerol production rate and glycerol yield on glucose (27 % and 23 % lower, respectively), was in line with data reported for a *gpd1* deletion mutant (122). After the diauxic shift, growth of the minimal CCM strain on ethanol, glycerol and organic acids produced during the glucose phase, proceeded at a 17 % lower rate than observed for the reference strain (Table 2). As a macroscopic characterization based on extracellular products might mask subtle differences of intracellular fluxes,

intracellular concentrations of CCM intermediates were measured during the mid-exponential growth phase on glucose. The concentrations of these metabolites hardly differed between the minimal CCM and the control strains (Table 3). These results indicated that a 32 % reduction of the complement of genes encoding CCM enzymes of *S. cerevisiae* had only a small impact on its physiology under standard laboratory conditions.

**Table 3 - Intracellular metabolite profiles of a 35-deletion, minimal CCM prototrophic *S. cerevisiae* strain in aerobic bioreactor batch cultures.** Intracellular metabolite contents were measured during the mid-exponential glucose phase of aerobic bioreactor batch cultures of *S. cerevisiae* CEN.PK113-7D (naïve reference) and IMX2538 (minimal CCM) (see Table 2 for other physiological data). Data represent average and standard deviation of data from analyses on three independent cultures for each strain. Fold differences that are statistically significant are indicated in bold and with an \* (two-tailed t-test, equal variances, P<0.05). # Below detection limit.

Metabolite	CEN.PK113-7D (naïve reference)	IMX2538 (Minimal CCM)	Fold difference
$\mu\text{mol} \cdot (\text{g biomass dry weight})^{-1}$			
<b>Glycolysis</b>			
Glucose 6-phosphate	4.13 ± 0.49	5.01 ± 0.38	1.2
Fructose 6-phosphate	0.38 ± 0.06	0.59 ± 0.08	<b>1.5*</b>
Fructose 1,6-bisphosphate	17.53 ± 0.82	20.52 ± 0.66	<b>1.2*</b>
Glyceraldehyde 3-phosphate	0.15 ± 0.01	0.19 ± 0.01	<b>1.3*</b>
Dihydroxyacetone phosphate	0.79 ± 0.13	2.08 ± 0.66	2.6
3-Phosphoglycerate	1.30 ± 0.12	2.52 ± 0.28	<b>1.9*</b>
2-Phosphoglycerate	0.16 ± 0.01	0.17 ± 0.01	1.0
Phosphoenolpyruvate	0.25 ± 0.01	0.12 ± 0.01	<b>0.5*</b>
<b>Trehalose synthesis</b>			
Trehalose 6-phosphate	9.34 ± 1.18	14.48 ± 1.67	<b>1.6*</b>
Trehalose <sup>#</sup>	0.20 ± 0.04	0.41 ± 0.09	<b>2.1*</b>
<b>Pentose phosphate pathway</b>			
6-Phosphogluconate	1.01 ± 0.02	1.11 ± 0.07	1.1
Ribose 5-phosphate	0.40 ± 0.02	0.51 ± 0.03	<b>1.3*</b>
Ribulose 5-phosphate	0.20 ± 0.03	0.33 ± 0.06	<b>1.6*</b>
Xylulose 5-phosphate	0.43 ± 0.07	0.72 ± 0.13	1.7
Sedoheptulose 7-phosphate	0.50 ± 0.06	0.54 ± 0.04	1.1
Erythrose 4-phosphate	0.004 ± 0.000	0.005 ± 0.000	<b>1.3*</b>
<b>TCA cycle and Glyoxylate cycle</b>			
Citrate	4.56 ± 0.65	5.80 ± 0.98	1.3
Isocitrate	0.02 ± 0.00	0.03 ± 0.00	<b>1.7*</b>
$\alpha$ -ketoglutarate	0.29 ± 0.02	0.67 ± 0.24	2.3

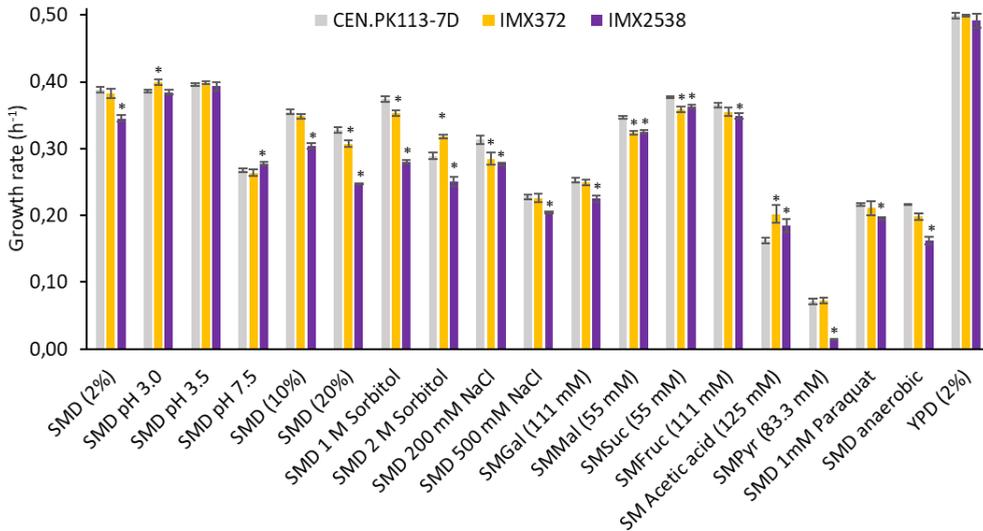
<b>Succinate</b>	0.37 ± 0.03	0.37 ± 0.12	1.0
<b>Fumarate</b>	0.10 ± 0.02	0.18 ± 0.06	1.8
<b>Malate</b>	0.60 ± 0.08	1.03 ± 0.19	<b>1.7*</b>
<b>Nucleotides and cofactors</b>			
<b>Energy charge</b>	0.86 ± 0.00	0.88 ± 0.01	1.0
<b>AMP</b>	0.13 ± 0.01	0.17 ± 0.01	<b>1.3*</b>
<b>ADP</b>	1.55 ± 0.08	1.43 ± 0.11	0.9
<b>ATP</b>	5.50 ± 0.16	5.58 ± 0.21	1.0
<b>UDP</b>	0.23 ± 0.01	0.22 ± 0.05	0.9
<b>UTP</b>	1.61 ± 0.06	1.57 ± 0.10	1.0
<b>GDP</b>	0.27 ± 0.02	0.21 ± 0.07	0.8
<b>GTP</b>	1.58 ± 0.06	1.49 ± 0.15	0.9
<b>CDP</b>	0.13 ± 0.02	0.15 ± 0.01	1.1
<b>CTP</b>	0.84 ± 0.02	0.82 ± 0.03	1.0
<b>NADH</b>	0.18 ± 0.02	0.15 ± 0.02	0.8
<b>NAD<sup>+</sup></b>	85.43 ± 1.90	91.82 ± 2.64	<b>1.1*</b>
<b>NADP<sup>+</sup></b>	8.49 ± 0.67	6.87 ± 0.25	<b>0.8*</b>
<b>Acetyl-CoA</b>	5.80 ± 0.18	5.38 ± 0.34	0.9
<b>FAD</b>	0.87 ± 0.12	0.54 ± 0.18	0.6

### Dissecting individual from synergistic responses to growth on a range of conditions

To explore genetic redundancy of CCM genes, the minimal CCM strain and the congenic reference strain CEN.PK113-7D were grown under a broad range of conditions. Some of these were chosen based on previously reported phenotypes (*e.g.* high osmolarity) or connection to CCM (*e.g.* growth on various carbon sources), while others subjected the strains to adverse conditions (*e.g.* acidic or alkaline pH).

Consistent with reports that deletion of *GPD1* causes decreased osmotolerance (123,125), the minimal CCM strain grew 13 to 25% slower than the reference strain exposed to high osmolarity, which was imposed by adding high concentrations of sorbitol (1M and 2M) or glucose (10 % and 20 % w/v). Construction and analysis of strains with different combinations of deletions in reference and CCM-minimization backgrounds confirmed that this growth reduction specifically resulted from *GPD1* deletion, rather than from synergistic effects (Figure 4 and Supplementary Figure S3). *MPC3* deletion has been reported to cause lower growth rates on glycerol or lactate as sole carbon source (77). Since *Mpc3* is a pyruvate transporter, growth was assessed directly on chemically defined medium with pyruvate as sole carbon source. The minimal CCM strain grew 79% slower than the control strain, however this could

surprisingly not be attributed to the *MPC3* deletion (Figure 4 and Supplementary Figure S3).



**Figure 4 - Specific growth rates of 35-deletion, prototrophic minimal CCM strain under a broad range of growth conditions.** Specific growth rates of the prototrophic *S. cerevisiae* strains CEN.PK113-7D (naive reference strain), IMX372 (minimal glycolysis (MG)) and IMX2538 (minimal CCM) under different growth conditions. Specific growth rates were measured in triplicate cultures using a growth profiler, except for those in SMPyr and SMD-anaerobic, which were measured in independent duplicate shake flask cultures. Abbreviations indicate the following growth conditions: SM, synthetic medium; SMD, synthetic medium with glucose; Gal, galactose as carbon source; Mal, maltose as carbon source; Suc, sucrose as carbon source; Fruc, fructose as carbon source; Pyr, pyruvate as carbon source; YPD, complex medium with glucose. Significant changes in growth rate of IMX372 and IMX2538 with respect to CEN.PK113-7D are indicated with a \* (two-tailed paired homoscedastic t-test  $p < 0.05$ ).

Combined deletion of *AAC1*, *AAC3*, *MPC3* and *SAL1* caused a 3 – 5 % and a 14 – 18 % decrease of growth rates on SMD and SME, respectively, thus marking their importance on these carbon sources. According to previous reports, individual deletion of these four genes does not affect growth on glucose and individual deletion of *AAC1*, *AAC3* and *SAL1* does not affect growth on ethanol (64,65,77,129). Deletion of *MPC3* has been reported to cause a decrease in growth rate on glycerol and lactate (77), and may therefore also be responsible for the lower growth rate on ethanol. Reintroduction of *MPC3* in strain IMX1984 (*glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup> aac1Δ aac3Δ sal1Δ mpc3Δ*) increased specific growth rate on SME by only 3 %, while reintroduction in IMX2519 (CCMin 5: *glyc<sup>min</sup> fer<sup>min</sup> ppp<sup>min</sup> tca<sup>min</sup> mc<sup>min</sup> fum<sup>min</sup> glyox<sup>min</sup> Ace<sup>min</sup> Glycerol<sup>min</sup>*) did not affect growth rate (Supplementary Figure S3). These results suggest that the observed impact of quadruple deletion of *AAC1*, *AAC3*, *MPC3* and *SAL1* on ethanol growth was caused by synergistic effects.

Some paralogs have been reported to be specifically important under anaerobic conditions (*AAC3*, *FRDS1*) (94,129). In line with these reports, while the MG strain showed the same growth rate as CEN.PK113-7D under anaerobic conditions, the minimal CCM strain showed a 25 % lower anaerobic growth rate (Figure 4). However, testing of deletions in the reference background indicated that this difference was not caused by the deletions of *AAC3* or *FRDS1* (Supplementary Figure S3).

Over the broad range of conditions tested, including several stress conditions, few differences in specific growth rate were observed (Figure 4). Combining the phenotype of strains with individual and clustered deletions enabled to identify synergistic interactions between minor paralogs.

## Discussion

Genetic reduction has been applied in several microorganisms (1), including *B. subtilis* (3), *E.coli* (2), *Lactococcus lactis* (6), *C. glutamicum* (4), *Streptomyces* species (7), *Pseudomonas* species (5) and *Schizosaccharomyces pombe* (130), with the purpose of discovering a minimal genome content and/or for engineering efficient cell factories. In *S. cerevisiae*, Marakami *et al.* reduced genome content by 5 % by deleting 15 terminal chromosomal regions (131). Moreover, the creation of a synthetic yeast genome in the Sc2.0 project is accompanied by an 8 % genome reduction by deletion of long terminal repeats, retrotransposons and introns; and engineering of a single chromosome yeast strain was characterized by a 9 % decrease in DNA (132,133). Genome reduction studies typically target two types of DNA sequences: non-expressed DNA (cryptic genes, mobile DNA) and irrelevant/non-essential genes. These DNA elements can be targeted by random strategies for which little knowledge is required, such as transposon mutagenesis or the elegant SCRaMble technique used for the recent reduction of left synthetic chromosome arm XII in *S. cerevisiae* (21). In the present study, knowledge-based reduction of the gene complement for CCM in *S. cerevisiae* was informed by gene expression data and previous phenotypic analysis on single-knockout mutants (21,131-133).

In this study, we reduced genetic complexity of central carbon metabolism (CCM) in *S. cerevisiae* by deletion of 35 genes encoding enzymes and transporters. This reduction corresponded to elimination of 32 % of the (iso)enzymes and transporters involved in the included processes, without major impacts on strain physiology, which was tested under a broad range of conditions (Figure 2-4). The present study built on earlier work by Solis-Escalante *et al.* (22), who eliminated 50 % of isoenzymes involved in glycolysis and ethanolic fermentation with a similar small impact on physiology. The attainable reduction of gene sets for enzymes and transporters involved in other CCM pathways differed, with 50 % for fumarate reduction and glycerol synthesis, 37 % for

the mitochondrial carriers, 36 % for the pentose-phosphate pathway, 23 % for the TCA cycle, 14% for the glyoxylate cycle, 20 % for the glycolysis – TCA cycle interface plus gluconeogenesis, and 8 % for acetyl-CoA metabolism. The lower attainable genetic reduction of the four latter pathways can be largely attributed to neofunctionalization and relocalization of enzymes during evolution.

Our results show that yeast CCM is remarkably robust to genetic reduction, in particular during growth on glucose, yeast's favorite carbon source, but also when challenged by a broad range of growth conditions. Notable exceptions were growth on pyruvate (79 % growth rate reduction), anaerobic growth on glucose (25 % reduction), growth on ethanol (24 % reduction) and growth at high osmolarity (between 13 % and 25 % lower specific growth rates). Growth-rate reductions on ethanol and at high osmolarity could be attributed to specific genes or gene combinations, while for growth on pyruvate and anaerobic growth some hypothetical targets could be excluded. The physiological role of most deleted paralogs remains elusive. Such a situation is exemplified by *TKL2* and *NQM1*, which are paralogs of the major PPP genes *TKL1* and *TAL1*, respectively. In *S. cerevisiae* strains engineered for l-arabinose utilization, their deletion was shown to lead to lower growth rates on this pentose sugar (33,134). Clearly, as pentoses are not natural carbon sources for *S. cerevisiae*, this role cannot have provided an evolutionary driving force for fixation of these paralogs in its genome, but it does indicate potential contribution to fitness under other, as yet unidentified growth conditions. Testing the minimal CCM strain under an even wider variety of environmental conditions, including dynamics in nutrient availability and other environmental parameters may reveal physiological roles of these and other paralogs. Alternatively, the mechanisms that fixed some paralogs during evolution may have been disrupted by relatively recent mutations or gene loss (135). Following this reasoning, absence of a clear phenotype of knock-out mutants may have captured a stage in the evolutionary trajectory of *S. cerevisiae* that will eventually lead to loss of a paralog, evolution towards complete sub-functionalization, or retention of functional overlap with asymmetric divergence (136).

In this first step towards the genetic minimization of CCM in yeast, choices had to be made on which pathways and genes were considered as part of the CCM and on criteria for redundancy. For instance, transport of NAD<sup>+</sup>, FAD<sup>+</sup>, ADP/ATP and Pi across the mitochondrial membrane was considered, while transport of NAD(P)H, which requires more complex shuttle systems (137,138) was not yet included. In addition, as *S. cerevisiae* cannot synthesize carnitine (139), the carnitine shuttle system transporting acetyl-CoA across compartments was not considered. Since *CRC1*, *CAT1*, *YAT1* and *YAT2* involved in this shuttle are dispensable for growth in the absence of carnitine (139), they can be considered for further genetic reduction of the CCM. Some genes required for anaerobic growth, such as *ADH3* (137,140) were also retained, but

could be removed if fast anaerobic growth is excluded as a criterion. Several other processes and pathways are of particular interest for development of strain platforms for modular engineering yeast CCM. In this context, glucose uptake, which involves a set of 20 hexose transporters (141) provides an interesting target for future experiments, whose minimization can benefit from a recently constructed Hxt<sup>0</sup> CRISPR kit (142). Another logical target for minimization is uptake and assimilation of (alternative) carbon sources and especially of maltose, whose metabolism is enabled by highly redundant subtelomeric genes (143,144).

Genetic reduction presents a first, indispensable step towards the construction of modular yeast strains for extensive remodeling of CCM. Current demands for economically competitive cell factories, with optimized titer, rate and yield (TRY), requires extensive remodeling of the CCM for the supply of precursors, (redox) cofactors and energy-rich molecules (145-147). For instance, the extensive remodeling of the native Entner-Doudoroff (ED) glycolytic pathway into the Embden-Meyerhof-Parnas (EMP) pathway improved carotenoid synthesis in *Pseudomonas putida* (148). Similarly, substantial efforts are invested in remodeling yeast CCM in *S. cerevisiae* to increase the supply of cytosolic acetyl-CoA, a precursor for a wide array of attractive biomolecules (Figure 1) (149). Also, production of complicated chemical structures like plant natural products in *S. cerevisiae* requires extensive remodeling of the entire central carbon metabolism (150-152). As demonstrated by Kuijpers *et al.* (24), genetic reduction facilitates the colocalization of sets of genes in 'pathway clusters' and strongly accelerates the genetic remodeling of these pathways. With this strategy, the 12 steps of glycolysis and ethanolic fermentation were rapidly and efficiently swapped with heterologous variants and enabled the implementation of an innocuous DNA and RNA watermarking method (153). A similar strategy can be considered for remodeling CCM, with the minimal CCM strain as starting point. As recently demonstrated, 44 transcriptional-unit sized DNA fragments can be assembled in *S. cerevisiae* into specialized, synthetic supernumerary chromosomes (26). Since the capacity of HR was not reached, assembly of synthetic chromosomes containing the set of 76 genes encoding the minimal CCM has now become a realistic objective. Subsequent CRISPR-Cas assisted removal of the duplicate CCM genes from their native locations could then generate powerful platforms for chromosome swapping and combinatorial CCM remodeling studies. The reduction of genetic complexity demonstrated in the present study therefore not only provides new insights in genetic redundancy of CCM, but also contributes to the eventual localization of all genes required for a minimized CCM on specialized, synthetic supernumerary chromosomes that allow for extensive, combinatorial remodeling of yeast metabolism for industrial applications.

## Materials and Methods

### Strains, media and maintenance

The *Saccharomyces cerevisiae* strains used in this study are all derived from the CEN.PK family (154,155) (Supplementary Table S3). The naïve, uracil auxotrophic and Cas9 containing strain, IMX581 (156) and the uracil auxotrophic MG (minimal glycolysis) strain IMX370 (22) were used for deletion of genes encoding enzymes or transporters involved in central carbon metabolism (CCM). The naïve uracil prototrophic strain CEN.PK113-7D was used for physiological comparison. Complex medium used for propagation of yeast strains consisted of 10 g L<sup>-1</sup> Bacto yeast extract, 20 g L<sup>-1</sup> Bacto peptone and 20 g L<sup>-1</sup> glucose (YPD), autoclaved at 110°C for 20 min. After transformation, yeast strains were selected in synthetic medium (SM) (157) containing 3.0 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 5.0 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 1.0 mL L<sup>-1</sup> trace elements autoclaved at 121°C for 20 min, whereafter 1.0 mL L<sup>-1</sup> of filter sterilized vitamin solution was added. Before autoclaving media were set to pH 6 by 1 M KOH addition. SM was supplemented with 20 g L<sup>-1</sup> glucose (SMD) or 2 % ethanol v/v (SME) for propagation and growth characterization. Synthetic medium was supplemented with 150 mg L<sup>-1</sup> uracil for uracil auxotrophic strains. For selection of transformants carrying the *amdS* selection marker (158), ammonium sulfate was replaced as nitrogen source with 10 mM of acetamide. For experiments on SM in the growth profiler and under anaerobic conditions, ammonium sulfate was replaced by 2.3 g L<sup>-1</sup> Urea. For both media in which ammonium sulfate was replaced, 6.6 g L<sup>-1</sup> K<sub>2</sub>SO<sub>4</sub> was added. Growth was performed in 500 mL shake flask containing 100 mL medium or in 100 mL shake flasks containing 20 mL medium at 30 °C and 200 rpm in an Innova 44 Incubator shaker (New Brunswick Scientific, Edison, NJ). Culture on solid media were incubated for 3-5 days at 30 °C.

CEN.PK113-7D, IMX372, IMX2538 and several intermediate strains were tested in the growth profiler on a variety of liquid media , containing: SM(urea) plus 2 % glucose (SMD), SMD at pH 3.0, 3.5 or 7.5, SM plus 10 % glucose (SMD (10 %)), SM plus 20 % glucose (SMD (20 %)), SMD plus 1 or 2M sorbitol, SMD plus 200 or 500 mM NaCl, SM plus 111 mM galactose (SMGal), SM plus 55 mM maltose (SMMal), SM plus 55mM sucrose (SMSuc), SM plus 111 mM fructose (SMFruc), SM plus 125 mM acetic acid, SMD plus 1 mM paraquat and YPD (2 % glucose). Growth on SM plus 83.3 mM pyruvic acid was performed in shake flasks. For anaerobic growth in shake flasks SMD was supplemented with 0.01 g L<sup>-1</sup> ergosterol and 0.42 g L<sup>-1</sup> Tween 80 dissolved in ethanol (SMD (2 %) anaerobic) (157).

Plasmids were propagated in and isolated from chemically competent *Escherichia coli* XL1-Blue cells, which were cultivated in Lysogeny Broth containing 10 g L<sup>-1</sup> Bacto

tryptone, 5.0 g L<sup>-1</sup> Bacto yeast extract and 5 g L<sup>-1</sup> NaCl supplemented with 100 mg L<sup>-1</sup> ampicillin (LB-amp) when required. *E. coli* was cultivated in 15 mL Greiner tubes containing 5 mL medium at 37 °C and 200 rpm in an Innova 4000 Incubator shaker (New Brunswick Scientific). Bacterial cultures on solid medium were incubated overnight at 37°C.

For solid medium 20 g L<sup>-1</sup> of agar was added before autoclaving. All *S. cerevisiae* and *E. coli* strains were stored at -80°C in 1 ml aliquots containing 30 % v/v glycerol in appropriate medium.

### **Molecular biology techniques**

Plasmids were isolated from *E. coli* using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MO) or the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, MA) according to the provided protocols. DNA fragments for plasmid construction or integrative DNA fragments used in yeast transformation were amplified using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) according to manufacturer's instructions, using PAGE-purified or desalted oligonucleotides (Sigma-Aldrich) depending on the application. Purification of genomic Polymerase Chain Reaction (PCR) amplified DNA was performed with the GenElute PCR Clean-Up kit (Sigma-Aldrich) or the GeneJET PCR Purification Kit (Thermo Fisher Scientific) if no a-specific products were present. When a-specific products were present or when DNA was amplified from plasmids, the DNA was purified by separation using electrophoresis on a 1 % (w/v) agarose gel (TopVision Agarose, Thermo Fisher Scientific) in 1x Tris-acetate-EDTA (TAE) buffer (Thermo Fisher Scientific) or on a 2 % (w/v) agarose gel (TopVision Agarose, Thermo Fisher Scientific) in 1x Tris-Borate-EDTA (TBE) buffer (Thermo Fisher Scientific) with subsequent purification with the Zymoclean Gel DNA Recovery kit (Zymo Research). Chemical transformation of *E. coli* XL1-Blue was performed by thawing of competent cells on ice, addition of DNA, followed by a heat shock for 40s at 42°C. Subsequently, cells were incubated on ice for 2 minutes and plated immediately on selective LB-amp plates and grown overnight at 37°C. Transformation of *S. cerevisiae* was performed using the lithium acetate/single-stranded carrier DNA/polyethylene glycol method (159). Single-cell lines were obtained by three consecutive re-streaks on selective solid medium. Yeast genomic DNA was extracted according to Looke *et al.* (160), the YeaStar Genomic DNA kit (Zymo Research, Irvine, CA) according to Protocol I supplied by the manufacturer or the QIAGEN Blood & Cell Culture Kit with 100/G or 20/G Genomic-tips (Qiagen, Hilden, Germany) following the manufacturer's recommendations. Verification of the accurate genotype of engineered *S. cerevisiae* strains and *E. coli* plasmids was done by diagnostic PCR before strain storage at -80°C. These diagnostic PCRs were performed using desalted oligonucleotides and the DreamTaq PCR master Mix (Thermo Fisher Scientific) according to the manufacturer's protocol.

## Plasmid and strain construction

Deletions were performed using CRISPR/Cas9. CRISPR/Cas9-based genome editing of *S. cerevisiae* was performed as described by Mans *et al.* (156) with minor alterations. Plasmids containing a single gRNA (Supplementary Table S4) were constructed via Gibson assembly with a backbone containing the marker cassette and one insert fragment containing the gRNA and the 2  $\mu$ m fragment. The backbone was amplified from a pMEL plasmid (156) with primers 5980 and 5792 and the insert fragment was amplified with a gRNA specific primer designed with the yeast restriction tool (156) and primer 5979 (primers in Supplementary Table S5). Plasmids containing two gRNAs were constructed using one backbone fragment and two insert fragments, each containing one gRNA and one half of the 2  $\mu$ m fragment. Backbones were PCR amplified from the pROS plasmids (156) with the double-binding primer 6005. Insert fragments were obtained with the gRNA specific primers together with either primer 5974 or primer 5975 (Primers in Supplementary Table S5). The backbone and gRNA insert fragment(s) were gel purified, DpnI digested (Thermo Fisher Scientific) and Gibson-assembled in a final volume of 5  $\mu$ L using NEBuilder HiFi DNA assembly master mix (NEB, Ipswich, MA), according to manufacturer's instructions. Assembled plasmids were transformed and subsequently isolated from LB-amp grown *E. coli*. Correct assembly was checked using diagnostic PCR (Supplementary Table S6).

IMK588 was constructed by integrating the *KanMX* marker at the *OAC1* locus of CEN.PK113-7D. The *KanMX* marker with homologous flanks to *OAC1* was amplified with primers 6358 and 6359 from pUG6 (Supplementary Table S4, S7 and S8).

In order to perform CRISPR editing in the MG strain (IMX370), *Cas9* was integrated by transforming a *Cas9* and *natNT2* DNA fragment, which can assemble by homologous recombination at the *CAN1* locus. The *Cas9* fragment (*can1* flank-*Cas9* expression cassette-SHR A) was PCR amplified with primers 2873 and 4653 from plasmid p414-*TEF1p-Cas9-CYC1t* (Supplementary Table S4 and S9). The *natNT2* fragment (SHR A-NatNT2 marker cassette-*can 1* flank) was amplified with primers 3093 and 5542 from plasmid pUG-*natNT2* (Supplementary Table S4 and S9). The *Cas9* containing MG strain was stocked as IMX1331.

For genome editing using CRISPR, *S. cerevisiae* strains were transformed with 1  $\mu$ g of each gRNA plasmid and 1  $\mu$ g of each 120 bp double-stranded DNA repair fragment. These repair fragments were made by annealing of complimentary oligonucleotides listed in Supplementary Table S7, and consisted 60 bp homology sequences immediately upstream of the start codon and downstream of the stop codon of the targeted gene, unless stated otherwise. Transformants were plated on selective medium. Gene deletion was verified by diagnostic colony PCR on randomly picked colonies by using the primers which bind outside of the targeted open reading frame

(Supplementary Table S8). gRNA plasmids were removed by growing the colonies in liquid YPD medium and subsequent plating on solid YPD medium. Plasmid removal was confirmed by growth on selective and non-selective solid media after which the strains were stored. For all transformations the corresponding gRNA plasmids and repair fragment are summarized in Supplementary Table S10.

To obtain a prototrophic strain with 35 deletions, a *URA3* transcriptional unit amplified from CEN.PK113-7D DNA (primers: 17752 and 17753, Supplementary Table S7) was integrated at the *GPP2* locus of strain IMX2520 (34 deletions). The flanks of the *URA3* repair fragment were homologous to the 60 bp immediately upstream and downstream of the *GPP2* ORF. The prototrophic 35 deletion strain IMX2538 was checked by diagnostic PCR, short-read sequencing and long-read nanopore sequencing. Integration of the *MPC3* transcriptional unit at the *X2* locus of IMX1984 and IMX2519 was achieved by amplifying the respective fragment from CEN.PK113-7D genomic DNA (primers:18025 and 18026 Supplementary Table S7) and integration by CRISPR/Cas9 using gRNA plasmid pUDR376, resulting in strains IMX2640 and IMX2641 respectively. Correct integration was verified by diagnostic PCR (Supplementary Table S8).

### **Sequencing**

High-quality genomic DNA of yeast for sequencing was extracted using the QIAGEN Blood & Cell Culture Kit with 100/G or 20/G Genomic-tips (QIAGEN) according to the manufacturer's instructions. DNA concentration was measured using the BR ds DNA kit (Invitrogen, Carlsbad, CA) and a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific). The purity was verified with a Nanodrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific).

#### Short read sequencing

IMX2538 (35 deletions prototrophic strain) was sequenced using 300 bp paired-end sequencing reads prepared with the MiSeq Reagent Kit v3 on an Illumina MiSeq sequencer (Illumina, San Diego, CA). To this end, extracted DNA was mechanically sheared to 550 bp with the M220 ultrasonicator (Covaris, Wolburn, MA) and subsequently, the TruSeq DNA PCR-Free Library Preparation kit (Illumina) was employed to make a six-strain library. The samples were quantified by qPCR on a Rotor-Gene Q PCR cycler (Qiagen) using the KAPA Library Quantification Kit (Kapa Biosystems, Wilmington, MA). Library integrity and fragment size were determined with a TapeStation 2200 (Agilent Technologies). Sequencing reads were mapped onto the CEN.PK113-7D (161) reference genome using the Burrows–Wheeler Alignment (BWA) tool (version 0.7.15) (162) and further processed using SAMtools (version 1.3.1) (163) and Pilon (with -vcf setting; version 1.18) (164) to identify single nucleotide polymorphisms (SNPs). The sequence was analyzed by visualizing the

generated .bam files in the Integrative Genomics Viewer (IGV) software (version 2.4.0) (165). Chromosomal copy number was estimated by the Magnolya algorithm (version 0.15) (166).

### Long read sequencing

High quality DNA of IMX2538 was isolated and checked on quantity and quality as described above. Furthermore, quality and integrity of DNA was checked with a TapeStation 2200 (Agilent Technologies, Santa Clara, CA). IMX2538 was sequenced in-house on a single R10 flow cell (FLO—MIN111) using the SQK-LSK109 sequencing kit (Oxford Nanopore Technologies, Oxford, United Kingdom), according to the manufacturer's instructions. With MinKnow (version 3.6.5, Oxford Nanopore Technologies) raw signal files were generated. Basecalling was performed by Guppy (version 4.0.11, Oxford Nanopore Technologies), followed by *de novo* assembly with Canu (version 2.0 (167)).

Short- and long-read sequencing data are available at NCBI under BioProject PRJNA757356.

### **Growth rate measurement in shake flasks**

The growth rate of the constructed strains was determined in 500 mL shake flasks containing 100 mL of SMD or SME medium. Wake-up cultures were inoculated with a 1 mL aliquot of a strain stored at -80 °C and grown until late-exponential phase. Pre-cultures were inoculated from the wake-up cultures and grown to mid-exponential phase. Finally, measuring cultures were inoculated in biological duplicate from the pre-culture at an initial OD<sub>660</sub> of 0.3. Cultures were monitored for OD<sub>660</sub> with a Jenway 7200 spectrophotometer in technical duplicate (Cole-Parmer, Vernon Hills, IL). A maximum specific growth rate ( $\mu_{\max}$ ) was calculated from at least five data points in the exponential phase with at least 2 doublings.

Anaerobic shake flask-based experiments were performed at 30 °C in a Bactron anaerobic chamber (Sheldon Manufacturing Inc., Cornelius, OR) with an atmosphere of 5 % (v/v) H<sub>2</sub>, 6 % (v/v) CO<sub>2</sub> and 89 % (v/v) N<sub>2</sub>, on a IKA KS 260 basic shaker at 200 rpm, using 50 mL shake flasks containing 30 mL SMD (2 %) anaerobic medium.

### **Growth rate measurement in microtiter plates**

Growth measurements of strains in microtiter plate with a Growth Profiler 960 (EnzyScreen BV, Heemstede, The Netherlands) were performed as described by Postma *et al.* (26).

## **Physiological characterization of CEN.PK113-7D and IMX2538 in bioreactor cultures**

Aerobic batch bioreactor cultures were performed in 2-L bioreactors (Applikon, Delft, The Netherlands). Bioreactors were filled with synthetic medium containing 5.0 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3.0 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, and 1.0 mL L<sup>-1</sup> trace elements. After heat sterilization, 20 g L<sup>-1</sup> glucose, 0.2 g L<sup>-1</sup> antifoam emulsion C (Sigma-Aldrich, St. Louis, MA), and filter-sterilized vitamins were added to complete the medium. Upon inoculation, bioreactors contained a working volume 1.4 L and the culture pH was maintained at 5.0 by automated addition of 2M H<sub>2</sub>SO<sub>4</sub> or 2M KOH. Temperature was kept stable at 30°C and mixing of the medium was performed at 800 rpm. The gas flow was set to 700 mL of air per minute to supply oxygen and remove produced carbon dioxide, and an overpressure of 0.3 bar was applied to the reactor. Dissolved oxygen tension was thus maintained for all reactors above 59 % over the whole duration of the batch cultivation. Off-gas was cooled to 2°C in a condenser on the bioreactor to prevent water evaporation, and further dried with a Permapure MD-110-48P-4 filter dryer (Permapure, Lakewood, NJ) for subsequent analysis of carbon dioxide and oxygen percentages by a MultiExact 4100 gas analyser (Servomex, Zoetermeer, The Netherlands). For both CEN.PK113-7D and IMX2538 reactors were run in biological triplicate and inoculated from exponentially growing shake flask cultures.

Optical densities were measured in technical triplicates on Jenway 7200 spectrophotometer (Cole-Parmer) at 660 nm, while cell dry weights were determined by filtration of 10 mL of well-mixed sample over dried PES membrane filters with a pore size of 0.45 µm (Pall Corporation, Port Washington, NY). Filters were washed with demineralized water and dried in a microwave oven for 20 minutes at 360 W.

Extracellular organic acids, sugars and ethanol were determined by high performance liquid chromatography (HPLC) analysis using an Aminex HPX-87H ion-exchange column (Agilent, Santa Clara) with 5 mM H<sub>2</sub>SO<sub>4</sub> as mobile phase and a flow rate of 0.6 mL min<sup>-1</sup> at 60°C. Glucose, glycerol, and ethanol were detected by a refractive-index detector (Agilent G1362A) and organic acids by a dual-wavelength absorbance detector (Agilent G1314F).

During mid-exponential growth in the glucose consumption phase, intracellular metabolite samples were taken with a filtration-based washing method according to Douma *et al.* (168) with some modifications. Briefly, approximately 3 mL of cell culture was sampled in 15 mL of 100 % methanol at -40°C. Biomass was washed with cooled 100 % methanol on a PES membrane with a pore size of 0.45 µm (Pall Corporation) which was pre-cooled and wetted with 100 % methanol at -40°C. Finally, metabolites were extracted with 75 % of boiling ethanol. 100 µL of <sup>13</sup>C cell extract was added to each tube as an internal standard for metabolite quantification (169). The intracellular

CCM metabolites, cofactors and nucleotides were derivatized and quantified as described by de Jonge *et al.* (170) and Niedenführ *et al.* (171).

## **Data availability**

All genomic data for this chapter have been deposited in the NCBI database (<https://www.ncbi.nlm.nih.gov/taxonomy>) under the BioProjectID PRJNA757356.

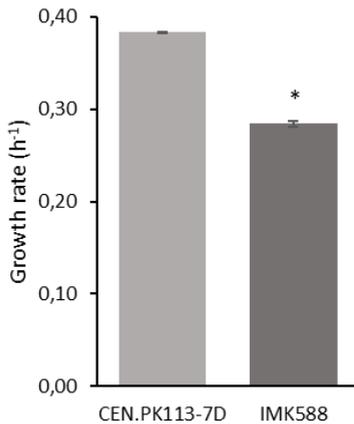
## **Acknowledgements**

We thank Pilar de la Torre and Marcel van den Broek for sequencing and bioinformatic analysis, Roel Sarelse, Ilse Pardijs and Lycka Kamoen for strain construction, Koen Verhagen, Patricia van Dam and Martin Pabst for analysis of intracellular metabolites, Erik de Hulster for assistance with sampling of the bioreactors, Marijke Luttik for the growth experiment on pyruvate medium, Sofia Dashko and Jean-Marc Daran for input during the literature analysis and Jack Pronk for feedback on an advanced version of the manuscript.

## Supplementary data

### Figure S1 - Growth rate of the *OAC1* deletion mutant

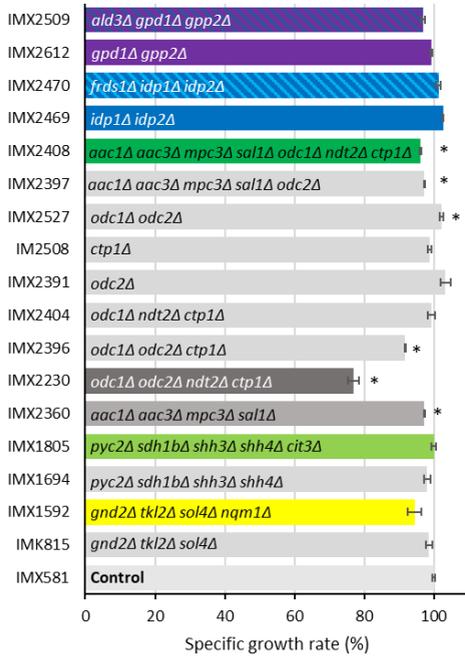
Maximum specific growth rate of *S. cerevisiae* CEN.PK113-7D (naïve reference strain) and the *OAC1* deletion mutant, *S. cerevisiae* IMK588, tested in shake flasks on selective SMD medium. Growth rates represent the average and standard deviation of two biological duplicates. \* IMK588 had a 26% slower growth rate with respect to CEN.PK113-7D (two-tailed paired homoscedastic t-test  $p < 0.05$ ).



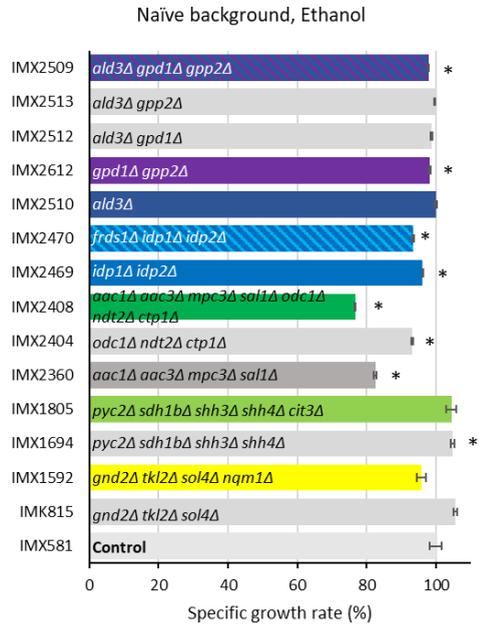
### Figure S2 - Growth rates of deletion strains

Maximum specific growth rates of all *S. cerevisiae* strains tested in shake flasks on selective SMD (A, B) or SME (B, D) supplemented with uracil. Represented are deletion strains in naïve background (A, B) and in the engineered background (C, D). Color coding is according to Figure 3 in the main manuscript. Growth rates represent the average and standard deviation of two biological duplicates. Growth rates of the deletion strains are expressed as % of the control strain IMX581. Significant changes in growth rate with respect to the control strain are indicated with a \* (two-tailed paired homoscedastic t-test  $p < 0.05$ ).

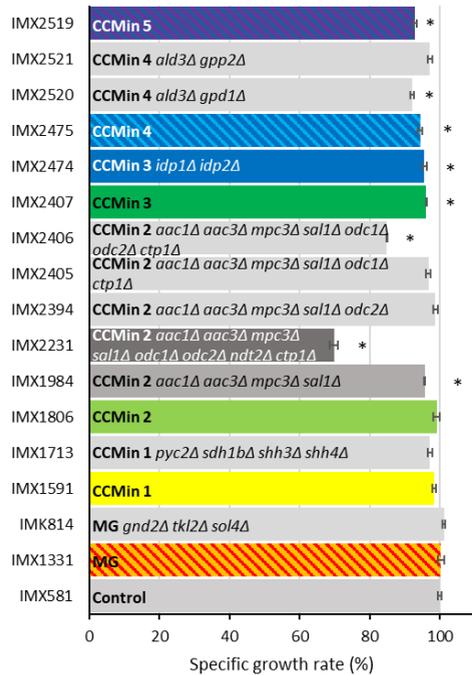
A) Naïve background, Glucose



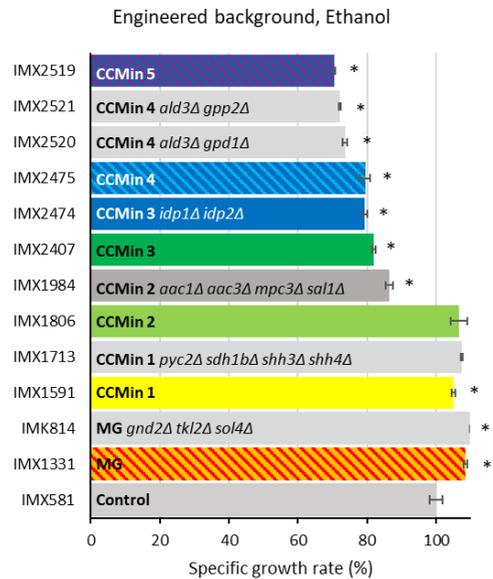
B) Naïve background, Ethanol



C) Engineered background, Glucose



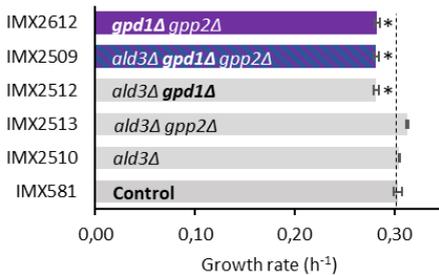
D) Engineered background, Ethanol



### Figure S3 - Growth rates in specific environments

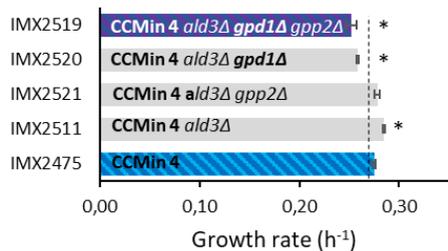
A, B, C and D) Specific growth rate determination on osmotic stress conditions (supplemented with uracil) measured with the growth profiler in biological triplicate. E) Specific growth rate of *FRDS1* and *AAC3* deletion mutants on SMD supplemented with Tween, ergosterol and uracil measured under anaerobic conditions in shake flasks in biological duplicate. F) Specific growth rate of *MPC3* complementation strains on SME supplemented with uracil in shake flask in biological duplicate. G) Specific growth rate of *MPC3* complementation strain on SM with 83.3 mM pyruvate supplemented with uracil in shake flask in biological duplicate. Significant changes in growth rate with respect to the parental strain are indicated with a \* (two-tailed paired homoscedastic t-test  $p < 0.05$ ).

A) 20% glucose, naïve background

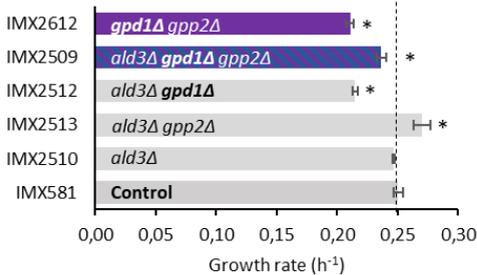


B)

20% glucose, engineered background

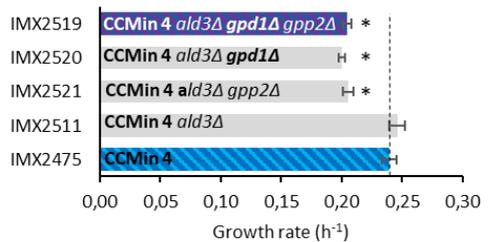


C) 2M sorbitol, naïve background



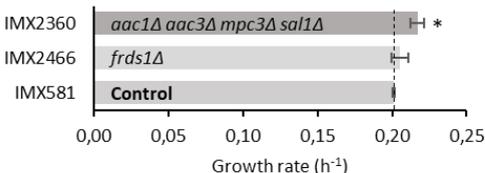
D)

2M sorbitol, engineered background



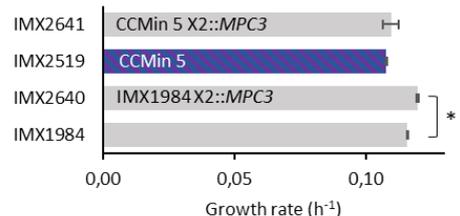
E)

SMD anaerobic, naïve background



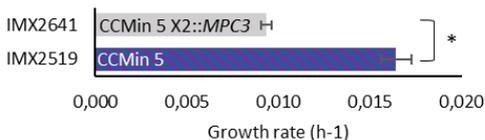
F)

SME, engineered background



G)

SMPyruvate, engineered background



**Table S1 Genetic characteristics of the paralogs considered for deletion**

Genes that were deleted in this study are marked in red. Protein similarity was based on global alignment using BLOSUM62 of the translated genes of the *S. cerevisiae* CEN.PK113-7D genome sequence (161). WGD pairs are indicated (172).

Pathway	Gene	% Protein Similarity	Type of duplication
Pentose phosphate pathway	<i>GND1</i> (YHR183W) / <i>GND2</i> (YGR256W)	87%	WGD
	<i>TKL1</i> (YPR074C) / <i>TKL2</i> (YBR117C)	71%	WGD
	<i>SOL3</i> (YHR163W) / <i>SOL4</i> (YGR248W)	47%	WGD
	<i>TAL1</i> (YLR354C) / <i>NQM1</i> (YGR043C)	68%	WGD
TCA cycle + anaplerotic reactions	<i>PYC1</i> (YGL062W) / <i>PYC2</i> (YBR218C)	92%	WGD
	<i>SDH1</i> (YKL148C) / <i>SDH1b</i> (YJL045W)	84%	WGD
	<i>SDH3</i> (YKL141W) / <i>SHH3</i> (YMR118C)	42%	WGD
	<i>SDH4</i> (YDR178W) / <i>SHH4</i> (YLR164W)	47%	WGD
	<i>CIT1</i> (YNR001C) / <i>CIT2</i> (YCR005C) / <i>CIT3</i> (YPR001W)	CIT1/CIT2=75% CIT1/CIT3=42%	CIT1/CIT2=WGD
Mitochondrial transporters	<i>AAC1</i> (YMR056C) / <i>AAC2</i> (YBL030C) / <i>AAC3</i> (YBR085W)	AAC2/AAC1=73% AAC2/AAC3=87%	AAC2/AAC3=WGD
	<i>SAL1</i> (YNL083W)		
	<i>MPC1</i> (YGL080W) / <i>MPC2</i> (YHR162W) / <i>MPC3</i> (YGR243W)	MPC2/MPC3=65% MPC2/MPC1=25%	MPC2/MPC3=WGD
	<i>CTP1</i> (YBR291C)		
	<i>ODC1</i> (YPL134C) / <i>ODC2</i> (YOR222W)	61%	WGD
	<i>NDT1</i> (YIL006W) / <i>NDT2</i> (YEL006W)	51%	WGD
Fumarate reductase	<i>OSM1</i> (YJR051W) / <i>FRDS1</i> (YEL047C)	59%	WGD
Glyoxylate and TCA cycle	<i>IDP1</i> (YDL066W) / <i>IDP2</i> (YLR174W) / <i>IDP3</i> (YNL009W)	IDP1/IDP3=65% IDP2/IDP3 =73%	IDP2/IDP3=WGD
acetyl-CoA synthesis	<i>ALD2</i> (YMR170C) / <i>ALD3</i> (YMR169C)	91%	

Glycerol synthesis	<i>GPD1</i> (YDL022W) / <i>GPD2</i> (YOL059W)	72%	WGD
	<i>GPP1</i> (YIL053W) / <i>GPP2</i> (YER062C)	92%	WGD

### Table S2 Mitochondrial carrier proteins

A list of 35 mitochondrial transporter identified by Palmieri *et al.* (62) and reviewed by Palmieri and Monné (63). In this list the systematic name, standard name, importance in this study and the function of the gene encoding for mitochondrial transporters are listed. NA, not assigned. \**CRC1* was not considered as involved in the CCM since there is no carnitine supplied to the cultures in this study and thus the carnitine shuttle is not active.

Systematic name	Standard name	Considered as involved in CCM?	Function
<b>YIL006w</b>	<i>NDT1</i>	YES	Mitochondrial NAD <sup>+</sup> transporter
<b>YEL006w</b>	<i>NDT2</i>	YES	Mitochondrial NAD <sup>+</sup> transporter
<b>YIL134w</b>	<i>FLX1</i>	YES	Mitochondrial FAD transporter
<b>YBR192w</b>	<i>RIM2</i>	NO	Mitochondrial pyrimidine nucleotide transporter
<b>YDL119c</b>	<i>HEM25</i>	NO	Mitochondrial glycine transporter
<b>YNL003c</b>	<i>SAM5</i>	NO	S-adenosylmethionine transporter of the mitochondrial inner membrane
<b>YKR052c</b>	<i>MRS4</i>	NO	Iron transporter of the mitochondrial carrier family (may transport other cations)
<b>YJL133w</b>	<i>MRS3</i>	NO	Iron transporter of the mitochondrial carrier family (May transport other cations)
<b>YGR257c</b>	<i>MTM1</i>	NO	pyridoxal 5'-phosphate (PLP) transporter
<b>YER053c</b>	<i>PIC2</i>	NO	Mitochondrial copper and phosphate carrier
<b>YJR077c</b>	<i>MIR1</i>	NO	Mitochondrial phosphate carrier
<b>YOR130c</b>	<i>ORT1</i>	NO	Ornithine transporter of the mitochondrial inner membrane (ornithine-proton exchange or ornithine-ornithine exchange also transports arginine and lysine)
<b>YOR100c</b>	<i>CRC1</i>	NO*	Mitochondrial inner membrane carnitine transporter. Transports carnitine, acetylcarnitin and propionylcarnitine .
<b>YBR104w</b>	<i>YMC2</i>	NO	Putative mitochondrial inner membrane transporter. Proposed role in oleate metabolism and glutamate biosynthesis.
<b>YPR058w</b>	<i>YMC1</i>	NO	Secondary mitochondrial inner membrane glycine transporter; required with HEM25 for the transport of glycine into mitochondria. Proposed

			role in oleate metabolism and glutamate biosynthesis.
<b>YPL134c</b>	<i>ODC1</i>	YES	2-oxodicarboxylate transporter (transports 2-oxoglutarate and oxodipate+ corresponding dicarboxylates and malate by counterexchange)
<b>YOR222w</b>	<i>ODC2</i>	YES	2-oxodicarboxylate transporter (transports oxoglutarate and oxodipate+ corresponding dicarboxylates and malate by counterexchange)
<b>YPR021c</b>	<i>AGC1</i>	NO	Mitochondrial amino acid transporter (transport aspartate and glutamate in uniport as well as in exchange mechanism)
<b>YJR095w</b>	<i>SFC1</i>	YES	Mitochondrial succinate-fumarate counter exchange transporter (fumerate to cytosol, succinate to mitochondria)
<b>YBR291c</b>	<i>CTP1</i>	YES	Mitochondrial inner membrane citrate transporter
<b>YFR045w</b>	NA	NO	Putative mitochondrial transport protein; null mutant is viable
<b>YMR241w</b>	<i>YHM2</i>	YES	Citrate and oxoglutarate carrier protein (Citrate exported and oxoglutarate imported) (oxaloacetate, succinate and fumerate to a lesser extend)
<b>YLR348c</b>	<i>DIC1</i>	YES	Mitochondrial dicarboxylate carrier (transports malate, succinate, malonate, inorganic phosphate by counter exchange mechanism. Also sulphate, thiosulphate)
<b>YKL120w</b>	<i>OAC1</i>	YES	Transports oxaloacetate and sulfate (unidirectional+counterexchange)
<b>YDL198c</b>	<i>GGC1</i>	NO	Mitochondrial GTP/GDP exchange transporter, essential for mitochondrial genome maintenance, has a role in mitochondrial iron transport
<b>YGR096w</b>	<i>TPC1</i>	YES	Mediates uptake of the essential cofactor thiamine pyrophosphate (ThPP) into mitochondria
<b>YMR056c</b>	<i>AAC1</i>	YES	Mitochondrial inner membrane ADP/ATP translocator
<b>YBL030c</b>	<i>AAC2</i>	YES	Mitochondrial inner membrane ADP/ATP translocator
<b>YBR085w</b>	<i>AAC3</i>	YES	Mitochondrial inner membrane ADP/ATP translocator
<b>YHR002w</b>	<i>LEU5</i>	YES	Involved in the accumulation of CoA in the mitochondrial matrix
<b>YPR011c</b>	NA	NO	Putative 5'-phosphosulfate (APS) and 3'-phospho-adenosine 5'-phosphosulfate (PAPS) transporter.

			Not enough information on the function to include in this study.
<b>YNL083w</b>	<i>SAL1</i>	YES	ADP/ATP transporter (activity of either Sal1p or Pet9p is critical for viability)
<b>YGL080w</b>	<i>MPC1</i>	YES	Conserved subunit of mitochondrial pyruvate carrier (MPC)
<b>YHR162w</b>	<i>MPC2</i>	YES	Highly conserved subunit of the mitochondrial pyruvate carrier (MPC)
<b>YGR243w</b>	<i>MPC3</i>	YES	Highly conserved subunit of the mitochondrial pyruvate carrier (MPC)
<b>YMR166C</b>	<i>MME1</i>	NO	Mitochondrial inner membrane transporter that exports magnesium

#### **Table S3 Strains**

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16538955>

#### **Table S4 Plasmids**

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16538955>

#### **Table S5 pROS/pMEL gRNA primers. gRNA sequence is underlined.**

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16538955>

#### **Table S6 List of primers used to confirm guideRNA plasmids**

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16538955>

#### **Table S7 List of primers used for making repair fragments for gene deletion or integration**

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16538955>

#### **Table S8 List of primers used the verify removal or insertion of a gene.**

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16538955>

#### **Table S9 List of primers to construct and verify strain IMX1331**

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16538955>

#### **Table S10 Strain transformations**

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16538955>

## References

1. Lara, A.R. and Gosset, G. (2019) *Minimal cells: design, construction, biotechnological applications*. Springer International Publishing.
2. Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., Ote, T., Yamakawa, T., Yamazaki, Y., Mori, H. *et al.* (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.*, **55**, 137-149.
3. Morimoto, T., Kadoya, R., Endo, K., Tohata, M., Sawada, K., Liu, S., Ozawa, T., Kodama, T., Kakeshita, H., Kageyama, Y. *et al.* (2008) Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res.*, **15**, 73-81.
4. Suzuki, N., Nonaka, H., Tsuge, Y., Inui, M. and Yukawa, H. (2005) New multiple-deletion method for the *Corynebacterium glutamicum* genome, using a mutant *lox* sequence. *Appl. Environ. Microbiol.*, **71**, 8472-8480.
5. Lieder, S., Nikel, P.I., de Lorenzo, V. and Takors, R. (2015) Genome reduction boosts heterologous gene expression in *Pseudomonas putida*. *Microb. Cell Fact.*, **14**, 23.
6. Zhu, D., Fu, Y., Liu, F., Xu, H., Saris, P.E.J. and Qiao, M. (2017) Enhanced heterologous protein productivity by genome reduction in *Lactococcus lactis* NZ9000. *Microb. Cell Fact.*, **16**, 1.
7. Komatsu, M., Uchiyama, T., Omura, S., Cane, D.E. and Ikeda, H. (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 2646-2651.
8. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L. *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253.
9. Mutschler, H., Robinson, T., Tang, T.D. and Wegner, S. (2019) Special issue on bottom-up synthetic biology. *Chembiochem*, **20**, 2533-2534.
10. Conant, G.C. and Wolfe, K.H. (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.*, **9**, 938-950.
11. Escalera-Fanjul, X., Quezada, H., Riego-Ruiz, L. and Gonzalez, A. (2019) Whole-genome duplication and yeast's fruitful way of life. *Trends Genet.*, **35**, 42-54.
12. Acharya, D. and Ghosh, T.C. (2016) Global analysis of human duplicated genes reveals the relative importance of whole-genome duplicates originated in the early vertebrate evolution. *BMC genomics*, **17**, 1-14.
13. Fares, M.A., Keane, O.M., Toft, C., Carretero-Paulet, L. and Jones, G.W. (2013) The roles of whole-genome and small-scale duplications in the functional specialization of *Saccharomyces cerevisiae* genes. *PLoS Genet.*, **9**, e1003176.
14. Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H. *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**, 901-906.
15. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387-391.
16. Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D. *et al.* (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science*, **353**, aaf1420.
17. Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Mattiazzi Usaj, M., van Leeuwen, J. *et al.* (2018) Systematic analysis of complex genetic interactions. *Science*, **360**, eaao1729.

18. Tong, A.H.Y., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghobizadeh, S., Hogue, C.W.V., Bussey, H. *et al.* (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**, 2364-2368.
19. Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M. *et al.* (2004) Global mapping of the yeast genetic interaction network. *Science*, **303**, 808-813.
20. Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S. *et al.* (2010) The genetic landscape of a cell. *Science*, **327**, 425-431.
21. Luo, Z., Yu, K., Xie, S., Monti, M., Schindler, D., Fang, Y., Zhao, S., Liang, Z., Jiang, S., Luan, M. *et al.* (2021) Compacting a synthetic yeast chromosome arm. *Genome Biol.*, **22**, 5.
22. Solis-Escalante, D., Kuijpers, N.G., Barrajon-Simancas, N., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2015) A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell*, **14**, 804-816.
23. Bradley, P.H., Gibney, P.A., Botstein, D., Troyanskaya, O.G. and Rabinowitz, J.D. (2019) Minor isozymes tailor yeast metabolism to carbon availability. *mSystems*, **4**, e00170-00118.
24. Kuijpers, N.G., Solis-Escalante, D., Luttkik, M.A., Bisschops, M.M., Boonekamp, F.J., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2016) Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 15060-15065.
25. Nielsen, J. and Keasling, Jay D. (2016) Engineering cellular metabolism. *Cell*, **164**, 1185-1197.
26. Postma, E.D., Dashko, S., van Breemen, L., Taylor Parkins, S.K., van den Broek, M., Daran, J.-M. and Daran-Lapujade, P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **49**, 1769-1783.
27. DeLoache, W.C., Russ, Z.N. and Dueber, J.E. (2016) Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways. *Nat. Commun.*, **7**, 11152.
28. Knijnenburg, T.A., Daran, J.-M.G., van den Broek, M.A., Daran-Lapujade, P.A., de Winde, J.H., Pronk, J.T., Reinders, M.J. and Wessels, L.F. (2009) Combinatorial effects of environmental parameters on transcriptional regulation in *Saccharomyces cerevisiae*: a quantitative analysis of a compendium of chemostat-based transcriptome data. *BMC genomics*, **10**, 53.
29. Gancedo, C. and Flores, C.L. (2008) Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.*, **72**, 197-210.
30. Jansen, M.L.A., Bracher, J.M., Papapetridis, I., Verhoeven, M.D., de Bruijn, H., de Waal, P.P., van Maris, A.J.A., Klaassen, P. and Pronk, J.T. (2017) *Saccharomyces cerevisiae* strains for second-generation ethanol production: from academic exploration to industrial implementation. *FEMS Yeast Res.*, **17**, fox044.
31. Lobo, Z. and Maitra, P.K. (1982) Pentose phosphate pathway mutants of yeast. *Mol. Gen. Genet.*, **185**, 367-368.
32. Schaaff-Gerstenschläger, I., Mannhaupt, G., Vetter, I., Zimmermann, F.K. and Feldmann, H. (1993) *TKL2*, a second transketolase gene of *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, **217**, 487-492.
33. Matsushika, A., Goshima, T., Fujii, T., Inoue, H., Sawayama, S. and Yano, S. (2012) Characterization of non-oxidative transaldolase and transketolase enzymes in the pentose phosphate pathway with regard to xylose utilization by recombinant *Saccharomyces cerevisiae*. *Enzyme Microb. Technol.*, **51**, 16-25.

34. Stanford, D., Whitney, M., Hurto, R., Eisaman, D., Shen, W.-C. and Hopper, A. (2004) Division of labor among the yeast Sol proteins implicated in tRNA nuclear export and carbohydrate metabolism. *Genetics*, **168**, 117-127.
35. Castelli, L.M., Lui, J., Campbell, S.G., Rowe, W., Zeef, L.A., Holmes, L.E., Hoyle, N.P., Bone, J., Selley, J.N., Sims, P.F. *et al.* (2011) Glucose depletion inhibits translation initiation via eIF4A loss and subsequent 48S preinitiation complex accumulation, while the pentose phosphate pathway is coordinately up-regulated. *Mol. Biol. Cell*, **22**, 3379-3393.
36. Barnett, J.A. (2003) A history of research on yeasts 6: the main respiratory pathway. *Yeast*, **20**, 1015-1044.
37. Cupp, J. and McAlister-Henn, L. (1991) NAD (+)-dependent isocitrate dehydrogenase. Cloning, nucleotide sequence, and disruption of the *IDH2* gene from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **266**, 22199-22205.
38. Cupp, J.R. and McAlister-Henn, L. (1992) Cloning and characterization of the gene encoding the IDH1 subunit of NAD (+)-dependent isocitrate dehydrogenase from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **267**, 16417-16423.
39. Dickinson, J.R., Roy, D.J. and Dawes, I.W. (1986) A mutation affecting lipoamide dehydrogenase, pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase activities in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **204**, 103-107.
40. Repetto, B. and Tzagoloff, A. (1989) Structure and regulation of *KGD1*, the structural gene for yeast alpha-ketoglutarate dehydrogenase. *Mol. Cell. Biol.*, **9**, 2695-2705.
41. Repetto, B. and Tzagoloff, A. (1990) Structure and regulation of *KGD2*, the structural gene for yeast dihydrolipoyl transsuccinylase. *Mol. Cell. Biol.*, **10**, 4221-4232.
42. Przybyla-Zawislak, B., Dennis, R.A., Zakharkin, S.O. and McCammon, M.T. (1998) Genes of succinyl-CoA ligase from *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, **258**, 736-743.
43. Bullis, B.L. and Lemire, B.D. (1994) Isolation and characterization of the *Saccharomyces cerevisiae SDH4* gene encoding a membrane anchor subunit of succinate dehydrogenase. *J. Biol. Chem.*, **269**, 6543-6549.
44. Chapman, K.B., Solomon, S.D. and Boeke, J.D. (1992) *SDH1*, the gene encoding the succinate dehydrogenase flavoprotein subunit from *Saccharomyces cerevisiae*. *Gene*, **118**, 131-136.
45. Colby, G., Ishii, Y. and Tzagoloff, A. (1998) Suppression of *sdh1* mutations by the *SDH1b* gene of *Saccharomyces cerevisiae*. *Yeast*, **14**, 1001-1006.
46. Gebert, N., Gebert, M., Oeljeklaus, S., von der Malsburg, K., Stroud, D.A., Kulawiak, B., Wirth, C., Zahedi, R.P., Dolezal, P., Wiese, S. *et al.* (2011) Dual function of Sdh3 in the respiratory chain and TIM22 protein translocase of the mitochondrial inner membrane. *Mol. Cell*, **44**, 811-818.
47. Lombardo, A., Carine, K. and Scheffler, I.E. (1990) Cloning and characterization of the iron-sulfur subunit gene of succinate dehydrogenase from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **265**, 10419-10423.
48. Szeto, S.S., Reinke, S.N., Oyedotun, K.S., Sykes, B.D. and Lemire, B.D. (2012) Expression of *Saccharomyces cerevisiae* Sdh3p and Sdh4p paralogs results in catalytically active succinate dehydrogenase isoenzymes. *J. Biol. Chem.*, **287**, 22509-22520.
49. Chang, Y.L., Hsieh, M.H., Chang, W.W., Wang, H.Y., Lin, M.C., Wang, C.P., Lou, P.J. and Teng, S.C. (2015) Instability of succinate dehydrogenase in SDHD polymorphism connects reactive oxygen species production to nuclear and mitochondrial genomic mutations in yeast. *Antioxid. Redox Signal.*, **22**, 587-602.
50. Jia, Y.K., Bécam, A.M. and Herbert, C. (1997) The *CIT3* gene of *Saccharomyces cerevisiae* encodes a second mitochondrial isoform of citrate synthase. *Mol. Microbiol.*, **24**, 53-59.
51. Haselbeck, R.J., Colman, R.F. and McAlister-Henn, L. (1992) Isolation and sequence of a cDNA encoding porcine mitochondrial NADP-specific isocitrate dehydrogenase. *Biochemistry*, **31**, 6219-6223.
52. Zhao, W.N. and McAlister-Henn, L. (1996) Expression and gene disruption analysis of the isocitrate dehydrogenase family in yeast. *Biochemistry*, **35**, 7873-7878.

53. Gangloff, S.P., Marguet, D. and Lauquin, G. (1990) Molecular cloning of the yeast mitochondrial aconitase gene (*ACO1*) and evidence of a synergistic regulation of expression by glucose plus glutamate. *Mol. Cell. Biol.*, **10**, 3551-3561.
54. Fazius, F., Shelest, E., Gebhardt, P. and Brock, M. (2012) The fungal  $\alpha$ -amino adipate pathway for lysine biosynthesis requires two enzymes of the aconitase family for the isomerization of homocitrate to homoisocitrate. *Mol. Microbiol.*, **86**, 1508-1530.
55. Pronk, J.T., Yde Steensma, H. and Van Dijken, J.P. (1996) Pyruvate metabolism in *Saccharomyces cerevisiae*. *Yeast*, **12**, 1607-1633.
56. Stucka, R., Dequin, S., Salmon, J.-M. and Gancedo, C. (1991) DNA sequences in chromosomes 11 and VII code for pyruvate carboxylase isoenzymes in *Saccharomyces cerevisiae*: analysis of pyruvate carboxylase-deficient strains. *Mol. Gen. Genet.*, **229**, 307-315.
57. Brewster, N.K., Val, D.L., Walker, M.E. and Wallace, J.C. (1994) Regulation of pyruvate carboxylase isozyme (*PYC1*, *PYC2*) gene expression in *Saccharomyces cerevisiae* during fermentative and nonfermentative growth. *Arch. Biochem. Biophys.*, **311**, 62-71.
58. Boles, E., de Jong-Gubbels, P. and Pronk, J.T. (1998) Identification and characterization of *MAE1*, the *Saccharomyces cerevisiae* structural gene encoding mitochondrial malic enzyme. *J. Bacteriol.*, **180**, 2875-2882.
59. Valdés-Hevia, M.D., de la Guerra, R. and Gancedo, C. (1989) Isolation and characterization of the gene encoding phosphoenolpyruvate carboxykinase from *Saccharomyces cerevisiae*. *FEBS Lett.*, **258**, 313-316.
60. Soontorngun, N., Larochelle, M., Drouin, S., Robert, F. and Turcotte, B. (2007) Regulation of gluconeogenesis in *Saccharomyces cerevisiae* is mediated by activator and repressor functions of Rds2. *Mol. Cell. Biol.*, **27**, 7895-7905.
61. Palmieri, F., Agrimi, G., Blanco, E., Castegna, A., Di Noia, M.A., Iacobazzi, V., Lasorsa, F.M., Marobbio, C.M., Palmieri, L., Scarcia, P. *et al.* (2006) Identification of mitochondrial carriers in *Saccharomyces cerevisiae* by transport assay of reconstituted recombinant proteins. *Biochim. Biophys. Acta.*, **1757**, 1249-1262.
62. Palmieri, L., Palmieri, F., Runswick, M.J. and Walker, J.E. (1996) Identification by bacterial expression and functional reconstitution of the yeast genomic sequence encoding the mitochondrial dicarboxylate carrier protein. *FEBS Lett.*, **399**, 299-302.
63. Palmieri, F. and Monné, M. (2016) Discoveries, metabolic roles and diseases of mitochondrial carriers: A review. *Biochim. Biophys. Acta.*, **1863**, 2362-2378.
64. Cavero, S., Traba, J., Del Arco, A. and Satrústegui, J. (2005) The calcium-dependent ATP-Mg/Pi mitochondrial carrier is a target of glucose-induced calcium signalling in *Saccharomyces cerevisiae*. *Biochem J.*, **392**, 537-544.
65. Smith, C.P. and Thorsness, P.E. (2008) The molecular basis for relative physiological functionality of the ADP/ATP carrier isoforms in *Saccharomyces cerevisiae*. *Genetics*, **179**, 1285-1299.
66. Adrian, G.S., McCammon, M.T., Montgomery, D.L. and Douglas, M.G. (1986) Sequences required for delivery and localization of the ADP/ATP translocator to the mitochondrial inner membrane. *Mol. Cell. Biol.*, **6**, 626-634.
67. Gawaz, M., Douglas, M.G. and Klingenberg, M. (1990) Structure-function studies of adenine nucleotide transport in mitochondria. II. Biochemical analysis of distinct AAC1 and AAC2 proteins in yeast. *J. Biol. Chem.*, **265**, 14202-14208.
68. Chen, X.J. (2004) Sal1p, a calcium-dependent carrier protein that suppresses an essential cellular function associated with the Aac2 isoform of ADP/ATP translocase in *Saccharomyces cerevisiae*. *Genetics*, **167**, 607-617.
69. Laco, J., Zeman, I., Pevala, V., Polcic, P. and Kolarov, J. (2010) Adenine nucleotide transport via Sal1 carrier compensates for the essential function of the mitochondrial ADP/ATP carrier. *FEMS Yeast Res.*, **10**, 290-296.

70. Todisco, S., Agrimi, G., Castegna, A. and Palmieri, F. (2006) Identification of the mitochondrial NAD<sup>+</sup> transporter in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **281**, 1524-1531.
71. Marobbio, C.M., Vozza, A., Harding, M., Bisaccia, F., Palmieri, F. and Walker, J.E. (2002) Identification and reconstitution of the yeast mitochondrial transporter for thiamine pyrophosphate. *EMBO J.*, **21**, 5653-5661.
72. Bafunno, V., Giancaspero, T.A., Brizio, C., Bufano, D., Passarella, S., Boles, E. and Barile, M. (2004) Riboflavin uptake and FAD synthesis in *Saccharomyces cerevisiae* mitochondria: involvement of the Flx1p carrier in FAD export. *J. Biol. Chem.*, **279**, 95-102.
73. Tzagoloff, A., Jang, J., Glerum, D.M. and Wu, M. (1996) *FLX1* codes for a carrier protein involved in maintaining a proper balance of flavin nucleotides in yeast mitochondria. *J. Biol. Chem.*, **271**, 7392-7397.
74. Prohl, C., Pelzer, W., Diekert, K., Kmita, H., Bedekovics, T., Kispal, G. and Lill, R. (2001) The yeast mitochondrial carrier Leu5p and its human homologue Graves' disease protein are required for accumulation of coenzyme A in the matrix. *Mol. Cell. Biol.*, **21**, 1089-1097.
75. Bender, T., Pena, G. and Martinou, J.C. (2015) Regulation of mitochondrial pyruvate uptake by alternative pyruvate carrier complexes. *EMBO J.*, **34**, 911-924.
76. Herzig, S., Raemy, E., Montessuit, S., Veuthey, J.L., Zamboni, N., Westermann, B., Kunji, E.R. and Martinou, J.C. (2012) Identification and functional expression of the mitochondrial pyruvate carrier. *Science*, **337**, 93-96.
77. Timón-Gómez, A., Proft, M. and Pascual-Ahuir, A. (2013) Differential regulation of mitochondrial pyruvate carrier genes modulates respiratory capacity and stress tolerance in yeast. *PLoS one*, **8**, e79405.
78. Palmieri, L., Lasorsa, F.M., De Palma, A., Palmieri, F., Runswick, M.J. and Walker, J.E. (1997) Identification of the yeast *ACR1* gene product as a succinate-fumarate transporter essential for growth on ethanol or acetate. *FEBS Lett.*, **417**, 114-118.
79. Palmieri, L., Vozza, A., Hönlinger, A., Dietmeier, K., Palmisano, A., Zara, V. and Palmieri, F. (1999) The mitochondrial dicarboxylate carrier is essential for the growth of *Saccharomyces cerevisiae* on ethanol or acetate as the sole carbon source. *Mol. Microbiol.*, **31**, 569-577.
80. Fernández, M., Fernández, E. and Rodicio, R. (1994) *ACR1*, a gene encoding a protein related to mitochondrial carriers, is essential for acetyl-CoA synthetase activity in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **242**, 727-735.
81. Palmieri, L., Vozza, A., Agrimi, G., De Marco, V., Runswick, M.J., Palmieri, F. and Walker, J.E. (1999) Identification of the yeast mitochondrial transporter for oxaloacetate and sulfate. *J. Biol. Chem.*, **274**, 22184-22190.
82. Marobbio, C.M.T., Giannuzzi, G., Paradies, E., Pierri, C.L. and Palmieri, F. (2008) alpha-Isopropylmalate, a leucine biosynthesis intermediate in yeast, is transported by the mitochondrial oxalacetate carrier. *J. Biol. Chem.*, **283**, 28445-28453.
83. Castegna, A., Scarcia, P., Agrimi, G., Palmieri, L., Rottensteiner, H., Spera, I., Germinario, L. and Palmieri, F. (2010) Identification and functional characterization of a novel mitochondrial carrier for citrate and oxoglutarate in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **285**, 17359-17370.
84. Tibbetts, A.S., Sun, Y., Lyon, N.A., Ghrist, A.C. and Trotter, P.J. (2002) Yeast mitochondrial oxodicarboxylate transporters are important for growth on oleic acid. *Arch. Biochem. Biophys.*, **406**, 96-104.
85. Scarcia, P., Palmieri, L., Agrimi, G., Palmieri, F. and Rottensteiner, H. (2017) Three mitochondrial transporters of *Saccharomyces cerevisiae* are essential for ammonium fixation and lysine biosynthesis in synthetic minimal medium. *Mol. Genet. Metab.*, **122**, 54-60.

86. Kaplan, R.S., Mayor, J.A., Gremse, D.A. and Wood, D.O. (1995) High level expression and characterization of the mitochondrial citrate transport protein from the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **270**, 4108-4114.
87. Kaplan, R.S., Mayor, J.A., Kakhniashvili, D., Gremse, D.A., Wood, D.O. and Nelson, D.R. (1996) Deletion of the nuclear gene encoding the mitochondrial citrate transport protein from *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, **226**, 657-662.
88. Cho, J.H., Ha, S.J., Kao, L.R., Megraw, T.L. and Chae, C.B. (1998) A novel DNA-binding protein bound to the mitochondrial inner membrane restores the null mutation of mitochondrial histone Abf2p in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **18**, 5712-5723.
89. Palmieri, L., Agrimi, G., Runswick, M.J., Fearnley, I.M., Palmieri, F. and Walker, J.E. (2001) Identification in *Saccharomyces cerevisiae* of two isoforms of a novel mitochondrial transporter for 2-oxoadipate and 2-oxoglutarate. *J. Biol. Chem.*, **276**, 1916-1922.
90. Camarasa, C., Faucet, V. and Dequin, S. (2007) Role in anaerobiosis of the isoenzymes for *Saccharomyces cerevisiae* fumarate reductase encoded by *OSM1* and *FRDS1*. *Yeast*, **24**, 391-401.
91. Jouhten, P. and Penttilä, M. (2014) Anaerobic Carbon Metabolism of *Saccharomyces cerevisiae*. In: Piškur, J. and Compagno, C. (eds.), *Molecular Mechanisms in Yeast Carbon Metabolism*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 57-82.
92. Liu, Z., Österlund, T., Hou, J., Petranovic, D. and Nielsen, J. (2013) Anaerobic  $\alpha$ -amylase production and secretion with fumarate as the final electron acceptor in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **79**, 2962-2967.
93. Neal, S.E., Dabir, D.V., Wijaya, J., Boon, C. and Koehler, C.M. (2017) Osm1 facilitates the transfer of electrons from Erv1 to fumarate in the redox-regulated import pathway in the mitochondrial intermembrane space. *Mol. Cell. Biol.*, **28**, 2773-2785.
94. Arikawa, Y., Enomoto, K., Muratsubaki, H. and Okazaki, M. (1998) Soluble fumarate reductase isoenzymes from *Saccharomyces cerevisiae* are required for anaerobic growth. *FEMS Microbiol. Lett.*, **165**, 111-116.
95. Kunze, M., Pracharoenwattana, I., Smith, S.M. and Hartig, A. (2006) A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochim. Biophys. Acta.*, **1763**, 1441-1452.
96. Xiberras, J., Klein, M. and Nevoigt, E. (2019) Glycerol as a substrate for *Saccharomyces cerevisiae* based bioprocesses - Knowledge gaps regarding the central carbon catabolism of this 'non-fermentable' carbon source. *Biotechnol. Adv.*, **37**, 107378.
97. Boubekour, S., Camougrand, N., Bunoust, O., Rigoulet, M. and Guérin, B. (2001) Participation of acetaldehyde dehydrogenases in ethanol and pyruvate metabolism of the yeast *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, **268**, 5057-5065.
98. Kurita, O. and Nishida, Y. (1999) Involvement of mitochondrial aldehyde dehydrogenase *ALD5* in maintenance of the mitochondrial electron transport chain in *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett.*, **181**, 281-287.
99. Meaden, P.G., Dickinson, F.M., Mifsud, A., Tessier, W., Westwater, J., Bussey, H. and Midgley, M. (1997) The *ALD6* gene of *Saccharomyces cerevisiae* encodes a cytosolic, Mg(2+)-activated acetaldehyde dehydrogenase. *Yeast*, **13**, 1319-1327.
100. White, W.H., Skatrud, P.L., Xue, Z. and Toyn, J.H. (2003) Specialization of function among aldehyde dehydrogenases: the *ALD2* and *ALD3* genes are required for beta-alanine biosynthesis in *Saccharomyces cerevisiae*. *Genetics*, **163**, 69-77.
101. Navarro-Aviño, J.P., Prasad, R., Miralles, V.J., Benito, R.M. and Serrano, R. (1999) A proposal for nomenclature of aldehyde dehydrogenases in *Saccharomyces cerevisiae* and characterization of the stress-inducible *ALD2* and *ALD3* genes. *Yeast*, **15**, 829-842.
102. Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y. and Nielsen, J. (2015) Microbial acetyl-CoA metabolism and metabolic engineering. *Metab. Eng.*, **28**, 28-42.

103. Chen, Y., Siewers, V. and Nielsen, J. (2012) Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in *Saccharomyces cerevisiae*. *PLoS one*, **7**, e42475.
104. Orlandi, I., Casatta, N. and Vai, M. (2012) Lack of Ach1 CoA-transferase triggers apoptosis and decreases chronological lifespan in yeast. *Front. Oncol.*, **2**, 67.
105. Takahashi, H., McCaffery, J.M., Irizarry, R.A. and Boeke, J.D. (2006) Nucleocytoplasmic acetyl-coenzyme A synthetase is required for histone acetylation and global transcription. *Mol. Cell*, **23**, 207-217.
106. van Rossum, H.M., Kozak, B.U., Niemeijer, M.S., Duine, H.J., Luttik, M.A.H., Boer, V.M., Kötter, P., Daran, J.-M.G., van Maris, A.J.A. and Pronk, J.T. (2016) Alternative reactions at the interface of glycolysis and citric acid cycle in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **16**.
107. Regev-Rudzki, N., Karniely, S., Ben-Haim, N.N. and Pines, O. (2005) Yeast aconitase in two locations and two metabolic pathways: seeing small amounts is believing. *Mol. Biol. Cell*, **16**, 4163-4171.
108. Taylor, K.M., Kaplan, C.P., Gao, X. and Baker, A. (1996) Localization and targeting of isocitrate lyases in *Saccharomyces cerevisiae*. *Biochem J.*, **319**, 255-262.
109. Chaves, R.S., Herrero, P., Ordiz, I., Angeles del Brio, M. and Moreno, F. (1997) Isocitrate lyase localisation in *Saccharomyces cerevisiae* cells. *Gene*, **198**, 165-169.
110. Kunze, M., Kragler, F., Binder, M., Hartig, A. and Gurvitz, A. (2002) Targeting of malate synthase 1 to the peroxisomes of *Saccharomyces cerevisiae* cells depends on growth on oleic acid medium. *Eur. J. Biochem.*, **269**, 915-922.
111. Hiltunen, J.K., Mursula, A.M., Rottensteiner, H., Wierenga, R.K., Kastaniotis, A.J. and Gurvitz, A. (2003) The biochemistry of peroxisomal beta-oxidation in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.*, **27**, 35-64.
112. Kim, K.S., Rosenkrantz, M.S. and Guarente, L. (1986) *Saccharomyces cerevisiae* contains two functional citrate synthase genes. *Mol. Cell. Biol.*, **6**, 1936-1942.
113. Hartig, A., Simon, M.M., Schuster, T., Daugherty, J.R., Yoo, H.S. and Cooper, T.G. (1992) Differentially regulated malate synthase genes participate in carbon and nitrogen metabolism of *S. cerevisiae*. *Nucleic Acids Res.*, **20**, 5677-5686.
114. Lee, Y.J., Jang, J.W., Kim, K.J. and Maeng, P.J. (2011) TCA cycle-independent acetate metabolism via the glyoxylate cycle in *Saccharomyces cerevisiae*. *Yeast*, **28**, 153-166.
115. Fernández, E., Moreno, F. and Rodicio, R. (1992) The *ICL1* gene from *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, **204**, 983-990.
116. Minard, K.I. and McAlister-Henn, L. (1991) Isolation, nucleotide sequence analysis, and disruption of the *MDH2* gene from *Saccharomyces cerevisiae*: evidence for three isozymes of yeast malate dehydrogenase. *Mol. Cell. Biol.*, **11**, 370-380.
117. Luttik, M.A., Kötter, P., Salomons, F.A., van der Klei, I.J., van Dijken, J.P. and Pronk, J.T. (2000) The *Saccharomyces cerevisiae* *ICL2* gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme A metabolism. *J. Bacteriol.*, **182**, 7007-7013.
118. Fernández, E., Fernández, M. and Rodicio, R. (1993) Two structural genes are encoding malate synthase isoenzymes in *Saccharomyces cerevisiae*. *FEBS Lett.*, **320**, 271-275.
119. Van Dijken, J.P. and Scheffers, W.A. (1986) Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiol. Rev.*, **1**, 199-224.
120. Ansell, R., Granath, K., Hohmann, S., Thevelein, J.M. and Adler, L. (1997) The two isoenzymes for yeast NAD<sup>+</sup>-dependent glycerol 3-phosphate dehydrogenase encoded by *GPD1* and *GPD2* have distinct roles in osmoadaptation and redox regulation. *EMBO J.*, **16**, 2179-2187.
121. Nevoigt, E. and Stahl, U. (1997) Osmoregulation and glycerol metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.*, **21**, 231-241.
122. Björkqvist, S., Ansell, R., Adler, L. and Lidén, G. (1997) Physiological response to anaerobicity of glycerol-3-phosphate dehydrogenase mutants of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **63**, 128-132.

123. Albertyn, J., Hohmann, S., Thevelein, J.M. and Prior, B.A. (1994) *GPD1*, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in *Saccharomyces cerevisiae*, and its expression is regulated by the high-osmolarity glycerol response pathway. *Mol. Cell. Biol.*, **14**, 4135-4144.
124. Al-Saryi, N.A., Al-Hejjaj, M.Y., van Roermund, C.W.T., Hulmes, G.E., Ekal, L., Payton, C., Wanders, R.J.A. and Hettema, E.H. (2017) Two NAD-linked redox shuttles maintain the peroxisomal redox balance in *Saccharomyces cerevisiae*. *Sci. Rep.*, **7**, 11868.
125. Eriksson, P., André, L., Ansell, R., Blomberg, A. and Adler, L. (1995) Cloning and characterization of *GPD2*, a second gene encoding sn-glycerol 3-phosphate dehydrogenase (NAD+) in *Saccharomyces cerevisiae*, and its comparison with *GPD1*. *Mol. Microbiol.*, **17**, 95-107.
126. Nissen, T.L., Hamann, C.W., Kielland-Brandt, M.C., Nielsen, J. and Villadsen, J. (2000) Anaerobic and aerobic batch cultivations of *Saccharomyces cerevisiae* mutants impaired in glycerol synthesis. *Yeast*, **16**, 463-474.
127. Norbeck, J., Pählman, A.K., Akhtar, N., Blomberg, A. and Adler, L. (1996) Purification and characterization of two isoenzymes of DL-glycerol-3-phosphatase from *Saccharomyces cerevisiae*. Identification of the corresponding *GPP1* and *GPP2* genes and evidence for osmotic regulation of Gpp2p expression by the osmosensing mitogen-activated protein kinase signal transduction pathway. *J. Biol. Chem.*, **271**, 13875-13881.
128. Pahlman, A.K., Granath, K., Ansell, R., Hohmann, S. and Adler, L. (2001) The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress. *J. Biol. Chem.*, **276**, 3555-3563.
129. Kolarov, J., Kolarova, N. and Nelson, N. (1990) A third ADP/ATP translocator gene in yeast. *J. Biol. Chem.*, **265**, 12711-12716.
130. Sasaki, M., Kumagai, H., Takegawa, K. and Tohda, H. (2013) Characterization of genome-reduced fission yeast strains. *Nucleic Acids Res.*, **41**, 5382-5399.
131. Murakami, K., Tao, E., Ito, Y., Sugiyama, M., Kaneko, Y., Harashima, S., Sumiya, T., Nakamura, A. and Nishizawa, M. (2007) Large scale deletions in the *Saccharomyces cerevisiae* genome create strains with altered regulation of carbon metabolism. *Appl. Microbiol. Biotechnol.*, **75**, 589-597.
132. Richardson, S.M., Mitchell, L.A., Stracquadiano, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y. *et al.* (2017) Design of a synthetic yeast genome. *Science*, **355**, 1040-1044.
133. Shao, Y., Lu, N., Wu, Z., Cai, C., Wang, S., Zhang, L.-L., Zhou, F., Xiao, S., Liu, L., Zeng, X. *et al.* (2018) Creating a functional single-chromosome yeast. *Nature*, **560**, 331-335.
134. Wisselink, H.W., Cipollina, C., Oud, B., Crimi, B., Heijnen, J.J., Pronk, J.T. and van Maris, A.J. (2010) Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered *Saccharomyces cerevisiae*. *Metab. Eng.*, **12**, 537-551.
135. Shen, X.-X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase, M.A.B., Wisecaver, J.H., Wang, M., Doering, D.T. *et al.* (2018) Tempo and mode of genome evolution in the budding yeast subphylum. *Cell*, **175**, 1533-1545.e1520.
136. Kuzmin, E., VanderSluis, B., Nguyen Ba, A.N., Wang, W., Koch, E.N., Usaj, M., Khmelinskii, A., Usaj, M.M., van Leeuwen, J., Kraus, O. *et al.* (2020) Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science*, **368**, eaaz5667.
137. Bakker, B.M., Overkamp, K.M., van Maris, A.J., Kötter, P., Luttik, M.A., van Dijken, J.P. and Pronk, J.T. (2001) Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.*, **25**, 15-37.
138. Miyagi, H., Kawai, S. and Murata, K. (2009) Two sources of mitochondrial NADPH in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **284**, 7553-7560.

139. Swiegers, J.H., Dippenaar, N., Pretorius, I.S. and Bauer, F.F. (2001) Carnitine-dependent metabolic activities in *Saccharomyces cerevisiae*: three carnitine acetyltransferases are essential in a carnitine-dependent strain. *Yeast*, **18**, 585-595.
140. Bakker, B.M., Bro, C., Kötter, P., Luttkik, M.A., van Dijken, J.P. and Pronk, J.T. (2000) The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in *Saccharomyces cerevisiae*. *J. Bacteriol.*, **182**, 4730-4737.
141. Wieczorke, R., Krampe, S., Weierstall, T., Freidel, K., Hollenberg, C.P. and Boles, E. (1999) Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.*, **464**, 123-128.
142. Wijsman, M., Swiat, M.A., Marques, W.L., Hettinga, J.K., van den Broek, M., Torre Cortés, P., Mans, R., Pronk, J.T., Daran, J.M. and Daran-Lapujade, P. (2019) A toolkit for rapid CRISPR-SpCas9 assisted construction of hexose-transport-deficient *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.*, **19**, foy107.
143. Brickwedde, A., Brouwers, N., van den Broek, M., Gallego Murillo, J.S., Fraiture, J.L., Pronk, J.T. and Daran, J.-M.G. (2018) Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357T. *Front. Microbiol.*, **9**.
144. Brown, C.A., Murray, A.W. and Verstrepen, K.J. (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr. Biol.*, **20**, 895-903.
145. Aslan, S., Noor, E. and Bar-Even, A. (2017) Holistic bioengineering: rewiring central metabolism for enhanced bioproduction. *Biochem J.*, **474**, 3935-3950.
146. Papagianni, M. (2012) Recent advances in engineering the central carbon metabolism of industrially important bacteria. *Microb. Cell Fact.*, **11**, 50.
147. François, J.M., Lachaux, C. and Morin, N. (2020) Synthetic biology applied to carbon conservative and carbon dioxide recycling pathways. *Front. Bioeng. Biotechnol.*, **7**, 446.
148. Sánchez-Pascuala, A., Fernández-Cabezón, L., de Lorenzo, V. and Nikel, P.I. (2019) Functional implementation of a linear glycolysis for sugar catabolism in *Pseudomonas putida*. *Metab. Eng.*, **54**, 200-211.
149. van Rossum, H.M., Kozak, B.U., Pronk, J.T. and van Maris, A.J.A. (2016) Engineering cytosolic acetyl-coenzyme A supply in *Saccharomyces cerevisiae*: Pathway stoichiometry, free-energy conservation and redox-cofactor balancing. *Metab. Eng.*, **36**, 99-115.
150. Liu, Q., Yu, T., Li, X., Chen, Y., Campbell, K., Nielsen, J. and Chen, Y. (2019) Rewiring carbon metabolism in yeast for high level production of aromatic chemicals. *Nat. Commun.*, **10**, 4976.
151. Meadows, A.L., Hawkins, K.M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., Dahl, R.H., Tai, A., Mahatdejkul-Meadows, T., Xu, L. *et al.* (2016) Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature*, **537**, 694-697.
152. Koopman, F., Beekwilder, J., Crimi, B., van Houwelingen, A., Hall, R.D., Bosch, D., van Maris, A.J., Pronk, J.T. and Daran, J.M. (2012) *De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. *Microb. Cell Fact.*, **11**, 155.
153. Boonekamp, F.J., Dashko, S., Duiker, D., Gehrmann, T., van den Broek, M., den Ridder, M., Pabst, M., Robert, V., Abeel, T., Postma, E.D. *et al.* (2020) Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences. *ACS Synth. Biol.*, **9**, 1361-1375.
154. Entian, K.-D. and Kötter, P. (2007) 25 Yeast genetic strain and plasmid collections. In: Stansfield, I. and Stark, M. J. R. (eds.), *Methods in Microbiology*. Academic Press, Vol. **36**, pp. 629-666.
155. Nijkamp, J.F., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttkik, M.A., Daran-Lapujade, P., Vongsangnak, W., Nielsen, J., Heijne, W.H. *et al.* (2012) *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.*, **11**, 36.

156. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J. and Daran, J.M. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **15**, 1-15.
157. Verduyn, C., Postma, E., Scheffers, W.A. and Van Dijken, J.P. (1992) Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, **8**, 501-517.
158. Solis-Escalante, D., Kuijpers, N.G., Bongaerts, N., Bolat, I., Bosman, L., Pronk, J.T., Daran, J.M. and Daran-Lapujade, P. (2013) *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **13**, 126-139.
159. Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.*, **350**, 87-96.
160. Looke, M., Kristjuhan, K. and Kristjuhan, A. (2011) Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques*, **50**, 325-328.
161. Salazar, A.N., Gorter de Vries, A.R., van den Broek, M., Wijsman, M., de la Torre Cortes, P., Brickwedde, A., Brouwers, N., Daran, J.G. and Abeel, T. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.*, **17**.
162. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.
163. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
164. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abuelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K. *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS one*, **9**, e112963.
165. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.*, **14**, 178-192.
166. Nijkamp, J.F., van den Broek, M.A., Geertman, J.M., Reinders, M.J., Daran, J.M. and de Ridder, D. (2012) *De novo* detection of copy number variation by co-assembly. *Bioinformatics*, **28**, 3195-3202.
167. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, **27**, 722-736.
168. Douma, R.D., de Jonge, L.P., Jonker, C.T., Seifar, R.M., Heijnen, J.J. and van Gulik, W.M. (2010) Intracellular metabolite determination in the presence of extracellular abundance: Application to the penicillin biosynthesis pathway in *Penicillium chrysogenum*. *Biotechnol. Bioeng.*, **107**, 105-115.
169. Wu, L., Mashego, M.R., van Dam, J.C., Proell, A.M., Vinke, J.L., Ras, C., van Winden, W.A., van Gulik, W.M. and Heijnen, J.J. (2005) Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly <sup>13</sup>C-labeled cell extracts as internal standards. *Anal. Biochem.*, **336**, 164-171.
170. de Jonge, L.P., Buijs, N.A., ten Pierick, A., Deshmukh, A., Zhao, Z., Kiel, J.A., Heijnen, J.J. and van Gulik, W.M. (2011) Scale-down of penicillin production in *Penicillium chrysogenum*. *Biotechnol. J.*, **6**, 944-958.
171. Niedenfür, S., ten Pierick, A., van Dam, P.T., Suarez-Mendez, C.A., Nöh, K. and Wahl, S.A. (2016) Natural isotope correction of MS/MS measurements for metabolomics and (<sup>13</sup>C) fluxomics. *Biotechnol. Bioeng.*, **113**, 1137-1147.
172. Byrne, K.P. and Wolfe, K.H. (2005) The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.*, **15**, 1456-1461.



# Chapter 3

## What's wrong with SwYG? Reaching the limits of *Saccharomyces cerevisiae* molecular and analytical toolbox

Eline D. Postma<sup>#</sup>

Ewout Knibbe<sup>#</sup>

Francine J. Boonekamp

Sofia Dashko

Jordi Geelhoed

Anne-Marijn Maat

Marcel van den Broek

Pascale Daran-Lapujade

<sup>#</sup>Eline D. Postma and Ewout Knibbe contributed equally to this chapter and should be considered co-first authors.

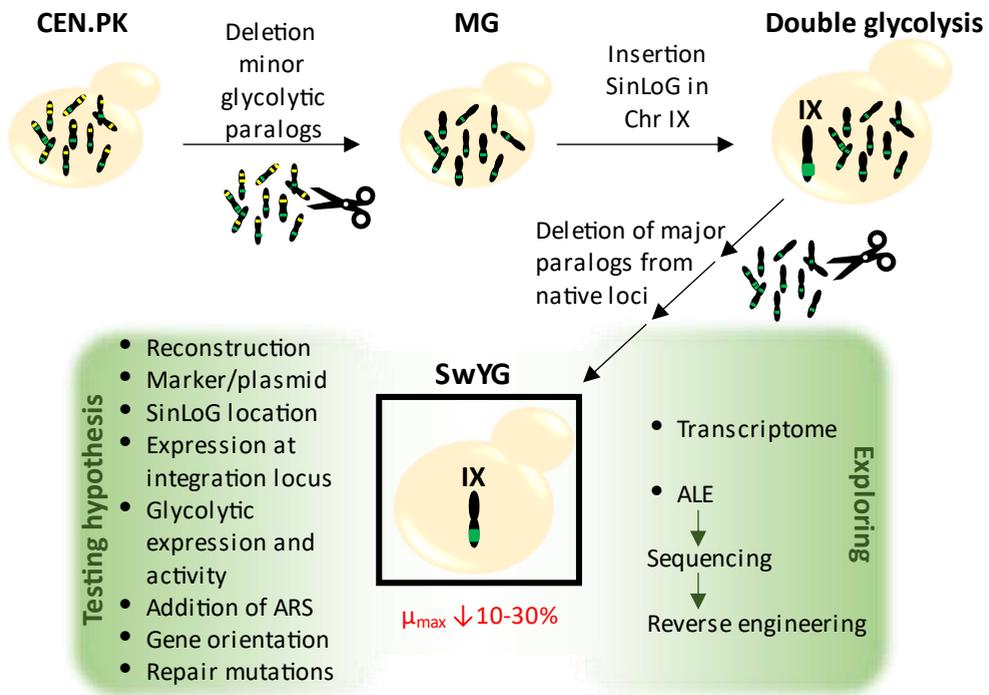
## Abstract

The construction of powerful cell factories requires extensive remodeling of microbial genomes, entailing many rounds of transformations to perform the large number of desired gene modifications. However, increasing the number of genetic interventions inevitably increases the occurrence of unwanted mutations and effects. Using glycolysis as paradigm, a previous study developed a *Saccharomyces cerevisiae* strain in which the glycolytic genes, relocalized to a single locus, can be easily swapped by any new design, thereby enabling fast and easy remodeling of the entire pathway. After 27 genetic modifications performed in 43 transformation rounds, the Switchable Yeast Glycolysis (SwYG) strain grew ca. 20 % slower than its ancestor with native glycolysis design. Exploring the cause of this slower growth rate, the present study reflects on the genetic and analytical challenges encountered by extensive strain construction programs and provides design guidelines for integration of large constructs in the yeast genome.

## Introduction

Converting microbial cells, optimized to operate in defined natural environments, into powerful factories with high product yields and productivity in harsh industrial environments requires extensive remodelling of their genome. Due to the limited predictability of biological systems, the construction of such cell factories requires the iteration of design-build-test-learn cycles (1). In the build phase of the cycle, microbial cells undergo many rounds of genetic manipulations meant to add new metabolic routes, but also to rewire native pathways and cellular machineries, indispensable to optimally supply energy-rich moieties, redox equivalents, precursors and co-factors to the newly added routes (2-4). Despite the recent developments in DNA editing with CRISPR-Cas technologies, extensive rewiring of native pathways and processes still presents a technical challenge. Taking as paradigm the popular model organism and synthetic biology platform *Saccharomyces cerevisiae*, rewiring entire native pathways is complicated by a high degree of genetic redundancy and the scattering of genes over the 12 Mb and 16 chromosome-containing genome. For instance, glycolysis and alcoholic fermentation, central pathways for carbon conversion, cover a set of twelve biochemical reactions important for industrial applications of this yeast. Most of these twelve reactions are catalyzed by multiple iso-enzymes, resulting in a set of 26 genes scattered over the yeast chromosomes. To facilitate remodeling of glycolysis and alcoholic fermentation in yeast, previous studies reported the construction of a strain in which genes encoding the 13 minor paralogs of glycolysis and fermentation were deleted and the remaining 13 major paralogs were relocalized to a single locus (5,6). In a process named pathway swapping, the resulting Switchable Yeast Glycolysis strain (SwYG) enables the two-step replacement of the entire glycolytic and fermentation pathways by any new design, and their translocation to any locus on native or synthetic chromosomes (5,7). The SwYG strain, harboring a minimized set of glycolytic and fermentation genes clustered on a single chromosomal locus, grew remarkably well but consistently displayed a 10 - 30 % reduction in growth rate in synthetic chemically defined medium with glucose as carbon source, as compared to a strain with a complete set of glycolytic enzymes scattered over the yeast chromosomes (5). This decrease in growth rate could not be attributed to the removal of the minor glycolytic paralogs, as a similar strain in which the same minimized set of genes was still present in their native loci grew just as fast as the non-minimized control strain, and therefore faster than SwYG (6). While the SwYG strain is a valuable addition to *S. cerevisiae* molecular toolbox, this decrease in growth rate was puzzling and could hold key information regarding the genetic design of large DNA constructs. If the co-localization of thirteen highly expressed genes had deleterious effects on the host, it should be taken into consideration for the design and engineering of cells factories, as well as the construction of synthetic chromosomes and genomes.

Many mechanisms could potentially cause the growth defect of the SwYG strain, and this study attempts to identify which ones are at play (Figure 1). The growth defect could be directly caused by the integration of the set of 13 clustered genes (called SinLoG for Single Locus Glycolysis) in the yeast chromosomes, by altering the expression either of genes in the SinLoG itself or of genes neighboring the SinLoG integration site. Alternatively, the SinLoG might affect DNA replication either by causing collision between the replication machinery and the polymerase processing the highly transcribed colocalized glycolytic genes, or by excessively increasing the distance between adjacent autonomously replicating sequences (ARS) (8,9). The growth defect of the SwYG strain could also be indirectly caused by the mutations voluntarily (targeted deletion of genes, presence of selection markers and plasmids) or involuntarily (random mutations) induced during strain construction. Using a combination of targeted genetic engineering, and explorative evolutionary and reverse



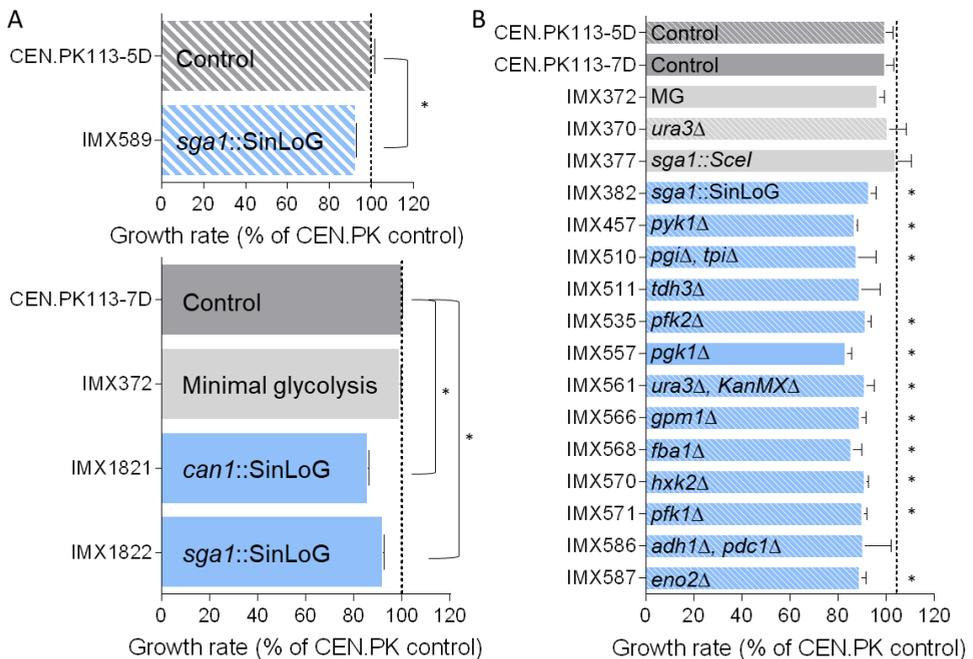
**Figure 1 - Schematic overview of the experimental approach.** Strain construction strategy leading to the Switchable Yeast Glycolysis (SwYG) strain with a Single Locus Glycolysis (SinLoG). First the Minimal Glycolysis (MG) strain was constructed by deletion of the minor glycolytic paralogs without measurable phenotypic effects (6). Subsequently the major glycolytic paralogs were integrated as SinLoG in chromosome IX and deleted from their native loci, resulting in the SwYG strain (5). The SwYG strain displayed a decreased growth rate as compared to the MG strain. Several hypotheses for the mechanism underlying this growth rate decrease were tested, as well as explorative transcriptomics and adaptive laboratory evolution (ALE).

engineering, this study makes full use of the rich *S. cerevisiae* molecular and analytical toolbox to directly test these hypotheses and to discover yet unpredicted mechanisms.

## Results

### Confirmation that SinLoG integration in the MG strain causes a decrease in yeast specific growth rate

First, the specific growth rate of the SwYG strain (IMX589) and its ancestor the MG strain was measured again, confirming that the SwYG strain grew slower than the MG strain harboring a minimal set of glycolytic genes scattered over the yeast genome (Figure 2). The SwYG strain is auxotrophic for uracil and harbors the *amdS* marker in the SinLoG. Repairing the uracil auxotrophy by integration of the *URA3* gene at the



**Figure 2 - Specific growth rates of single locus glycolysis strains.** The specific growth rates of various strains with a SinLoG integration are shown in blue and reference strains are in grey. Prototrophic strains are shown as filled column while uracil auxotrophic strains are depicted with dashed columns. All growth rates are normalized to the CEN.PK113-7D (prototrophic) or CEN.PK113-5D (uracil auxotrophic) control strains from the same experiment. **A)** Specific growth rates determined in duplicate shake flask cultures for several reference strains and single locus glycolysis strains. **B)** Growth rates of intermediate strains with glycolytic gene deletions during strain construction of the single locus glycolysis strains. Each modification follows from the previous one from top (MG strain) to bottom (SwYG strain). Growth rates are measured in 96-wells plates in at least culture triplicate. Statistically significant differences relative to the auxotrophic or prototrophic CEN.PK reference strain are indicated (\*  $P < 0.05$ , t-test 2-tailed, homoscedastic).

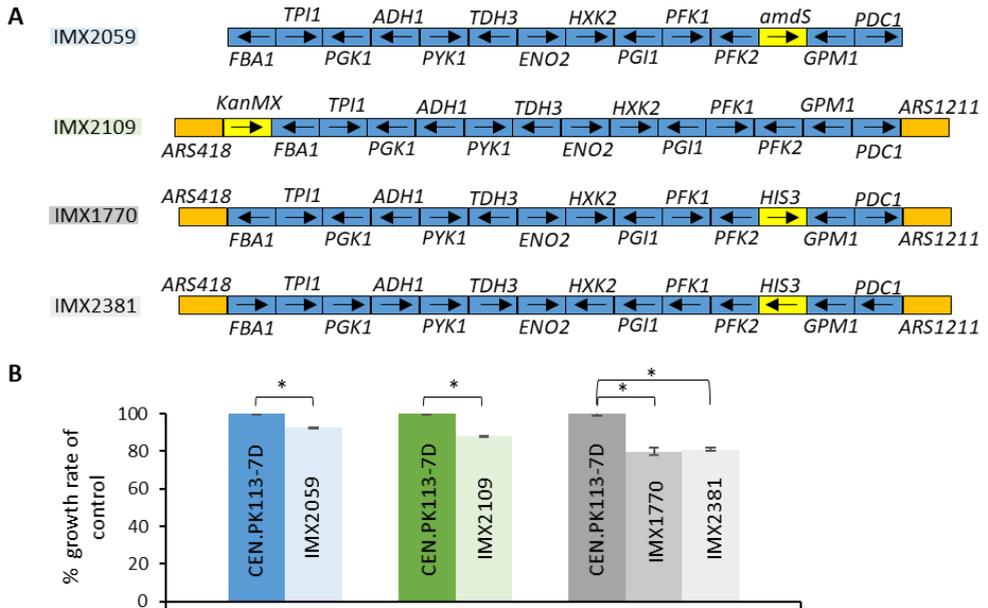
*TDH1* locus and removing the *amdS* marker did not affect growth rate, indicating that selection markers were not accountable for the SwYG strain slower growth phenotype (strain IMX1822, Figure 2). Re-localization of glycolytic genes to a single locus required 18 transformation rounds, starting by the integration of the SinLoG and followed by the sequential deletion of the 13 glycolytic genes from their native loci (Figure 1). To test which steps were involved in the SwYG strain slower growth, the specific growth rate of all relevant intermediate strains in the construction steps between the MG and the SwYG strains were measured. It revealed that, after integration of the SinLoG in the MG strain the following deletion steps did not significantly affect growth. It was further confirmed that the growth rate decrease caused by SinLoG integration was not an artefact, by repeating the transformation of the MG strain with the SinLoG DNA parts (strain IMX382, Supplementary Figure S1). Altogether these results confirmed that the integration of the SinLoG cassette caused a specific growth rate reduction and suggested that it was the most probable cause for the SwYG strain's slower growth rate phenotype.

### **SwYG decreased growth rate is not caused by secondary effects at the integration site**

While the SinLoG seemed to be responsible for the SwYG strain's reduced growth rate, this effect might not be a direct consequence of the genetic clustering of the glycolytic genes. It could result from unexpected genetic effects at the integration site. To explore this hypothesis, prototrophic strains with the SinLoG integrated in the *CAN1* (chromosome V, IMX1821) or *SGA1* locus (chromosome IX, IMX1822) were constructed. However, this resulted in a similar growth rate as measured before for prototrophic SwYG strains IMX605 and IMX606 and auxotrophic strain IMX589 ((5) and Figure 2). The SinLoG is a 35 kb long DNA stretch, considering that autonomously replicating sequences (ARS) are distributed on average every 30 kb across the yeast genome, integration of the SinLoG might hinder replication (5,9). However, flanking the SinLoG by additional ARS sequences (strain IMX2109 Figure 3, strain lineage in Supplementary Figure S2) (7), had no positive effect on the growth rate of the SwYG strain. The presence of the SinLoG might also impact the expression of genes surrounding its integration site, however their transcript levels were largely unaffected by the SinLoG presence (Supplementary Figure S3). Yeast appeared therefore extremely robust to the disruption of its chromosomes by integration of a 35 kb highly transcribed DNA sequence.

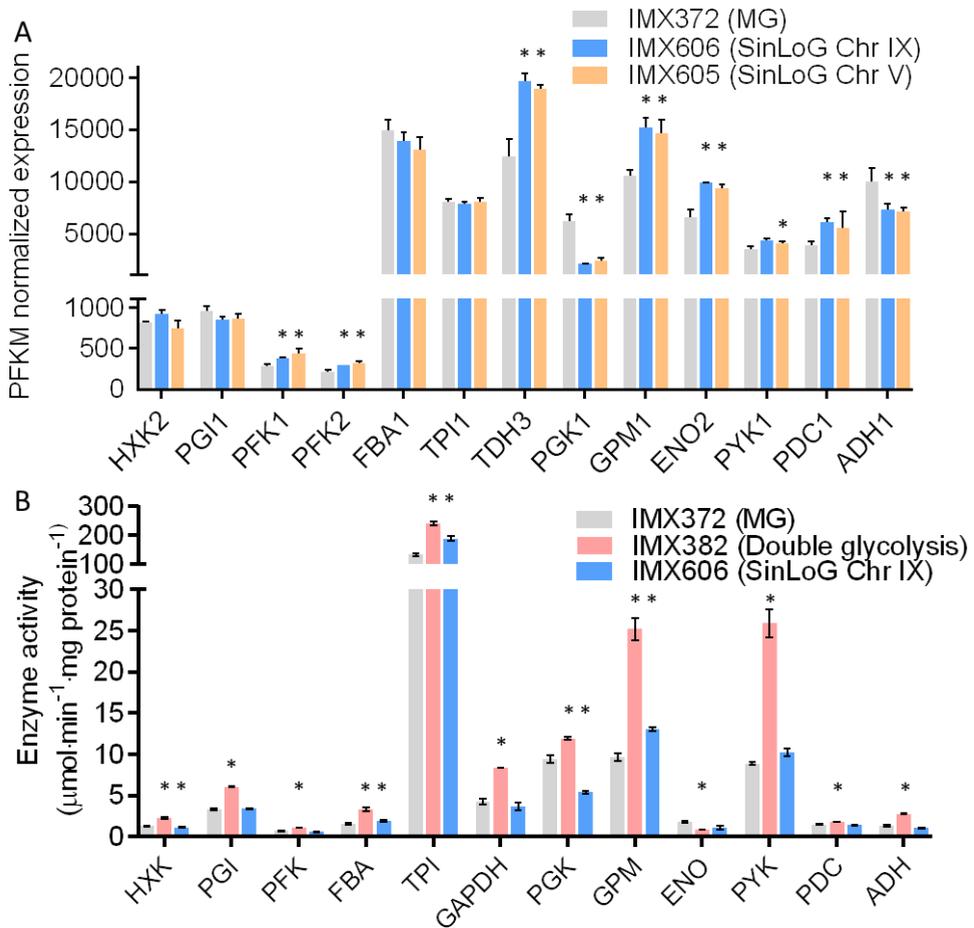
### **SwYG decreased growth rate is not caused by expression of genes from the SinLoG**

As the indirect effects considered so far were not responsible for the SwYG strain's slower growth rate, possible direct effects of the SinLoG itself were explored. Glycolysis is an essential pathway for growth on glucose as sole carbon source, altering its



**Figure 3 - Effects of ARS sequences and gene orientation.** **A)** Configurations of single locus glycolysis gene clusters with ARS sequences and various gene orientations. **B)** Specific growth rate of strains with SinLoG designed represented in panel A. Bars represent the average and standard deviation of duplicate shake flask cultures expressed as a percentage of the control strain CEN.PK113-7D. Statistically significant differences are indicated (\*  $P < 0.05$ , t-test 2-tailed, homoscedastic).

expression might therefore affect yeast growth. Transcript levels of the glycolytic genes and *in vitro* activity of the glycolytic enzymes were measured in the SwYG and control strains. Despite the re-localization and clustering of the glycolytic genes, their expression was very similar in the SwYG strain compared to the minimal glycolysis strain IMX372 (MG), with only *PGK1* expression changing more than 2-fold (Figure 4A). Accordingly, activity of the glycolytic enzymes was largely comparable to the MG control strain. Only P<sub>gk1</sub> activity was significantly decreased by 43 % in SwYG as compared to MG (Figure 4B). However, this reduced catalytic activity of P<sub>gk1</sub> is most likely not responsible for the lower growth rate of the SwYG strain as, like most glycolytic enzymes, P<sub>gk1</sub> is present with a large overcapacity and even with reduced activity it operates far from saturation (10). Indeed, for P<sub>gk1</sub>, the degree of saturation estimated from the glucose uptake rate and enzyme activity only increased from 19 % in the MG strain to 26 % in the SwYG strain. Next to their key role in central carbon metabolism, several glycolytic enzymes (hexokinase, aldolase and enolase) are also involved in moonlighting functions outside glycolysis (11), however, since glycolytic transcript levels and *in vitro* activity are similar whether the glycolytic genes are



**Figure 4 - Transcription and enzyme activity of the glycolytic genes in the single locus glycolysis strains.** A) PFKM-normalized transcript levels of the glycolytic genes in strains IMX372, IMX606 and IMX605 are compared. Error bars represent the standard error of the mean of independent culture duplicates. \* indicates significant difference compared to IMX372 ( $P_{\text{adj}} < 0.05$ ). B) Enzyme activities measured in cell-free extracts of the minimal glycolysis strain IMX372, the double glycolysis strain IMX382 and single locus glycolysis strain IMX606 (SwYG, SinLoG-IX). The data for IMX372 and IMX606 shown here is reproduced from (5) and is the average of at least four replicates. IMX382 was measured in duplicates. Error bars represent the standard error of the mean. \* indicates significant difference to IMX372 ( $P < 0.05$ , student t-test, 2 tailed, homoscedastic).

scattered over the chromosomes or co-localized, there is no evidence suggesting that moonlighting functions play a role in the SwYG strain reduced growth rate.

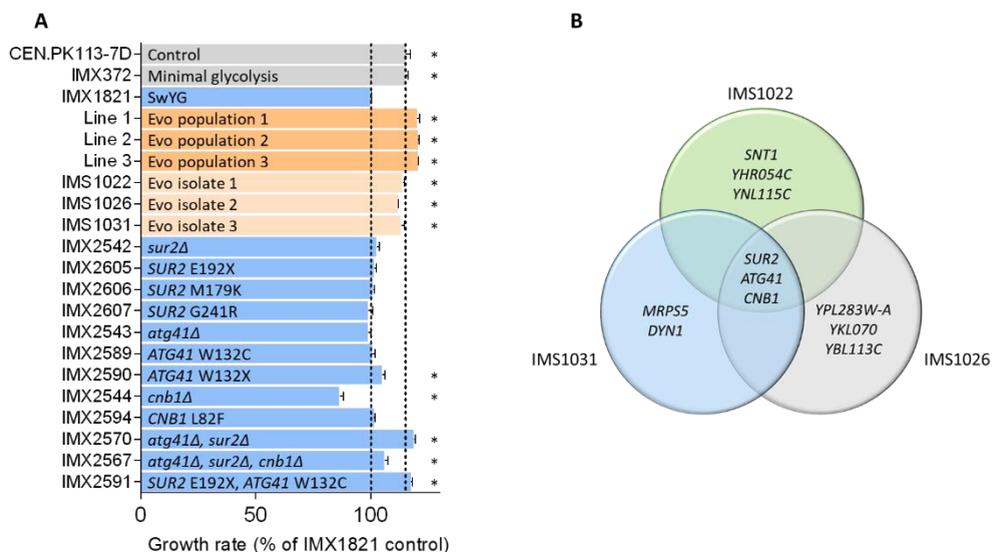
There is yet another aspect to consider for the genetic co-localization of genes such as the glycolytic genes. Glycolytic genes are intensively and continuously transcribed in

yeast and their transcripts are amongst the most abundant. While the strong and constitutive glycolytic promoters are an asset for yeast molecular biology toolbox, their co-localization might be deleterious for the host. During cell division, the replication machinery needs to access DNA, which might lead to head-on collisions with RNA polymerases in the SinLoG transcriptional hotspot (8). In such a scenario, DNA replication might be stalled, and even lead to DNA damage, which could result in a slower growth rate. Since co-directional collisions are less detrimental than head-on collisions, aligning the orientation of transcription units with the nearest ARS sequences could attenuate conflicts between replication and transcription. This possibility was assessed by the design, construction and growth rate measurement of two strains with modified transcription-unit's orientation, one without (IMX1770) and one with co-directionality of transcription and replication (IMX2381). These strains showed no significant difference in growth rate (Figure 3). The SwYG strain (auxotrophic IMX589 and prototrophic IMX1821) did not display alterations in the duration of the S phase of the cell cycle, which could be expected if DNA replication was competing with transcription (Supplementary Figure S4). Although these results do not entirely exclude potential effects of the SinLoG on DNA replication, they strongly suggest that it is not the main cause of the SwYG strain's growth defect.

### **Discovery-driven approaches to elucidate the molecular basis of SwYG slow growth rate**

While the targeted engineering strategy enabled the elimination of factors that were not causal for the SwYG strain's slower growth rate, it failed to directly identify the responsible factors. In an attempt to identify these mechanisms, gene expression profiles in the SwYG strains with SinLoG in chromosome V and IX and their ancestor the MG strain were mined. There again, while genes were up- and down-regulated in the strains with a SinLoG as compared to the MG strain, transcriptome analysis did not identify specific genes, sets of genes or mechanisms potentially involved in the SwYG strain's slow growth (Supplementary Figure S5).

As transcriptome analysis proved inconclusive, an adaptive laboratory evolution approach was undertaken. To increase the growth rate of SwYG, repeated transfers of IMX1821 and the control strain CEN.PK113-7D were performed on chemically defined glucose medium for approximately 525 generations (Figure 1). Three independent evolution lines successfully led to SwYG populations with growth rates comparable to that of the reference strain CEN.PK113-7D (Figure 5A and Supplementary Figure S6). Three single colony isolates were selected from independent evolution lines of SwYG and the control strain, and their genome was sequenced. All three independent single locus glycolysis evolution lines carried mutations in *ATG41*, *CNB1* and *SUR2* (Figure 5B and for strain lineage Supplementary Figure S6). *CNB1*, encoding a regulatory subunit of calcineurin, was also mutated in one of the evolution lines of the CEN.PK113-7D



**Figure 5 - Physiological and genetic characterization of the evolved and reverse engineered strains derived from the SwYG strain.** A) Specific growth rates of the evolved population of the SwYG strain, single colony isolates of the evolved populations and reverse engineered strains, shown as percentage of reference SwYG strain IMX1821. Control strains without a single locus glycolysis are shown in grey, evolved strains in orange and engineered SwYG strains are shown in blue, dashed lines indicate growth rates of the reference strains IMX1821 and CEN.PK113-7D. Significant differences compared to strain IMX1821 are indicated (\*,  $P < 0.05$ , t-test, homoscedastic). B) Venn diagram showing genes in which mutations were identified in the evolved strains derived from IMX1821.

control strain, and was most likely not the causal mutation for the SwYG strain phenotype (Supplementary Figure S7). Introducing the *CNB1* mutation of the evolved strains in the non-evolved SwYG strain confirmed this hypothesis (Figure 5). Conversely, *ATG41* and *SUR2* were not mutated in the control strain evolution lines and could therefore be involved in the slower growth phenotype of the SwYG strain. *SUR2* is involved in sphingolipids biosynthesis and encodes a sphinganine C4-hydroxylase (12-14). Sphingolipids are important membrane components and act as messengers and bioactive molecules in cell growth, senescence, autophagy and apoptosis (15). In all three evolution lines the Sur2 fatty acid hydroxylase domain (amino acids 162 to 297) was mutated. In one of the evolved strains the mutation caused the occurrence of an early stop codon, suggesting that the mutations most likely caused a loss of function of this non-essential gene (16). Both removal and reverse-engineering of the *SUR2* mutations were performed in the SwYG parental strain IMX1821, but did not result in a measurable phenotypic effect (Figure 5). The exact function of *ATG41* has not been identified, but it is associated with the formation of the autophagosome (17). Similar to Sur2, several mutations found in Atg41 truncated the protein, removing its extreme

C-terminal domain (amino acids 127-136). This region is essential for the interaction with protein partners (more particularly Atg9) and thus correct subcellular distribution and autophagy activity (17). As *SUR2*, *ATG41* was therefore deleted both in the SwYG parental strain IMX1821 and modified with the mutations found in the evolved strain. While individual mutations hardly affected the growth rate of the SwYG strain, the combined deletion or mutation of both *SUR2* and *ATG41* successfully restored the growth rate of the evolved strains (Figure 5). These results demonstrated that mutations in both genes are required to fix the slower growth rate phenotype of the SwYG strain.

## Discussion

With the ambition to consolidate the SwYG strain as platform for glycolysis swapping and to identify design principles for integration of large, modular pathways in *S. cerevisiae*, we aimed to characterize and identify underlying factors of the slower growth phenotype observed upon genetic clustering of the glycolytic pathway in the SwYG strain.

This in-depth study revealed that genetic factors directly or indirectly related to the integration of the SinLoG were not responsible for the SwYG strain slow growth rate, a promising outcome for integration of large DNA constructs in yeast. ALE rather identified the involvement of autophagy and sphingolipids biosynthesis in the slower growth rate phenotype, a response difficult to reconcile with the genetic interventions performed in the SwYG strain. These results might yet be explained if we consider an alternative theory, initially discarded due to its unlikelihood, but that might, in the light of all the data obtained in this study, explain the SwYG strain phenotype. The SinLoG integration in the MG strain causes a duplication of the glycolysis and fermentation genes. Glycolysis and fermentation proteins are amongst the most abundant proteins in *S. cerevisiae* cells, representing around 15% of total cell proteins under fermentative conditions (18). Integration of a second copy of these genes might increase the abundance of these proteins and, considering that proteins are extremely costly to synthesize, cause an energetic burden to the cells, reflected in a decreased growth rate. Enzyme assays confirmed the increased specific activity of most of these enzymes by 1.5- to 2-fold in the double glycolysis strain (IMX382, Figure 4). This increase in activity, with the reasonable assumption that it represents an equivalent increase in protein abundance, should result in a decrease in growth rate. Indeed, individual overexpression of glycolytic enzymes has been shown to lead to decreased growth rates when around 15% of the total cellular protein is made up of gratuitous glycolytic protein (19). Similarly, overexpression of fluorescent proteins to high percentages of the proteome (estimated around 10-15%) has been shown to lead to a growth rate

decrease of up to 15% from the cost of protein production (20). This overexpression alone therefore could explain the decrease in growth rate measured for the double glycolysis strain (IMX382, Figure 2).

Building on the realistic working hypothesis that protein burden is the major cause of the decreased growth rate in the double glycolysis strain, subsequent removal of the gene duplicates in this strain is expected to alleviate this burden and to restore growth rate to the levels of the MG strain. However, deletion of the 13 duplicated genes from the double glycolysis strain did not alter growth rate (Figure 2). This stability in growth rate upon deletion of the gene duplicates could reflect the compensation of the expected positive effect of sequential deletion by the occurrence of deleterious events during strain construction. While analysis of the genome sequence and transcriptome of the SwYG strain indicated that the targeted genetic modifications did not have measurable direct or indirect deleterious effects on yeast physiology, the 18 transformation rounds necessary to engineer IMX382 into the SwYG strain could have caused unwanted mutations in the yeast genome. Accordingly, the sequence of the SwYG strain revealed six mutations in coding regions of the native genome (Supplementary Table S1), three of which led to amino acid changes in Opt1 (plasma membrane oligopeptide transporter), Cwc25 (essential protein, component of spliceosomal complex implicated in the catalytic step of pre-mRNA splicing) and more interestingly Vps15 a serine/threonine protein kinase, key component in the synthesis of phosphatidylinositol 3-phosphate (PI(3)P) required for autophagosome formation (21-25). Bringing together the present results, a scenario describing the potential mechanisms involved in the SwYG strain slow growth phenotype emerges. During the construction of the SwYG strain from the double glycolysis strain, a mutation in *VPS15* occurred and negatively affected growth rate. The deletion of *VPS15* as well as the other main component of the PI3-kinase complex, *VPS34*, has been reported to lead to growth defects and cause aberrant endosome morphology (26,27). Further deletion of glycolytic genes in this impaired, *VPS15*-mutated intermediate strain would then fail to restore the growth rate to the level of the MG strain. During ALE, mutations in *ATG41*, an important protein for autophagosome formation (15,17), would counterbalance the deleterious effect of the mutated *VPS15*. The role of *SUR2* mutations in restoring the growth rate of the SwYG strain might appear more puzzling, however, evidence for the role of ceramides, important membrane lipids, in autophagy is growing (12-15,28). An additional, particularly attractive scenario in the context of engineering glycolysis, involves Pgc1. In mammals the Pgc1 homologue, also called Pgc1, plays a key role in the formation and activation of the complex responsible for PI(3)P synthesis and autophagosome formation (29). The PI3-kinase complex is largely conserved between mammalian cells and yeast and the human Pgc1 interaction partner, Beclin1, complements the function of its yeast homolog Vps30 (30). The potential involvement

of the yeast Pgc1 in PI(3)P synthesis and autophagy has not been explored yet, but it is not unlikely considering the conservation of moonlighting functions of several glycolytic proteins between yeast and mammals (31-33). Remarkably, the activity of Pgc1 is two-fold lower in SwYG strains as compared to the minimal glycolysis control strain (Figure 4). While this lower activity of an enzyme operating far from saturation is most likely not hindering the glycolytic flux, it might be deleterious for a potential moonlighting function. If such a moonlighting exists, deletion of the second copy of Pgc1 in the double glycolysis strain (strain IMX557, Figure 2) and ensuing lower Pgc1 abundance might have a negative effect on autophagosome formation and therefore growth. Further research will be required to test these hypotheses, starting by the repair of *VPS15* in the SwYG strain, and the integration of a second copy of Pgc1.

While *S. cerevisiae* is a fantastic microbial cell factory and model organism, with a remarkable genetic accessibility, the present study illustrates the challenges faced by intensive strain construction. It also illustrates the limits of analytical tools and highlights the need to combine multiple techniques to identify the molecular basis of mild phenotypic responses. Still, this study enabled a first, in-depth investigation of the impact of integration of large DNA constructs in *S. cerevisiae* native chromosomes, thereby identifying design guidelines for pathway transplantation, but also pathway design. The *SGA1* and *CAN1* loci proved to be robust integration sites for constructs of 30-40 kb, and can most likely harbor longer constructs. Addition of ARS was not required, although the spacing between adjacent ARS was increased from 38 kb to 73 kb, an observation in line with previous reports (34). Transcriptional hotspots, different from the native genetic configuration with a mosaic of genes with different transcription levels, did not harm yeast physiology, and the orientation of transcription did not require careful design. As in most organisms, in *S. cerevisiae* transformation is mutagenic, it is therefore recommended to keep the number of transformations to a minimum, an achievement made possible using CRISPR-Cas mediated multiplex editing. Finally, efforts to relocalize genes for genome modularization should consider watermarking genes, such that only the desired (native) copy can be exclusively targeted and edited (35). This approach minimizes the risk of excessive removal of DNA sequences during gene deletion. Applicable to native and synthetic chromosomes, these design guidelines should facilitate future large scale strain engineering endeavors.

## **Materials and Methods**

### **Media, strain cultivation and stocking**

The yeast strains used in this study are all derived from the CEN.PK lineage (Supplementary Table S2) (36). For non-selective growth, *S. cerevisiae* strains were

cultivated on Yeast extract Peptone Dextrose (YPD) medium, consisting of 10 g L<sup>-1</sup> Bacto yeast extract, 20 g L<sup>-1</sup> Bacto peptone and 20 g L<sup>-1</sup> glucose. For selective growth, Synthetic Medium (SM) was used, containing: 3 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 5 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mL L<sup>-1</sup> of a trace element solution (37). YPD and SM medium were set to pH 6 by addition of KOH and 20 g L<sup>-1</sup> Bacto agar was added in case of solid medium. YPD and SM media were autoclaved for 20 min at 110°C and 121°C, respectively. Finally, for Synthetic Medium with glucose (SMD), 1 mL L<sup>-1</sup> of a filter sterilized vitamin solution and 20 g L<sup>-1</sup> of glucose separately autoclaved for 20 min at 110°C was added. When needed, SMD was supplemented with 125 mg L<sup>-1</sup> histidine and/or 150 mg L<sup>-1</sup> uracil. When the dominant markers *amdS*, *KanMX* or *hphNT1* were used with SM, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was omitted from the medium and replaced by 6.6 g L<sup>-1</sup> K<sub>2</sub>SO<sub>4</sub>. For *amdS* selection, 1.8 g L<sup>-1</sup> filter sterilized acetamide was used as nitrogen source. For selection of *KanMX* or *hphNT1*, 2.3 g L<sup>-1</sup> urea was used as nitrogen source instead and 200 mg L<sup>-1</sup> G418 and 200 mg L<sup>-1</sup> hygromycin (Hyg) were added to the medium, respectively. For YPD medium, the G418 and hygromycin were directly added to the medium. Yeast cultures were grown at 30°C at 200 rpm in 50-/100-/500-mL shake flasks containing respectively 10-/20-/100 mL of medium, in an Innova 44 incubator shaker (New Brunswick Scientific, Edison, NJ). Yeast on solid media were incubated until single colonies were visible (approximately 3 days). *Escherichia coli* XL1-Blue strains with amp<sup>R</sup> containing plasmids were cultivated on Lysogeny Broth (LB) medium, consisting of: 10 g L<sup>-1</sup> tryptone, 5.0 g L<sup>-1</sup> yeast extract, and 5 g L<sup>-1</sup> NaCl, supplemented with 100 mg mL<sup>-1</sup> ampicillin. *E. coli* cultivations were performed at 37°C and 200 rpm in an Innova 4000 incubator (New Brunswick Scientific) shaker, in 15 mL Greiner tubes containing 5 mL medium. Cultures on solid LB medium, containing 20 g L<sup>-1</sup> Bacto agar, were incubated overnight at 37°C.

*S. cerevisiae* and *E. coli* strains were stored at -80°C in 1 mL aliquots of appropriate medium containing 30 % (v/v) glycerol.

### **Molecular biology techniques**

DNA used for strain construction purposes was amplified by PCR using Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA), according to the manufacturer's instructions. All diagnostic PCRs were performed with DreamTaq PCR Master Mix (Thermo Fisher Scientific), following the supplier's instruction. Primers were ordered desalted unless used for amplification of open reading frames (ORFs) in which case PAGE purified primers were used (Sigma-Aldrich, St. Louis, MO).

DNA was purified using the GenElute PCR Clean-Up kit (Sigma-Aldrich) or the GeneJET PCR Purification Kit (Thermo Fisher Scientific) when no unspecific bands were present. In case of unspecific bands, the DNA was separated on a 1 % or 2 % (w/v) agarose (TopVision Agarose, Thermo Fisher Scientific) gel in 1x Tris-acetate-EDTA

(TAE) buffer (Thermo Fisher Scientific) or 1x Tris-Borate-EDTA (TBE) (Thermo Fisher Scientific) buffer. 10  $\mu\text{L}$  L<sup>-1</sup> SERVA (SERVA Electrophoresis GmbH, Heidelberg, Germany) was added to the gel for DNA staining. As size standards, the GeneRuler DNA Ladder mix (Sigma-Aldrich) or GeneRuler DNA Ladder 50bp (Sigma-Aldrich) were used. DNA was isolated from gel using the GenElute PCR Clean-Up kit (Sigma-Aldrich). The purity and quantity of DNA was assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). For more precise DNA quantification the Qubit dsDNA BR Assay kit (Thermo Fisher Scientific) in combination with the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) was used. Digestion of DNA with DpnI was performed with the FastDigest restriction enzyme (Thermo Fisher Scientific), following the manufacturer's protocol. Gibson assembly for construction of gRNA plasmids was performed with Gibson assembly master mix 2x (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions but scaled down to a final volume of 5  $\mu\text{L}$ .

DNA for plasmid assembly was transformed into *E.coli* XL1-Blue by chemical transformation as described by Inoue *et al.* (38). Plasmids were isolated using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MO) or the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) and verified by diagnostic PCR or restriction analysis. *S. cerevisiae* was transformed with the lithium acetate/polyethylene glycol method (39). Single colony isolates were obtained by three consecutive re-streaks on selective solid medium. Yeast DNA was isolated by either boiling in 0.02 N NaOH, the protocol described by Looke *et al.* (40), the YeaStar genomic DNA kit (Zymo Research, Irvine, CA) or the QIAGEN Blood & Cell Culture Kit with 100/G Genomic-tips (Qiagen, Hilden, Germany) depending on the desired DNA purity.

### **Adaptive laboratory evolution**

CEN.PK113-7D and IMX1821 were evolved for faster growth on aerobic SMD medium at 30°C and 150 rpm in 100 mL shake flasks with 20 mL medium. The strains were inoculated from a freezer stock and subsequently transferred every 24 hours, by a 1:100 dilution in fresh medium. For both strains, three separate evolution lines were maintained for 79 transfers, when all IMX1821 evolution cultures had reached the approximate growth rate of wild type CEN.PK113-7D. The first 23 transfers of the CEN.PK113-7D evolution line were reported previously (41). From each evolution line, single colonies were isolated. The single cell lines that showed a growth rate representative of the evolved population were stored at -80°C and whole-genome sequenced (Supplementary Table S2).

### **Plasmid Construction**

To target the Cas9 protein to a desired locus and induce a double strand break, gRNA plasmids were constructed according to Mans *et al.* (42) with minor alterations. To

summarize, plasmids containing two gRNAs were Gibson assembled from a backbone amplified with primer 6005 from a pROS plasmid, and two insert fragments amplified with primers 5974 and 5975 and a respective gRNA specific primer. In Supplementary Table S3A all constructed plasmids are listed. Supplementary Table S4A-B contains all primers to construct and verify the gRNA plasmids.

### Strain construction

All strains constructed in this study originate from IMX589, the uracil auxotrophic SwYG strain, constructed by Kuijpers *et al.* (5) (Supplementary Table S2). This strain is characterized by the deletion of the minor paralogs of glycolysis and the relocalization of the major paralogs of glycolysis to the *SGA1* locus on chromosome IX. Some of the strains used in this study to test hypotheses were previously constructed and described (5,7,35,43). All strains constructed by CRISPR-Cas9 have been made as described by Mans *et al.* (42) and all strains were verified by diagnostic PCR (primers in Supplementary Table S4).

To construct a strain in which the transcription of the glycolytic genes was aligned with replication an *in silico* design was made (Figure 3). The watermarked glycolytic genes and auxiliary parts were amplified from template DNA with SHR containing primers to allow for homologous recombination (Supplementary Tables S3B, S4C and S5). Strain IMX1338 was transformed with 1.5 µg of plasmid p426-SNR52p-gRNA.CAN1.Y-SUP4t (Supplementary Table S3) and 200 fmol of each fragment (13 glycolytic genes: *FBA1*, *TPI1*, *PGK1*, *ADH1*, *PYK1*, *TDH3*, *ENO2*, *HXK2*, *PGI1*, *PFK1*, *PFK2*, *GPM1*, *PDC1*, the *HIS3* marker, *ARS418* and *ARS1211*) to integrate an altered orientation SinLoG in the *CAN1* locus (Chr. V). The gRNA plasmid was removed by non-selective growth and the uracil auxotrophic strain was stocked as IMX2359. To construct IMX2381, the glycolytic cassette in the *sga1* locus on Chr. IX was deleted from IMX2359 using gRNA plasmid pUDR413 and repaired with a *KIURA3* transcriptional unit. The *KIURA3* fragment was amplified from the pMEL10 plasmid with primers containing *sga1* flanks (Supplementary Table S3 and S4E). The gRNA plasmid was removed before stocking.

For reverse engineering of the mutations in *ATG41*, *SUR2* and *CNB1* found in the IMX1821 evolution lines, the three genes were first deleted by insertion of a synthetic gRNA and PAM sequence. To this end, IMX1821 was separately transformed with pUDR748, pUDR749 or pUDR750, to delete *ATG41*, *CNB1* and *SUR2*, respectively. 120 bp of corresponding repair fragments made by annealing of complementary primers containing the synthetic gRNA and PAM sequence were also supplied (Supplementary Table S4F). gRNA plasmids were removed, and strains were stocked as IMX2542, IMX2543 and IMX2544 for the deletion of *SUR2*, *ATG41* and *CNB1*, respectively. The mutated *SUR2* gene was amplified from evolution strains IMS1022, IMS1026 and IMS1031 with primer pair 17559/17560 (Supplementary Table S4G). Strain IMX2542

was transformed with plasmid pUDR114 to cut in the synthetic gRNA and separate repair fragments, resulting after plasmid removal in strains IMX2605 (*SUR2<sup>E192stop</sup>*), IMX2606 (*SUR2<sup>M179K</sup>*) and IMX2607 (*SUR2<sup>G251R</sup>*). Similarly, the *ATG41* and *CNB1* genes were amplified from the evolution lines (primers in Supplementary Table S4G) and transformed together with plasmid pUDR114 in strain IMX2543 and IMX2544, respectively. These transformations resulted in strains IMX2589 (*ATG41<sup>W132C</sup>*), IMX2590 (*ATG41<sup>W132stop</sup>*) and IMX2594 (*CNB1<sup>L82F</sup>*). For double deletion and replacement with a synthetic gRNA and PAM of *ATG41* and *SUR2*, strain IMX1821 was transformed with gRNA plasmid pUDR751 and corresponding 120 bp repair fragments (Supplementary Table S4F). After plasmid removal the resulting strain was stocked as IMX2570. Transformation of this strain with pUDR114 to cut the synthetic gRNAs and the amplified repair fragment (primers in Supplementary Table S4G) of *ATG41* and *SUR2* from evolution strain IMS1022, resulted in strain IMX2591 (*ATG41<sup>W132C</sup> SUR2<sup>E192stop</sup>*). Finally the triple deletion strain IMX2567 was made by transforming IMX2544 (*CNB1::(synthetic gRNA)*) with gRNA plasmid pUDR751 targeting *ATG41* and *SUR2* and the corresponding repair fragments with synthetic gRNA and PAM (Supplementary Table S4F).

## **Growth rate determination**

### Shake flasks

The growth of strains was monitored from 100 mL SMD cultures in 500 mL shake flasks by following the optical density at 660 nm ( $OD_{660}$ ) with a JENWAY 7200 spectrophotometer (Cole-Parmer, Stone, UK). Maximum specific growth rates ( $\mu_{max}$ ) were determined in at least biological duplicates from a minimal of six time points in exponential phase. To ensure that the cultures were in exponentially phase for growth rate measurements, two successive pre-cultures were performed prior to inoculation in the measuring flasks.

### 96-well format

Pre-cultures inoculated from glycerol-stocks were performed in 1.5 mL of YPD medium in 12-wells plate, and grown at 30°C and 800 rpm in a thermoshaker (Grant-bio PHMP-4, United Kingdom) until stationary phase. Of each strain 20  $\mu$ L were transferred to fresh SMD (with uracil) medium in a new 12-well plate and grown until mid-exponential phase. Each strain was then transferred with six biological replicates to a 96-well plate (CELLSTAR, Greiner Bio-One). Both the 12-well and 96-well plates were covered with sterile polyester acrylate sealing tape (Thermo Scientific). The  $OD_{660}$  from the cultures in the 96-well plates were determined every 20 min using a plate reader (TECAN infinite M200 Pro. Tecan, Männedorf, Switzerland). The growth rate was determined from at least 6 points in exponential phase.

### Ploidy staining and cell cycle phase estimation

Circa  $10^7$  cells were sampled from mid-exponential shake-flasks cultures on YPD and centrifuged (5 min at 4700 g). The pellet was washed with demineralized water, centrifuged again and resuspended in 800  $\mu$ l 70% ethanol while vortexing. Another 800  $\mu$ l 70% ethanol was added and fixed cells were stored at 4 °C until analysis. Staining of cells with SYTOX<sup>®</sup> Green Nucleic Acid Stain (Invitrogen S7020) was performed as described earlier (44). Samples were analyzed on a BD Accuri C6 flow cytometer with a 488 nm laser (BD Biosciences, Breda, The Netherlands). Fluorescence intensity was represented using FlowJo (V. 10.6.1, FlowJo, LLC, Ashland, OR, USA).

For cell cycle analysis, stained cells were selected from all events based on fluorescence at 533/30 nm. Subsequently the G<sub>1</sub> and G<sub>2</sub> populations corresponding to 1n and 2n were selected with ellipsoid gates based on forward scatter and fluorescence. All cells not corresponding to G<sub>1</sub> or G<sub>2</sub> were designated to be in the S phase.

### **Sequencing**

#### Whole genome sequencing and alignment

High quality genomic DNA was isolated with the QIAGEN Blood & Cell Culture Kit with 100/G Genomic-tips (Qiagen) from IMX1821 and evolution strains IMS1022, IMS1026, IMS1031, IMS1043, IMS1048 and IMS1053. The strains were sequenced in-house using an Illumina MiSeq Sequencer (Illumina, San Diego, CA) as described previously (7,35).

For long-read DNA sequencing of IMX589, genomic DNA was loaded on a R9 flowcell and sequenced on a MinION (Oxford Nanopore, sequencing kit SQK-LSK109). Guppy (Oxford Nanopore, version 3.9.5\_GPU) was used for basecalling. The resulting fastq files were filtered on length (>1 kb) and *de novo* assembled with Canu (version 1.8) (45). Resulting contigs were error-corrected by first mapping the Illumina paired-end read library of IMX589 with Burrows-Wheeler Algorithm (BWA mem, version 0.7.15) (46), converting the alignments to binary alignment format (BAM, samtools version 1.3.1) (47) and further processed with Pilon (48).

A *de novo* assembled reference genome was constructed for IMX589 (auxotrophic SwYG) using MinION and Miseq data (5). Using the Burrows-Wheeler Alignment (BWA) Tool (49) (version 0.7.15), sequencing data of IMX1821, IMS1022, IMS1026 and IMS1031 was aligned to the IMX589 reference genome and sequencing data of IMS1043, IMS1048 and IMS1053 was aligned to a CEN.PK113-7D reference (50). The data was further processed using SAMTools (49) (version 1.3.1) and single nucleotide polymorphisms (SNPs) were determined using Pilon (with -vcf setting; version 1.18) (48). The BWA.bam output file was visualized using the Integrative Genomics Viewer (version 2.4.0) (51), and copy numbers were estimated using Magnolya (version 0.15)

(52). SNPs were compared between CENPK113-7D and the three CEN.PK113-7D evolution lines (IMS1043, IMS1048 and IMS1053) and IMX1821 and the three SwYG evolution lines (IMS1022, IMS1026 and IMS1031).

#### Sanger sequencing

Sanger sequencing (Baseclear, Leiden, The Netherlands) was performed to check the reverse engineering of SNPs in *ATG41*, *SUR2*, *CNB1*, *VPS15*, *OPT1* and *CWC25*.

#### RNA sequencing and analysis

RNA sequencing results from the MG strain (IMX372) grown in aerobic SMD bioreactor cultures were obtained from Solis-Escalante *et al.* (6). For the SwYG strains with the glycolytic cassette in Chr. IX (IMX606) and in Chr. V (IMX605), RNA was sampled from the biological duplicate aerobic SMD bioreactor cultures run by Kuijpers *et al.* (5). The sampling and processing of the transcriptome samples from these bioreactors was executed as described by Solis-Escalante *et al.* (6). Library preparation and RNA sequencing were performed by Novogen. Sequencing of samples IMX605 and IMX606 was done with Illumina paired end 150 bp sequencing read system (PE150) using a 250 bp insert strand specific library which was prepared by Novogen.

Sequencing reads of samples IMX606, IMX605 and IMX372 were aligned to the CEN.PK113-7D genome using STAR (53). Expression was quantified by featureCounts (version 1.6.0) (54). Normalized FPKM counts were obtained by applying the rpkM function from the edgeR package (55). Differential expression analysis was done by applying DESeq2 (56).

## **Data availability**

All genomic data for this chapter have been deposited in the NCBI database (<https://www.ncbi.nlm.nih.gov/tudelft.idm.oclc.org/>) under the BioProjectID PRJNA757374.

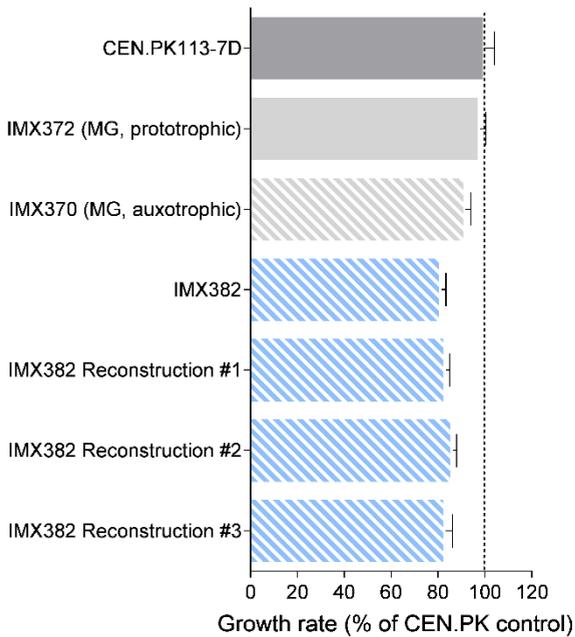
## **Acknowledgments**

We want to thank M.A.H. Luttik for cell cycle and ploidy measurements.

## Supplementary data

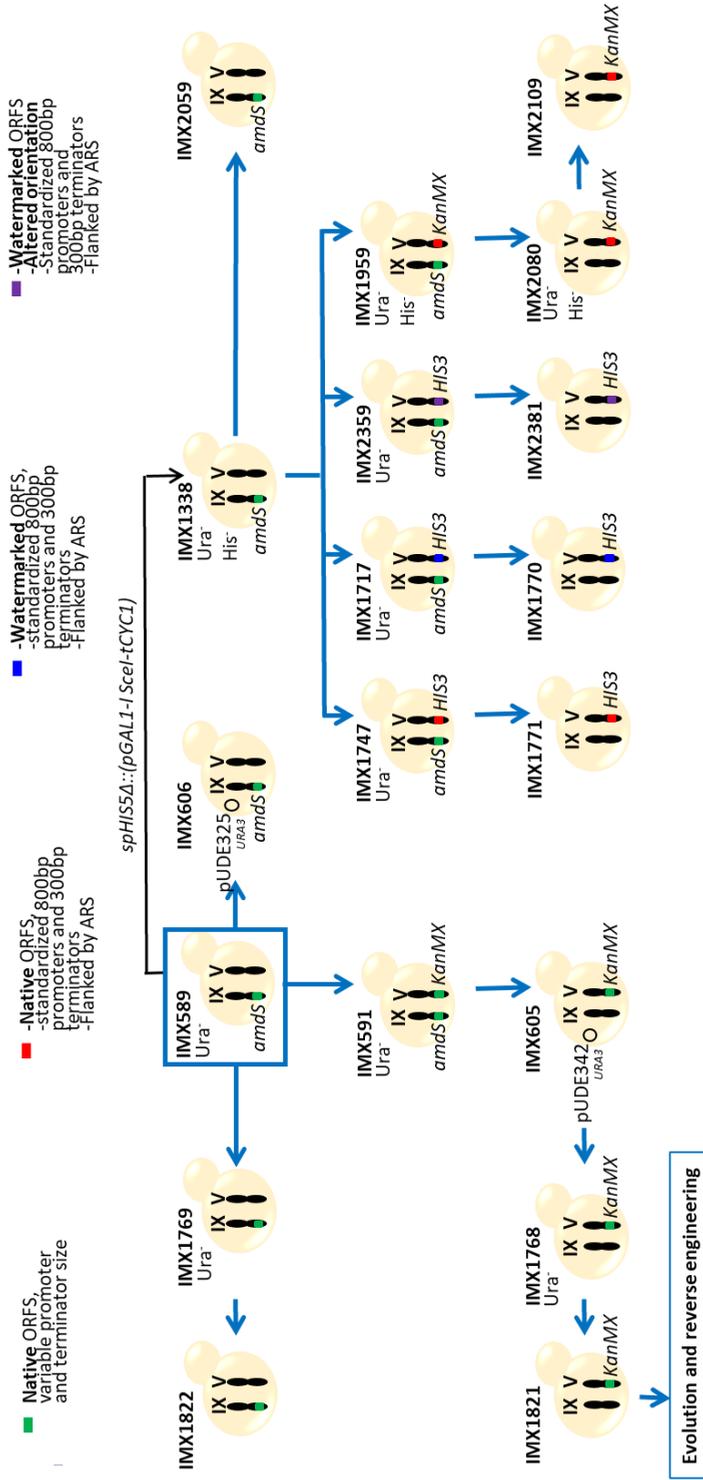
### Figure S1: Specific growth rates of yeast strains with Minimal Glycolysis background after integration of the single locus glycolysis

Reconstruction of strain IMX382 (SinLoG integration in *sga1* in IMX370) led to a similar growth rate defect. All growth rates are normalized to the CEN.PK113-7D control strain from the same experiment, prototrophic strains are shown as filled bars, uracil auxotrophic strains with dashed fill. Statistically significant differences relative to the auxotrophic MG strain reference strain are indicated (\*  $P < 0.05$ , t-test 2-tailed, homoscedastic).



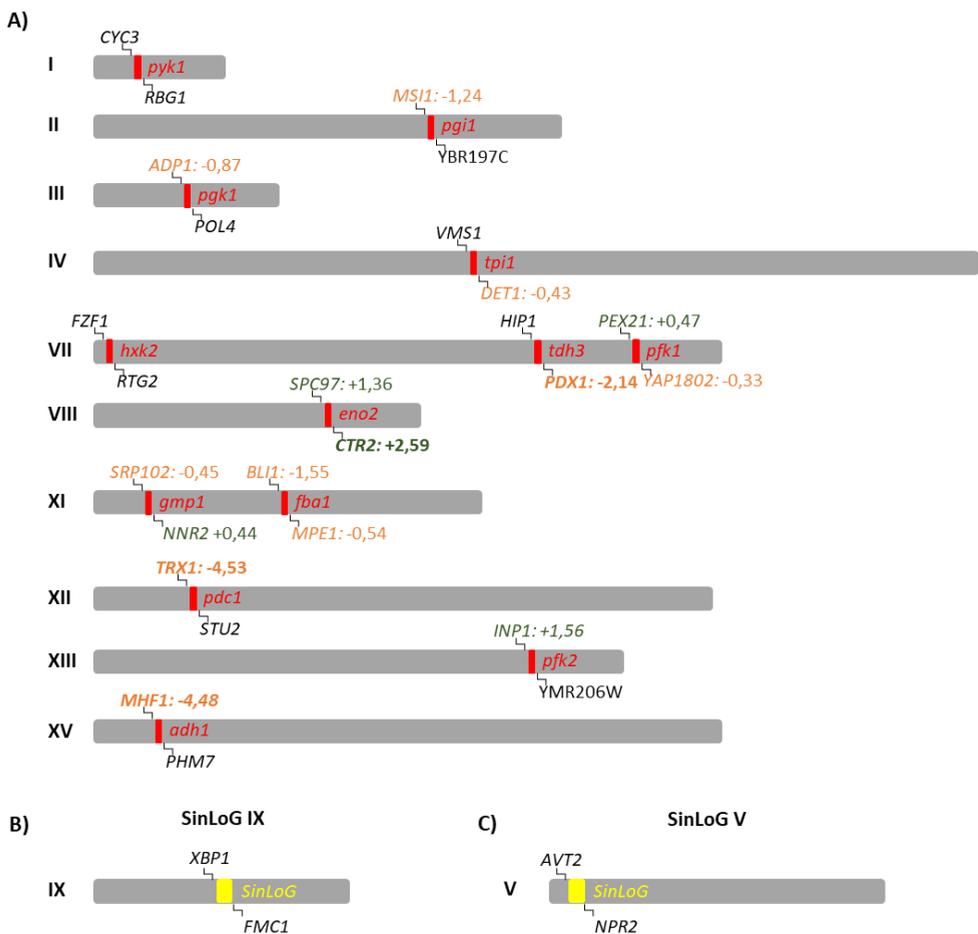
### Figure S2: Strain lineage of the single locus glycolysis strains

The lineage of strain construction is indicated for the main single locus glycolysis strains used in this study. Various designs of SinLoG with changes in promoters and terminators, watermarking, orientation of genes, and presence of ARS sequences are indicated by color.



**Figure S3: Relative transcription of genes neighboring the deletion loci of native glycolytic genes**

**A)** Relative expression in strain IMX606 (SinLoG-Chr IX) relative to the MG strain IMX372 of the genes neighboring the deleted glycolytic loci is indicated. Genes which are decreased in expression are marked in orange, genes which are increased in expression are marked in green and genes which are not expressed or did not change in expression are marked black (Mean PFKM\_MG>10 or Mean PFKM IMX606>10, padj<0.05). Genes with log<sub>2</sub>FoldChange>2 are marked in bold. **B,C)** Expression of genes neighboring the SinLoG loci on Chr. IX (IMX606) and Chr. V (IMX605). **D)** Function and PFKM count of IMX372 (MG) of genes that are changed in expression. Genes were considered not expressed if PFKM<10 in all strains (7); high expression was considered for PFKM>300.

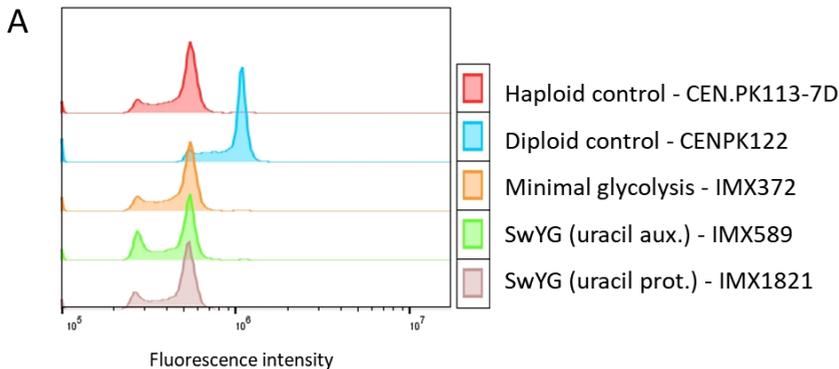


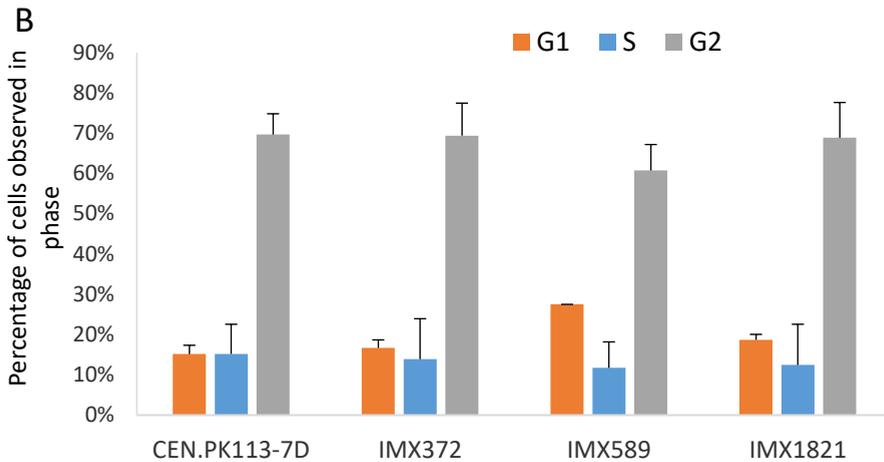
D)

Gene	Function	PFKM normalized counts
<i>MSI1</i>	Subunit of chromatin assembly factor I (CAF-1); chromatin assembly by CAF-1 affects multiple processes	53
<i>ADP1</i>	Putative ATP-dependent permease of the ABC transporter family	32
<i>DET1</i>	Acid phosphatase; involved in the non-vesicular transport of sterols in both directions between the endoplasmic reticulum and plasma membrane; deletion confers sensitivity to nickel	89
<i>PDX1</i>	E3-binding protein of the mitochondrial pyruvate dehydrogenase complex	105
<i>PEX21</i>	Peroxin required for peroxisomal matrix protein targeting	125
<i>YAP1802</i>	Protein of the AP180 family, involved in clathrin cage assembly	53
<i>SPC97</i>	Component of the microtubule-nucleating Tub4p (gamma-tubulin) complex	18
<i>CTR2</i>	Low-affinity copper transporter of the vacuolar membrane	148
<i>SRP102</i>	Signal recognition particle (SRP) receptor beta subunit; involved in SRP-dependent protein targeting	125
<i>NNR2</i>	Widely-conserved NADHX dehydratase; converts (S)-NADHX to NADH in ATP-dependent manner	45
<i>BLI1</i>	Subunit of the BLOC-1 complex involved in endosomal maturation	61
<i>MPE1</i>	Subunit of CPF cleavage and polyadenylation factor and E3 Ub-ligase; involved in 3' end formation of mRNA via cleavage and polyadenylation of pre-mRNA	51
<i>TRX1</i>	Cytoplasmic thioredoxin isoenzyme; part of the thioredoxin system that protects cells from oxidative and reductive stress	1580
<i>INP1</i>	Peripheral membrane protein of peroxisomes; involved in peroxisomal inheritance	16
<i>MHF1</i>	Component of the MFH histone-fold complex	115

**Figure S4: Ploidy and cell cycle distribution estimation by flow cytometric analysis**

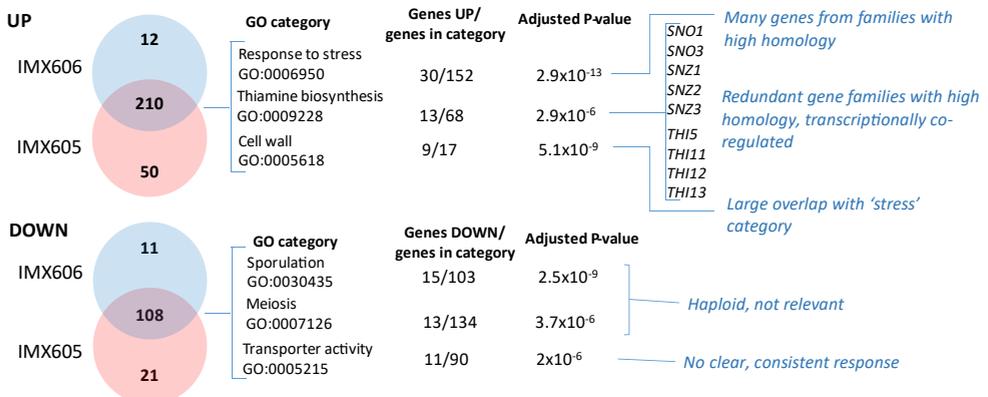
**A)** Ploidy measurement based on DNA staining and flow cytometry. All analyzed single locus glycolysis strains match the haploid reference. **B)** Cell cycle distribution based on the DNA content shown in panel **A**. The estimated percentage of cells in each cell cycle phase is shown. Error bars represent standard deviation between duplicate samples. The single locus glycolysis strain IMX589 (Chr IX) shows a slight increase in cells in G1 compared to the CEN.PK113-7D and IMX372 (minimal glycolysis) control strains.





**Figure S5: Global transcription effects from the single locus glycolysis**

Global transcriptional changes in strains IMX605 (SinLoG on Chr. V) and IMX606 (SinLoG on Chr. IX) compared to the minimal glycolysis strain IMX372. Since both strains carrying a SinLoG share a similar growth rate enrichment analysis was performed on significantly up and down regulated genes in both strains to identify possible causal factors for this growth defect. Up- and down regulated genes with an adjusted p-value below 0.05 and a fold change greater than 2 were selected and doublons were removed. Enrichment analysis was performed with FunSpec (<http://funspec.med.utoronto.ca/cgi-bin/funspec>), GO terms with a Bonferroni-corrected p-value <  $10^{-5}$  are reported. Although several categories are significantly enriched, no link to processes related to the integration of the SinLoG were identified.



**Selection criteria for genes:**

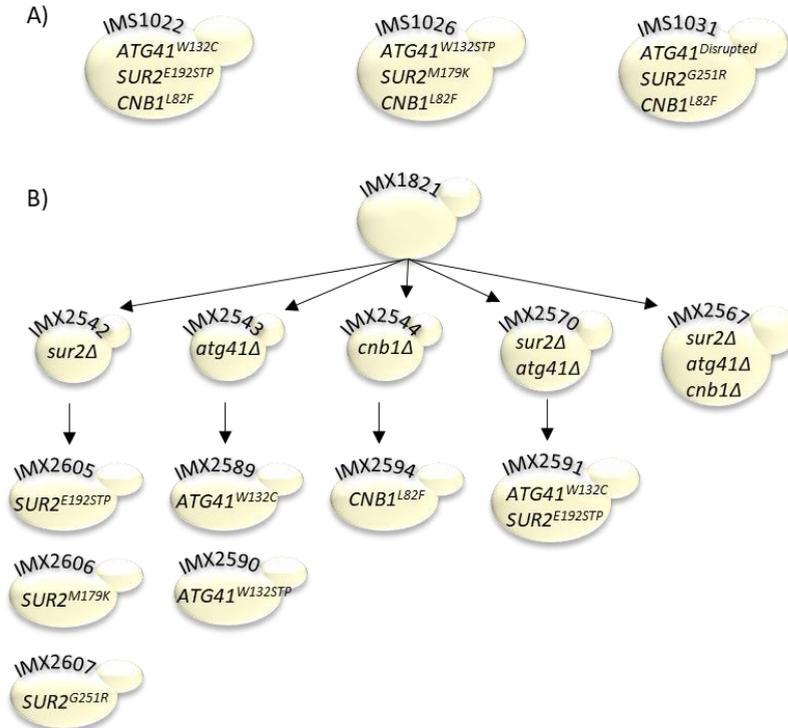
- Adjusted p-value < 0.05
- $2 < FC < 0.5$
- (removed doublons)

**Selection criteria for categories:**

- Enrichment analysis performed with FunSpec (<http://funspec.med.utoronto.ca/cgi-bin/funspec>)
- Bonferroni-corrected p-value <  $10^{-5}$

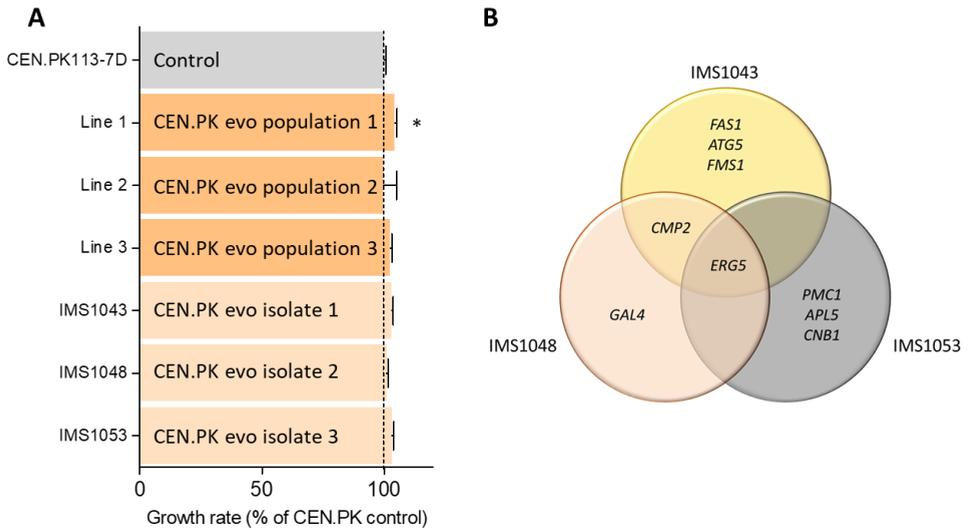
**Figure S6: Strain lineage of evolved isolates of the SwYG strain and reverse engineered strains**

**A)** Mutations in *ATG41*, *SUR2* and *CNB1* in evolved strains IMS1022, IMS1026 and IMS1031.  
**B)** Reverse engineering strategy of the *SUR2*, *ATG41* and *CNB1* mutations in the SwYG strain IMX1821.



**Figure S7: Physiological and genetic characterization of populations and single colony isolates after evolution of the CEN.PK113-7D reference strain.**

**A)** Specific growth rate CEN.PK113-7D and derived strains after evolution, represented as evolved population and single colony isolates. Growth rates are shown as percentage of the unevolved reference strain. No significant increase in growth rate was found except a slight increase for the population of line 1. Bars represent the average growth rate and standard deviation of duplicate shake flask cultures. Significant differences as compared to unevolved control are indicated (\*,  $P < 0.05$ , t-test, homoscedastic). **B)** Venn diagram showing genes in which mutations were identified in the evolved strains derived from CEN.PK113-7D.



**Table S1 Single nucleotide polymorphisms in the auxotrophic SwYG strain (IMX589) compared to MG strain (IMX372)**

Information from Kuijpers *et al.* (5)

Mutations (SNP)	Genome	SinloG
All	15	9
In ORF	6 <sup>#</sup>	1*
MisSense	3 ( <i>VPS15</i> , <i>OPT1</i> , <i>CWC25</i> )	0

### # Information on the 6 in ORF SNPs in the genome

Systematic name	Name	Type	Amino Acid change
YBR079W	<i>VPS15</i>	Non-synonymous	E474K
YJL212C	<i>OPT1</i>	Non-synonymous	I463T
YNL245C	<i>CWC25</i>	Non-synonymous	P62L
YDL079C	<i>MRK1</i>	Synonymous	I190I
YLR180W	<i>SAM1</i>	Synonymous	V217V
YNL262W	<i>POL2</i>	Synonymous	F1536F

### \* Information on the single SNP in the SinLoG

Systematic name	Name	Type	Amino Acid change
YOL086C	<i>ADH1</i>	Synonymous	A180A

### Table S2: List of strains used in this study

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16539840>

### Table S3 List of plasmids used in this study

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16539840>

### Table S4 List of primers used in this study

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16539840>

### Table S5 Configuration of SinLoG in IMX2359 and IMX2381

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16539840>

## References

1. Nielsen, J. and Keasling, J.D. (2016) Engineering cellular metabolism. *Cell*, **164**, 1185-1197.
2. Galanie, S., Thodey, K., Trenchard, I.J., Filsinger Interrante, M. and Smolke, C.D. (2015) Complete biosynthesis of opioids in yeast. *Science*, **349**, 1095-1100.
3. Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D. *et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496**, 528-532.
4. Kondo, A., Ishii, J., Hara, K.Y., Hasunuma, T. and Matsuda, F. (2013) Development of microbial cell factories for bio-refinery through synthetic bioengineering. *J. Biotechnol.*, **163**, 204-216.
5. Kuijpers, N.G., Solis-Escalante, D., Luttk, M.A., Bisschops, M.M., Boonekamp, F.J., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2016) Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 15060-15065.
6. Solis-Escalante, D., Kuijpers, N.G., Barrajon-Simancas, N., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2015) A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell*, **14**, 804-816.
7. Postma, E.D., Dashko, S., van Breemen, L., Taylor Parkins, S.K., van den Broek, M., Daran, J.-M. and Daran-Lapujade, P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **49**, 1769-1783.
8. Hamperl, S. and Cimprich, K.A. (2016) Conflict resolution in the genome: how transcription and replication make it work. *Cell*, **167**, 1455-1467.
9. Dhar, M.K., Sehgal, S. and Kaul, S. (2012) Structure, replication efficiency and fragility of yeast ARS elements. *Res. Microbiol.*, **163**, 243-253.
10. Tai, S.L., Daran-Lapujade, P., Luttk, M.A., Walsh, M.C., Diderich, J.A., Krijger, G.C., van Gulik, W.M., Pronk, J.T. and Daran, J.M. (2007) Control of the glycolytic flux in *Saccharomyces cerevisiae* grown at low temperature: a multi-level analysis in anaerobic chemostat cultures. *J. Biol. Chem.*, **282**, 10243-10251.
11. Gancedo, C. and Flores, C.L. (2008) Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.*, **72**, 197-210.
12. Haak, D., Gable, K., Beeler, T. and Dunn, T. (1997) Hydroxylation of *Saccharomyces cerevisiae* ceramides requires Sur2p and Scs7p. *J. Biol. Chem.*, **272**, 29704-29710.
13. Grilley, M.M., Stock, S.D., Dickson, R.C., Lester, R.L. and Takemoto, J.Y. (1998) Syringomycin action gene *SYR2* Is essential for sphingolipid 4-hydroxylation in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **273**, 11062-11068.
14. Megyeri, M., Prasad, R., Volpert, G., Sliwa-Gonzalez, A., Haribowo, A.G., Aguilera-Romero, A., Riezman, H., Barral, Y., Futerman, A.H. and Schuldiner, M. (2019) Yeast ceramide synthases, Lag1 and Lac1, have distinct substrate specificity. *J. Cell Sci.*, **132**.
15. Li, Y., Li, S., Qin, X., Hou, W., Dong, H., Yao, L. and Xiong, L. (2014) The pleiotropic roles of sphingolipid signaling in autophagy. *Cell Death Dis.*, **5**, e1245-e1245.
16. Desfarges, L., Durrrens, P., Juguelin, H., Cassagne, C., Bonneu, M. and Aigle, M. (1993) Yeast mutants affected in viability upon starvation have a modified phospholipid composition. *Yeast*, **9**, 267-277.
17. Yao, Z., Delorme-Axford, E., Backues, S.K. and Klionsky, D.J. (2015) Atg41/Icy2 regulates autophagosome formation. *Autophagy*, **11**, 2288-2299.
18. Metztl-Raz, E., Kafri, M., Yaakov, G., Soifer, I., Gurvich, Y. and Barkai, N. (2017) Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife*, **6**, e28034.

19. Eguchi, Y., Makanae, K., Hasunuma, T., Ishibashi, Y., Kito, K. and Moriya, H. (2018) Estimating the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins. *eLife*, **7**, e34595.
20. Kafri, M., Metzler-Raz, E., Jona, G. and Barkai, N. (2016) The cost of protein production. *Cell Rep.*, **14**, 22-31.
21. Chiu, Y.F., Liu, Y.C., Chiang, T.W., Yeh, T.C., Tseng, C.K., Wu, N.Y. and Cheng, S.C. (2009) Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction. *Mol. Cell. Biol.*, **29**, 5671-5678.
22. Tseng, C.K., Chung, C.S., Chen, H.C. and Cheng, S.C. (2017) A central role of Cwc25 in spliceosome dynamics during the catalytic phase of pre-mRNA splicing. *RNA*, **23**, 546-556.
23. Osawa, H., Stacey, G. and Gassmann, W. (2006) *ScOPT1* and *AtOPT4* function as proton-coupled oligopeptide transporters with broad but distinct substrate specificities. *Biochem J.*, **393**, 267-275.
24. Stack, J.H., DeWald, D.B., Takegawa, K. and Emr, S.D. (1995) Vesicle-mediated protein transport: regulatory interactions between the Vps15 protein kinase and the Vps34 PtdIns 3-kinase essential for protein sorting to the vacuole in yeast. *J. Cell Biol.*, **129**, 321-334.
25. Obara, K., Noda, T., Niimi, K. and Ohsumi, Y. (2008) Transport of phosphatidylinositol 3-phosphate into the vacuole via autophagic membranes in *Saccharomyces cerevisiae*. *Genes to cells : devoted to molecular & cellular mechanisms*, **13**, 537-547.
26. Grunau, S., Lay, D., Mindthoff, S., Platta, H.W., Girzalsky, W., Just, W.W. and Erdmann, R. (2011) The phosphoinositide 3-kinase Vps34p is required for pexophagy in *Saccharomyces cerevisiae*. *Biochem J.*, **434**, 161-170.
27. Laidlaw, K.M., Paine, K.M., Bisinski, D.D., Calder, G. and MacDonald, C. (2021) Endosomal recycling to the surface mediated by Gpa1 and PI3-Kinase is inhibited by glucose starvation. *bioRxiv*.
28. Yamagata, M., Obara, K. and Kihara, A. (2011) Sphingolipid synthesis is involved in autophagy in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, **410**, 786-791.
29. Qian, X., Li, X., Cai, Q., Zhang, C., Yu, Q., Jiang, Y., Lee, J.H., Hawke, D., Wang, Y., Xia, Y. *et al.* (2017) Phosphoglycerate kinase 1 phosphorylates beclin1 to induce autophagy. *Mol. Cell*, **65**, 917-931.e916.
30. Liang, X.H., Jackson, S., Seaman, M., Brown, K., Kempkes, B., Hibshoosh, H. and Levine, B. (1999) Induction of autophagy and inhibition of tumorigenesis by beclin 1. *Nature*, **402**, 672-676.
31. Lu, M., Ammar, D., Ives, H., Albrecht, F. and Gluck, S.L. (2007) Physical interaction between aldolase and vacuolar H<sup>+</sup>-ATPase is essential for the assembly and activity of the proton pump. *J. Biol. Chem.*, **282**, 24495-24503.
32. Lu, M., Sautin, Y.Y., Holliday, L.S. and Gluck, S.L. (2004) The glycolytic enzyme aldolase mediates assembly, expression, and activity of vacuolar H<sup>+</sup>-ATPase. *J. Biol. Chem.*, **279**, 8732-8739.
33. Mayordomo, I. and Sanz, P. (2001) Human pancreatic glucokinase (GlkB) complements the glucose signalling defect of *Saccharomyces cerevisiae* *hvk2* mutants. *Yeast*, **18**, 1309-1316.
34. Karas, B.J., Suzuki, Y. and Weyman, P.D. (2015) Strategies for cloning and manipulating natural and synthetic chromosomes. *Chromosome Res.*, **23**, 57-68.
35. Boonekamp, F.J., Dashko, S., Duiker, D., Gehrmann, T., van den Broek, M., den Ridder, M., Pabst, M., Robert, V., Abeel, T., Postma, E.D. *et al.* (2020) Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences. *ACS Synth. Biol.*, **9**, 1361-1375.
36. Entian, K.-D. and Kötter, P. (2007) 25 Yeast genetic strain and plasmid collections. In: Stansfield, I. and Stark, M. J. R. (eds.), *Methods in Microbiology*. Academic Press, Vol. **36**, pp. 629-666.

37. Verduyn, C., Postma, E., Scheffers, W.A. and Van Dijken, J.P. (1992) Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, **8**, 501-517.
38. Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23-28.
39. Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.*, **350**, 87-96.
40. Looke, M., Kristjuhan, K. and Kristjuhan, A. (2011) Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques*, **50**, 325-328.
41. Perli, T., Moonen, D.P.I., van den Broek, M., Pronk, J.T. and Daran, J.M. (2020) Adaptive laboratory evolution and reverse engineering of single-vitamin prototrophies in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **86**.
42. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J. and Daran, J.M. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **15**, 1-15.
43. Boonekamp, F.J. (2020) Modular engineering of synthetic glycolytic pathways in *Saccharomyces cerevisiae*, <https://doi.org/10.4233/uuid:4cd6d858-5f09-4567-86e8-f9f61ca7941f>.
44. Haase, S.B. and Reed, S.I. (2002) Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle*, **1**, 132-136.
45. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, **27**, 722-736.
46. Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint*, 1303.3997.
47. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
48. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K. *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS one*, **9**, e112963.
49. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.
50. Salazar, A.N., Gorter de Vries, A.R., van den Broek, M., Wijsman, M., de la Torre Cortes, P., Brickwedde, A., Brouwers, N., Daran, J.G. and Abeel, T. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.*, **17**.
51. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178-192.
52. Nijkamp, J.F., van den Broek, M.A., Geertman, J.M., Reinders, M.J., Daran, J.M. and de Ridder, D. (2012) *De novo* detection of copy number variation by co-assembly. *Bioinformatics*, **28**, 3195-3202.
53. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
54. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923-930.

55. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139-140.
56. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

# Chapter 4

## A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces* *cerevisiae*

Eline D. Postma

Sofia Dashko

Lars van Breemen

Shannara K. Taylor Parkins

Marcel van den Broek

Jean-Marc Daran

Pascale Daran-Lapujade

## Abstract

The construction of microbial cell factories for sustainable production of chemicals and pharmaceuticals requires extensive genome engineering. Using *Saccharomyces cerevisiae*, this study proposes synthetic neochromosomes as orthogonal expression platforms for rewiring native cellular processes and implementing new functionalities. Capitalizing the powerful homologous recombination capability of *S. cerevisiae*, modular neochromosomes of 50 and 100 kb were fully assembled *de novo* from up to 44 transcriptional-unit-sized fragments in a single transformation. These assemblies were remarkably efficient and faithful to their *in silico* design. Neochromosomes made of non-coding DNA were stably replicated and segregated irrespective of their size without affecting the physiology of their host. These non-coding neochromosomes were successfully used as landing pad and as exclusive expression platform for the essential glycolytic pathway. This work pushes the limit of DNA assembly in *S. cerevisiae* and paves the way for *de novo* designer chromosomes as modular genome engineering platforms in *S. cerevisiae*.

## Introduction

Microbial cell factories have an important role to play in the development of a sustainable and environmentally friendly biobased economy, as already exemplified by the ca. 100 billion litres of bioethanol (1) and half of the world's insulin (2) annually produced by *Saccharomyces cerevisiae*, more commonly known as baker's yeast. However, the construction of novel microbial production hosts for chemicals and pharmaceuticals is impeded by the poor economic viability of bio-based processes. Improving microbial process profitability and competition with current (petrochemical based) production methods requires powerful microbial cell factories that produce (novel) chemicals from non-native feedstock (3) with high product yields and productivity in harsh industrial conditions. This can only be accomplished by extensive genome engineering that not only adds new functionalities to the host microbe, but also deeply rewires its native cellular and metabolic processes (4,5).

Although the construction of synthetic cell factories with designer genomes, tailor-made for the optimum production of certain chemicals, might become feasible in the future, it is currently far from reach. One reason is our still limited understanding of the absolute requirements of life, as demonstrated by the recent reconstruction of a minimal *Mycoplasma* genome in which one third of the essential genes are not functionally characterized (6). The second major hurdle in the synthesis of designer genomes is DNA synthesis. The maximum size for the chemical synthesis of ssDNA of a quality that can be used for biotechnological application is 200 bp (7,8). Despite recent developments in stitching together these ssDNA parts into genes, routine synthesis of designer genomes would be too cumbersome and expensive (8,9). Indeed, although numerous fast (seamless) *in vitro* assembly methods (10) have been developed such as: Gibson assembly, Golden Gate, ligase cycling reaction, seamless ligation cloning extract and circular polymerase extension cloning, these methods are intrinsically limited in the number of fragments that can be assembled as well as the final size of the assembled DNA. Furthermore, *Escherichia coli*, used as a propagation host (11), cannot faithfully replicate genome-size DNA constructs (12). It is *S. cerevisiae* that ultimately enabled the complete assembly and replication of the 580 kb *Mycoplasma* genome (12). The homologous recombination machinery of *S. cerevisiae* was able to assemble 25 overlapping DNA fragments of ~24 kb into the *Mycoplasma* genome as well as assemble 38 overlapping ssDNA pieces of 200 bp (13,14). This high efficiency and fidelity of homologous recombination has undoubtedly contributed to *S. cerevisiae* popularity as eukaryotic model and cell factory. Yet, the full extent of *in vivo* assembly capabilities in terms of size and number of fragments has not been explored. Introduction of large genetic constructs in *S. cerevisiae* typically takes place by integration in existing chromosomes, without much considerations for potential

impact on the host chromosome architecture. The recent introduction of the noscapine pathway, for example, required introduction of 31 (heterologous) genes in nine different chromosomal loci (4). Integration of large constructs might alter the chromosome structure and affect the expression of genes surrounding the newly added construct, as well as replication of the host chromosome. It might also trigger unwanted chromosomal rearrangements between native chromosomes. It is long known that *S. cerevisiae* can stably replicate and segregate Yeast Artificial Chromosomes (YACs) (15). In the present study, we explore the potential of *in vivo* assembled, supernumerary chromosomes for modular genome engineering in *S. cerevisiae*.

To enable easy remodeling of existing functions, Kuijpers *et al.* (16) developed the pathway swapping concept in *S. cerevisiae*. The genetic reduction and relocalization of the entire Embden–Meyerhoff–Parnas pathway of glycolysis to a single chromosomal locus enabled to swap this essential pathway to any new (heterologous) design in two simple steps (16,17). Combined to an orthogonal expression platform, in the form of a supernumerary synthetic neochromosome, pathway swapping would offer the possibility to make the introduction and expression of large product pathways in combination with the rewiring of essential native pathways, more efficient and predictable. Unlike the construction of chromosomes based on an existing structure for the *Mycoplasma* and *S. cerevisiae* genome projects (Sc2.0) (18,19), the neochromosomes assembly design should be based on modularity to allow for easy change in makeup and configuration of native and heterologous pathways.

The present study explores the possibility for a modular assembly of a designer, supernumerary neochromosome for rational engineering of the yeast genome. A pipeline was developed for easy and rapid design, *in vivo* assembly and verification of neochromosomes. The limits of *in vivo* assembly in terms of number and size of assembled DNA fragments, efficiency and fidelity were explored. The stability of the neochromosomes and their impact on yeast physiology were probed. Finally, the ability of supernumerary synthetic chromosomes to serve as landing pad for metabolic pathways and to carry essential pathways was investigated. This study paves the way for the implementation of supernumerary, synthetic chromosomes as modular expression platforms for native and heterologous functions.

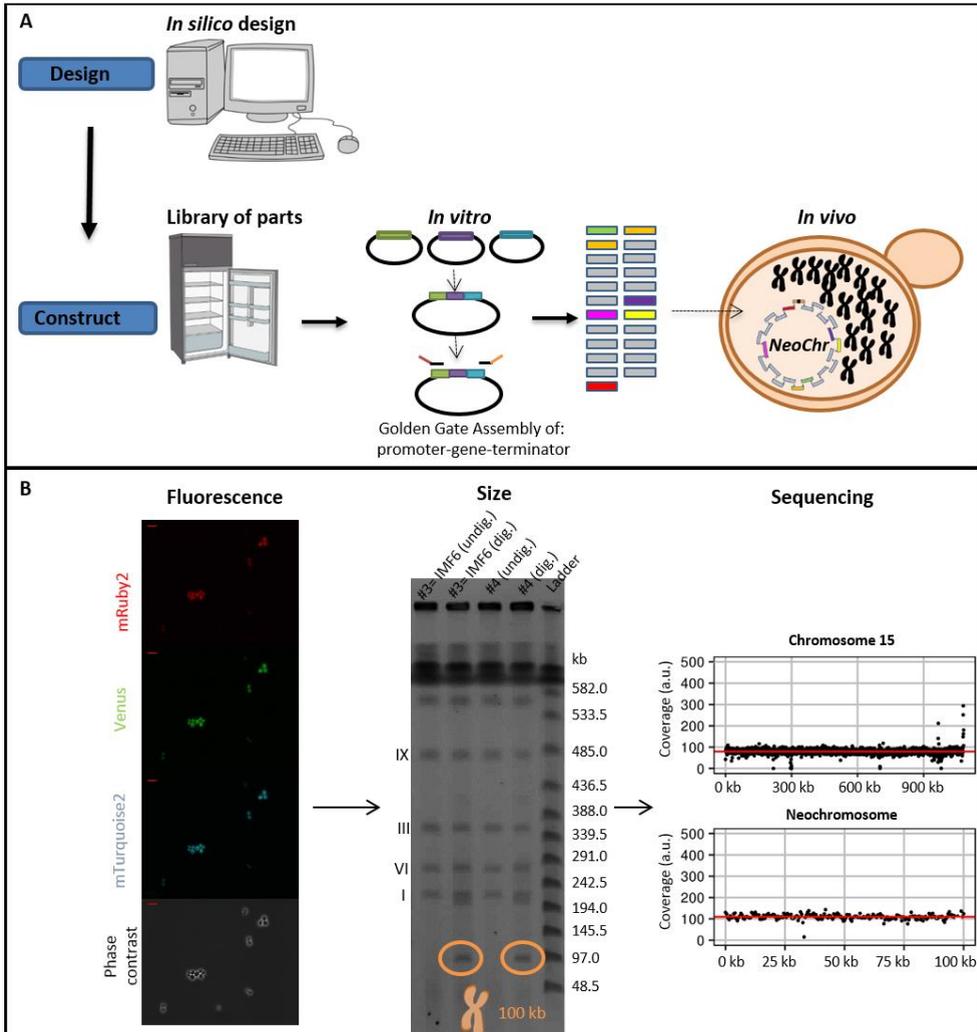
## Results

### Design considerations and proof of concept for the modular, *de novo* assembly of supernumerary synthetic neochromosomes

Establishment of a workflow for fast and flexible construction of neochromosomes should include a set of rules to enable easy design, assembly and screening of

neochromosomes with different configurations of genes and auxiliary parts. This flexibility, required for pathway optimization in future strain construction programs, gave the possibility to explore structural requirements for the assembly and maintenance of the neochromosomes in the present study. Indeed, in this nascent field, very little is known about neochromosome design requirements and even less is known about neochromosome stability and impact on physiology. The second most important construction criterion was the minimization of experimental steps to construct the neochromosomes. Accordingly, the chosen construction workflow exploits the strengths of *in vitro* and *in vivo* assembly, by first using Golden gate cloning for stitching functional DNA parts together (*e.g.* construction of transcription units from promoter, gene and terminator), then assembling *de novo* these *in vitro* stitched parts using *S. cerevisiae* highly efficient homologous recombination (Figure 1A). This workflow also minimized the number of PCR amplification steps, that are notoriously error-prone, to a single one.

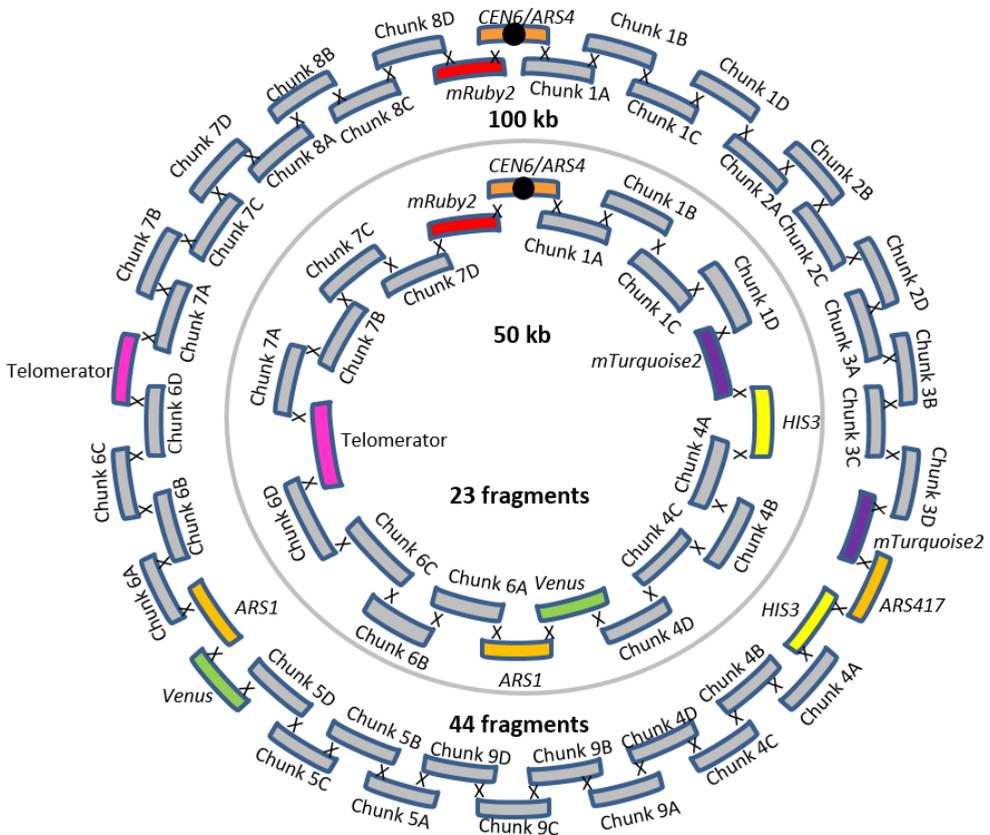
*In vivo* neochromosome assembly was promoted by framing the DNA fragments transformed to yeast with short overhangs with no homology with the yeast genome (60 bp long called SHRs (20)), that can be easily and cheaply added by PCR. The neochromosomes were designed to carry replicative fragments, including a centromere and autonomously replicating sequences (ARS) spaced every 30–40 kb (21), and markers to facilitate the selection of transformants with correctly assembled neochromosomes. To evaluate the efficiency of our neochromosome construction workflow, two test chromosomes of 50 and 100 kb were designed and assembled. To mimic assembly of pathways, typically composed of transcription units of ca. 2–3 kb, but to avoid potential interference by gene expression from the neochromosomes, the test neochromosomes were assembled from 2.5 kb *E. coli* DNA fragments. Although these DNA fragments contain *E. coli* genes, they are not expected to be expressed in *S. cerevisiae* due to fundamental differences in prokaryotic and eukaryotic transcription machineries (22,23). Specifically, the test neochromosomes comprised: *CEN6/ARS4* (24), *ARS1* (and also *ARS417* for 100 kb neochromosome), two auxotrophic selection markers (*HIS3* and *URA3*), three fluorescent proteins (Venus, mRuby2 and mTurquoise2), and 2.5 kb non-coding *E. coli* DNA fragments used to reach the desired chromosome size (called filler fragments) and the telomerator (25). The telomerator allows for *in vivo* chromosome linearization and was included for potential future use. This resulted in 23 fragments for the assembly of the 50 kb neochromosome and 44 fragments for the 100 kb neochromosome (Figure 2), which were transformed in fixed molar amounts per fragment (therefore approximately twice as much DNA was transformed for the 100 kb neochromosome with respect to the 50 kb neochromosome). Despite the large number of DNA fragments, transformation on selective growth medium (histidine and uracil deficient) resulted in



**Figure 1 - Modular genome engineering strategy.** Schematic representation of the design and construction of neochromosomes (A), as well as screening based on fluorescence, size and sequencing (B). The data in panel B represent the results pertaining to the assembly of the 100 kb neochromosome in strain IMF6. The red bar in the top left of the fluorescence microscopy pictures (40x) represents 10  $\mu$ m. The CHEF gel used to determine the size of the neochromosomes, contains two transformants (#3 and #4) of the 100 kb neochromosome transformation. The neochromosomes are either undigested (undig.) by I-SceI and therefore in their circular form, or in-plug linearized by I-SceI digestion (dig.) and therefore in linear form, which can be visualised on CHEF gel.

a large number of colonies (ca. 2000 colonies and 275 colonies, for the 50 and 100 kb neochromosome respectively). Four colonies of each transformation were checked by fluorescence microscopy and all expressed the three fluorescent markers (Figure 1B and Supplementary Figure S1). For both test chromosomes, one of these

four colonies was randomly selected and tested by diagnostic PCR of the assembly junctions (Supplementary Figure S2), which revealed the correct assembly of the 50 and 100 kb neochromosomes. The genome of two transformants harboring the 100 kb neochromosome (IMF1 and IMF6) and one transformant harboring the 50 kb neochromosome (IMF2) were sequenced. While IMF2 and IMF6 were faithfully assembled, harboring all fragments of the *in silico* design in the correct configuration, IMF1 was not. According to the sequencing data, IMF1 carried all expected fragments but was harboring an internal deletion in a filler fragment (6C) and a duplication of a 35 kb region (Supplementary Figure S3). This duplication, covering the middle of *mTurquoise2* to the middle of *Venus*, was confirmed by karyotyping using in-plug linearization of the neochromosome (Supplementary Figures S4 and S5) and most likely resulted from a recombination event between the *mTurquoise2* and



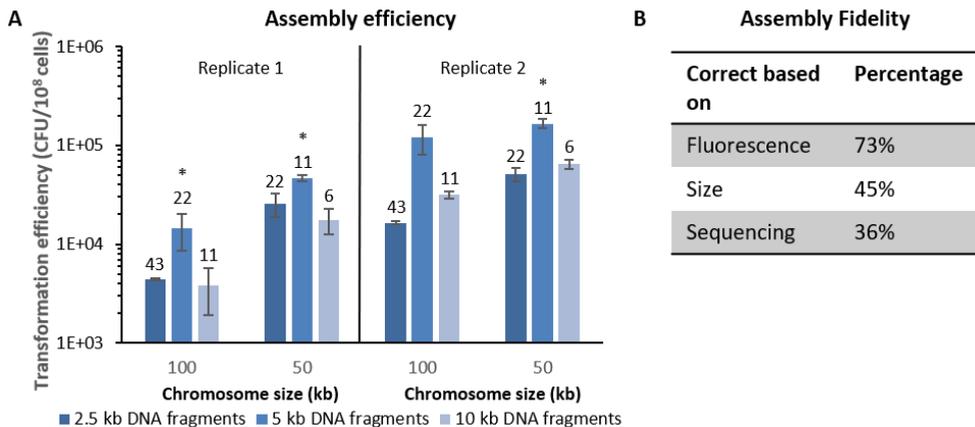
**Figure 2 - 50 and 100 kb neochromosome designs.** The outer circle depicts the *in vivo* assembly of the 100 kb neochromosome from 44 fragments present in strain IMF6. The inner circle depicts the *in vivo* assembly of the 50 kb neochromosome from 23 fragments present in strain IMF2.

the *Venus* genes, that have a nucleotide identity of 97% (26-28). Long-read sequencing of IMF2 and IMF6 further confirmed that in these strains no duplication or recombination events had occurred (Supplementary Figure S6). Neochromosomes can therefore be efficiently and faithfully assembled from as many as 44 fragments, despite the presence of an internal homologous region of approximately 700 bp, as well as regions homologous to the native chromosomes (*pTEF1*, *pTEF2*, *pCCW12*, *tSSA1*, *ARS*, *HIS3* and *URA3* expression cassette). The 60 bp SHR was therefore sufficient to promote the efficient assembly of the neochromosomes.

### **Exploring optimal fragment size and number for efficient, modular *in vivo* assembly of neochromosomes**

Although the assembly of the 50 and 100 kb neochromosomes resulted in a large number of colonies, doubling the number of transformed fragments caused a seven-fold decrease in colony count (ca. 2000 colonies and 275 colonies, respectively). This suggested that the number of fragments in the neochromosome design might become a limiting factor for the construction of larger neochromosomes harboring many native and non-native pathways. To gain more insight into transformation efficiency (*i.e.* CFU/10<sup>8</sup> transformed cells) and to test to what extent efficiency might be influenced by fragment size and/or the number of transformed fragments, new 50 and 100 kb neochromosomes were designed and assembled using DNA filler fragments of three different sizes: 2.5, 5 and 10 kb, transformed in equimolar amounts. The transformations resulted in a large number of colonies (10<sup>3</sup>–10<sup>4</sup> upon plating of the whole transformation mixture), and was most efficient with 5 kb fragments. On average transformation with 5 kb fragments was four-fold more efficient than 2.5 kb fragments and three-fold more efficient than with 10 kb fragments (one-tailed paired *t*-tests  $P \leq 0.05$ ), but fragments of 2.5 and 10 kb led to the same number of transformants (one-tailed paired *t*-test  $P > 0.05$ ) (Figure 3A). The number of fragments transformed to yeast can affect *in vivo* assembly in different ways. A large number of fragments can negatively affect the efficiency of *in vivo* assembly as the number of repair events that cells have to perform increases with the number of fragments. Additionally, the probability that all transformed fragments access the nucleus is negatively correlated to the number of fragments (14). Fragment size is also expected to affect *in vivo* assembly efficiency, considering that cell entry might be hindered for large DNA parts. The intermediate fragment size of 5 kb might present a good compromise between these antagonistic factors. For each specific fragment size, doubling the number of fragments to assemble significantly decreased transformation efficiency by ca. 3-fold (2-tailed paired *t*-test  $P < 0.01$ ) (Figure 3A).

Next to transformation efficiency, fidelity of assembly (*i.e.* percentage of the transformants harboring all transformed fragments in the correct order) is of paramount importance for functional pathway construction. Verification of the



**Figure 3 - Transformation efficiency and fidelity.** (A) Transformation efficiencies for the *in vivo* assembly of 100 and 50 kb synthetic chromosomes. Both neochromosomes have been assembled with 2.5 kb (dark blue), 5 kb (lighter blue) and 10 kb (lightest blue) fragments. The number above each bar indicates the number of assembled fragments. Data were normalized to  $10^8$  cells using controls on YPD plates, performed in biological duplicates (replicate 1 and 2) in technical triplicates for replicate 1 and technical duplicates for replicate 2. The asterisks indicates that transformation with 5 kb fragments resulted in a significantly different transformation efficiency compared to both neighboring efficiencies (two-tailed paired homoscedastic t-test  $P < 0.05$ ). (B) Assembly fidelity. *In vivo* assembly of the 100 kb NeoChr12 from 43 fragments. Eleven colonies that grew on selective medium were tested for mRuby2 and mTurquoise2 fluorescence. Neochromosomes were subsequently screened based on their size on CHEF gel and lastly by Illumina sequencing. The percentage of correct colonies for each screening round is indicated.

neochromosome assembly by PCR, as performed in most other synthetic neochromosome assemblies (14,18), is not only too labor intensive considering the large number of assembled fragments, but it is also unreliable. Indeed, PCR erroneously indicated a correct configuration for IMF1. Sequencing dozens of chromosomes by next generation sequencing is also time-consuming and costly. We therefore set up a screening pipeline consisting in a first selection based on the presence of auxotrophic markers distributed on the neochromosome, followed by fluorescence-activated cell sorting (FACS) of transformants expressing the fluorescent markers also distributed evenly on the neochromosome, the subsequent filtering based on chromosome size by karyotyping and a final confirmation of correct assembly by sequencing (Figure 1B).

To test the fidelity of assembly and assess the ease of our screening pipeline a novel neochromosome design was made (Supplementary Figure S7). The 100 kb NeoChr12 chromosome was similar to the 100 kb neochromosome carried by IMF6, except that only two fluorescent markers, *mRuby2* and *mTurquoise2*, were included, to prevent unwanted recombinations between the highly similar *mTurquoise2* and *Venus* genes, and the strongly expressed *mRuby2* was relocated away from the centromere. From

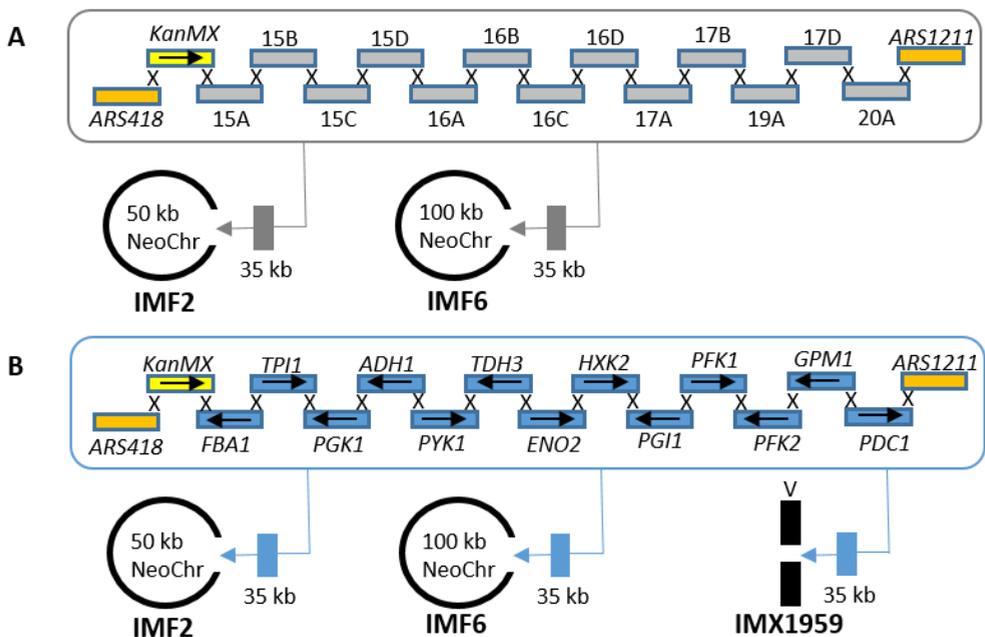
the transformation of 43 fragments of ~2.5 kb, eleven colonies were screened for correct neochromosome assembly. Eight (73%, Figure 3B and Supplementary Figure S8) showed double fluorescence, of which five displayed the correct neochromosome size by karyotyping (45% of all transformants, Supplementary Figure S9). Of these five correctly sized neochromosomes, Illumina whole genome sequencing revealed that four contained all expected fragments (Supplementary Figure S3). Therefore, a remarkably high assembly fidelity of 36% was reached when assembling 43 fragments in *S. cerevisiae*. It should be noted however that although sequencing data showed coverage for all fragments in the four correct colonies, two showed uniform coverage of all fragments (NeoChr12.4 and NeoChr12.8) while in the other two, the coverage of a small number of fragments differed from the rest of the neochromosome (four and two *E. coli* filler fragments with lower coverage for NeoChr12.1 and NeoChr12.3 respectively). This coverage variation most likely results from population heterogeneity, with subpopulations carrying different neochromosome configurations. This heterogeneity might have occurred early during neochromosome assembly, even though the strains were selected after single colony isolation, or later by recombination events occurring during strain propagation. Sequencing of IMF2 and IMF6, carrying the 50 and 100 kb original neochromosome design, confirmed a uniform coverage for all fragments in these strains. To investigate the possibility of recombination events during strain propagation, a single colony of IMF6 (100 kb original neochromosome) and of IMF23 (100 kb improved neochromosome), was inoculated in liquid selective medium and transferred in biological triplicate to fresh medium every day for fourteen days. Long-read sequencing from the first and last liquid cultures revealed that for both strains the neochromosome sequence from starting and end population was identical and faithful to the *in silico* design (Supplementary Figures S6 and S10). In conclusion, neochromosomes are genetically stable during propagation and population heterogeneity most likely occurs from neochromosome assembly and insufficient pure culture isolation.

It is important that assembly of the synthetic chromosomes does not result in a large number of mutations in the native genome or in the assembled synthetic chromosomes. Sequencing data revealed that the assembly of the neochromosomes resulted in a low number of mutations in the native genome as well as in the assembled neochromosome. NeoChr12.1 (IMF24), NeoChr12.3 (IMF25), NeoChr12.4 (IMF26) and NeoChr12.8 (IMF23) contained 3, 1, 13 and 12 SNPs in the neochromosome, respectively (between 0% and 58% of all mutations identified in the neochromosomes were located in SHRs, most likely originating from amplification primers (20)); and the native genome contained 1, 0, 2 and 0 mutations in coding regions, respectively (Supplementary Tables S1 and S2). This was similar to the previously constructed IMF2 (50 kb) and IMF6 (100 kb) strain which only harbored no and a single mutation

in coding regions of the native genome; and 5 and 16 mutations in the neochromosomes respectively (Supplementary Tables S1 and S2).

### Application of synthetic chromosomes as DNA landing pads

Supernumerary neochromosomes can be attractive landing pads to add, remove or modify functionalities. While recent chromosome engineering efforts have demonstrated that native *S. cerevisiae* chromosomes are extremely robust to large changes in their size and number (29,30), very little is known about *S. cerevisiae* tolerance to *in vivo* synthetic chromosome editing (19). The 50 and 100 kb neochromosomes carried by strains IMF2 and IMF6 respectively, were used as landing pads for the integration of an additional 35 kb of non-coding DNA (Figure 4). *mTurquoise2* was targeted by CRISPR/Cas9 for introducing a double strand break and replaced by the 35 kb construct *in vivo*-assembled from 16 DNA fragments, resulting in two chromosomes of 85 and 135 kb (Supplementary Figure S11). FACS (loss of mTurquoise2 fluorescence), diagnostic PCR, karyotyping and short- and



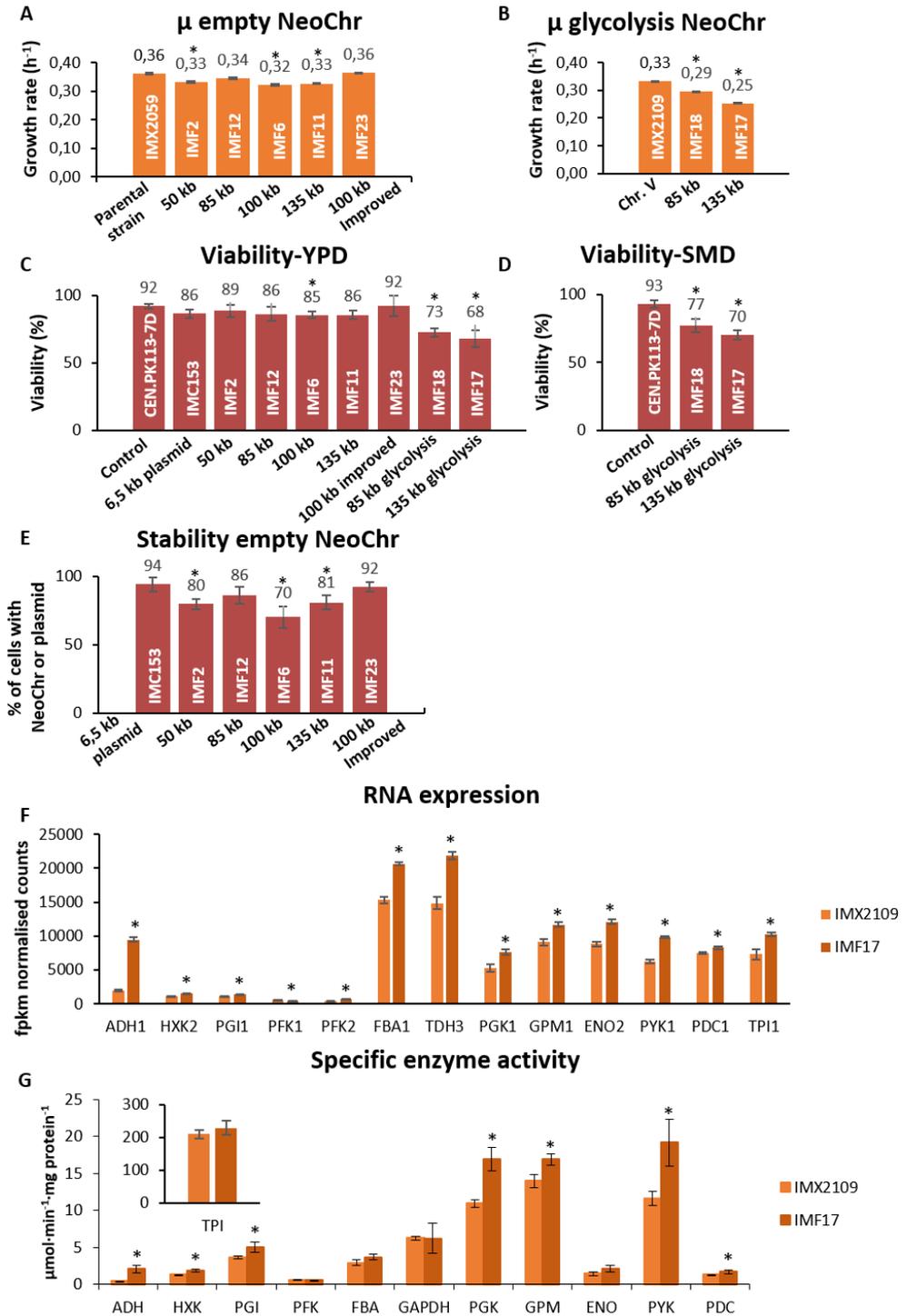
**Figure 4 - Neochromosomes as landing pads.** (A) 35 kb of non-coding DNA was simultaneously *in vivo* assembled and inserted by CRISPR/Cas9 at the *mTurquoise2* locus of the 100 kb neochromosome in IMF6 and of the 50 kb neochromosome in IMF2, resulting in 135 kb (IMF11) and 85 kb (IMF12) neochromosomes, respectively. (B) 35 kb of glycolytic genes were simultaneously *in vivo* assembled and inserted using CRISPR/Cas9 at the *mTurquoise2* locus of the 100 kb neochromosome in IMF6 and of the 50 kb neochromosome in IMF2 resulting in 135 kb (IMF13) and 85 kb (IMF14) neochromosomes, respectively. As control, the same glycolytic cassettes were integrated at the native *CAN1* locus of chromosome V resulting in strain IMX1959.

long-read sequencing revealed correct integration of all fragments (Supplementary Figures S3, S6, S12, S13).

Physiological characterization on selective SMD medium of the initial strains harboring the four differently sized neochromosomes (50 kb in strain IMF2, 85 kb in strain IMF12, 100 kb in strain IMF6 and 135 kb in strain IMF11) revealed that they grew on average 8% slower than their prototrophic parental strain (IMX2059), irrespective of neochromosome size (Figure 5). Genome sequencing did not reveal mutations that could explain the decrease in growth rate (Supplementary Tables S1 and S2). Conversely, IMF23 (100 kb NeoChr12.8), with the improved neochromosome design, in which the highly expressed *mRuby2* was re-located from its place adjacent to the centromere and in which *Venus* (highly homologous to *mTurquoise2*) was removed, showed the same growth rate as the parental strain. In addition, in comparison to the prototrophic control strain IMX2059, IMF23 showed no or very little variation in growth rate under a range of carbon sources (galactose, maltose and sucrose) and stress conditions (pH-, oxidative-, osmotic-, salt- and temperature-stress) (Supplementary Figure S14).

As *S. cerevisiae* is tolerant to variations in the number of native chromosomes (29-31), the observed slower growth phenotype for four of the neochromosome strains might be specific to the neochromosome design. It has been previously shown that the size of artificial chromosomes affects their stability (32-35). In selective medium, neochromosome loss would result in cells unable to grow and thereby in reduced growth rate. To evaluate the mitotic stability of the neochromosomes, IMF2, IMF12, IMF6, IMF11 and IMF23 were sequentially transferred in selective media and plated on both non-selective (YPD) and selective media (SMD). Along four successive culture transfers (ca. 25 generations), the viability of the strains on non-selective medium was similar to that of the prototrophic control strain CEN.PK113-7D (Figure 5), or a control strain containing a 6.5 kb centromeric plasmid (IMC153) (one-way ANOVA with Post-Hoc Tukey-Kramer,  $P > 0.05$ ).

Plating on selective medium revealed that for the four neochromosomes with the initial design (strains IMF2, IMF12, IMF6, IMF11), 14% to 30% of the population lost their neochromosome (Figure 5). This percentage remained stable over 25 generations (Supplementary Figure S15), and most likely explained the reduced growth rate measured for neochromosome-bearing strains. Indeed, the neochromosome with the improved design (IMF23), which did not show a reduced growth phenotype was highly stable with 92% of the population retaining the neochromosome, which is the same as the control strain IMC153, carrying a 6.5 kb centromeric plasmid with the same auxotrophic markers as located on the neochromosomes, and with the same parental strain as the neochromosome strains (Figure 5). Contrary to earlier reports (32,35) no

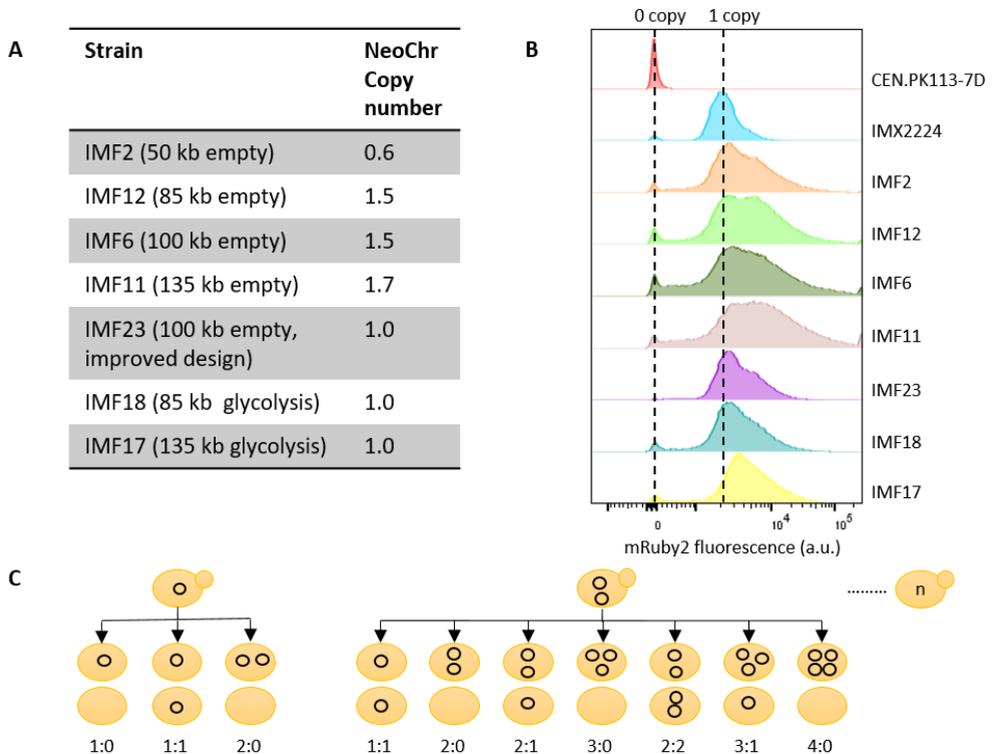


**Figure 5 - Physiological characterization of strains carrying synthetic chromosomes.** (A and B) Specific growth rate of strains carrying empty neochromosomes (Panel A: IMF2, IMF12, IMF6, IMF11 and IMF23) and glycolytic neochromosomes (Panel B: IMF18 and IMF17). Growth rates represent the average and standard deviation of biological duplicates except for the parental strain which cultures were performed in biological quadruplicate. (C and D) Viability measured as number of colonies on YPD (C) or SMD (D) sorted by FACS, divided by the total number of possible colonies (96), for the empty and glycolytic neochromosomes. Data represent biological duplicates and are averaged from 4 days of measurement for all strains except for the 100 kb improved design (two samples at day 1 and day 4). (E) Stability measured as the number of transformants on selective plates (SMD) divided by the number of colonies on non-selective plates (YPD). For each strain, the stability represents the average of 4 days of measurement in biological duplicates, except for 100 kb improved design (IMF23) for which 2 days of measurement were used (day 1 and day 4). (F) Transcript levels of the glycolytic genes from IMF17 (135 kb glycolysis neochromosome) and control strain IMX2109 expressing glycolysis from native chromosome V, grown in aerobic batch cultures. Transcript levels and standard deviations are from biological triplicates. (G) Specific activity of glycolytic enzymes in IMF17 and control strain IMX2109 from aerobic batch cultures. Activities were measured at least in biological duplicates. For panels A, B, C, D and E all significant differences with respect to the first bar are indicated with an asterisk (one-way ANOVA with Post-Hoc Tukey-Kramer,  $P < 0.05$ ). For panels F and G, the asterisk indicates whether transcript levels or enzyme activities of IMF17 are significantly different with respect to IMX2109 (two-tailed paired homoscedastic  $t$ -test  $P < 0.05$ ).

correlation was observed between neochromosome size and stability (strains IMF2, IMF12, IMF6 and IMF11 with increasing neochromosome size showed stability of 80%, 86%, 70% and 81%, respectively).

Therefore, empty neochromosomes can be very stable. The strains with the four neochromosomes (IMF2, IMF12, IMF6, IMF11) with the initial design showed decreased neochromosome stability as well as a decreased growth rate. In contrast, the improved neochromosome design was stable and resulted in a strain (IMF23) with a growth rate identical to that of the control strain (IMX2059). Deletion of *mRuby2* (IMF37) or *Venus* (IMF38), or both *mRuby2* and *Venus* (IMF39) from the 100 kb neochromosome in IMF6 did not increase the growth rate (Supplementary Figure S16). Therefore, homology between *mTurquoise2* and *Venus* (potential cause of recombination and loop-out of genetic elements essential for neochromosome maintenance) and/or strong expression of *mRuby2* close to the centromere (potentially leading to centromere destabilization), cannot account for the reduced growth rate of IMF2, IMF12, IMF6 and IMF11. It is noteworthy that *mRuby2* and *Venus* are not the only differences between the stable and unstable neochromosome designs, as they carried different filler fragments. One or more of these *E. coli* filler fragments might cause neochromosome instability. This could result in a mother cell retaining one copy of the neochromosome and the daughter cell zero copies (1:0 segregation) or the mother retaining two copies and the daughter zero (2:0 segregation) (Figure 6C). From the coverage of reads belonging to the neochromosomes, we estimated that average copy numbers of 0.6, 1.5, 1.5, 1.7, for the

50 kb (IMF2), 85 kb (IMF12), 100 kb (IMF6) and 135 kb (IMF11) neochromosomes, respectively (Figure 6A and Supplementary Figure S3). Since reads coverage was uniform along the neochromosomes, the most likely explanation for these non-integer average neochromosome copy numbers far from one, is an heterogeneous population with cells carrying none, one, two, or more neochromosome copies. Remarkably, the improved neochromosome design (strain IMF23) showed an average copy number of one, confirming its stable design. Copy number variations were confirmed by monitoring mRuby2 fluorescence, which is present on all neochromosomes. For the neochromosomes with less optimal design, the larger the neochromosome, the higher the fraction of cells with high fluorescence (indicating multiple copies, Figure 6B). For the new design (IMF23) most cells showed the same fluorescence intensity as the control in which *mRuby2* (with the same promoter and terminator) was integrated in

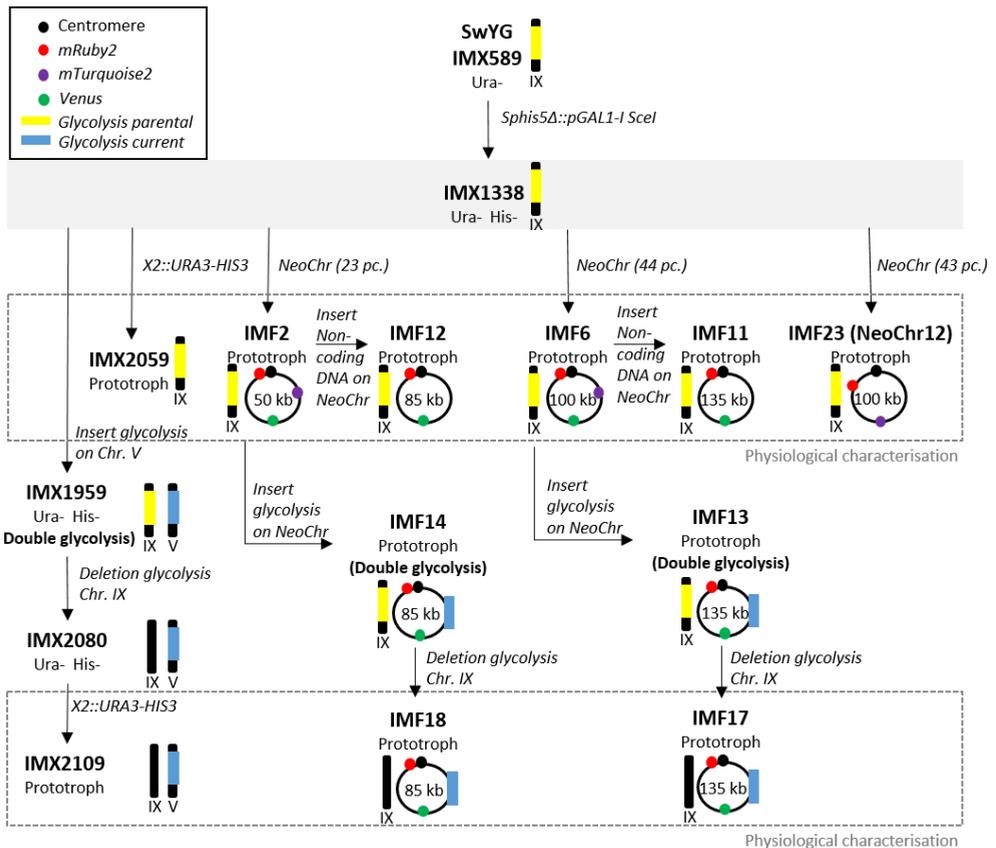


**Figure 6 - Quantification of neochromosome copy number.** (A) neochromosome copy number as determined by whole genome sequencing. (B) neochromosome copy number estimation based on mRuby2 fluorescence. CEN.PK113-7D with no copies of *mRuby2* was used as negative control and IMX2224 with a single copy of *mRuby2* integrated in the genome as positive control. (C) Schematic overview of the potential segregation of circular neochromosome upon cell division.

single copy in the genome of the neochromosome parental strain (IMX2224) (Figure 6B).

### **Synthetic chromosomes as platforms for exclusive expression of (essential) pathways**

Neochromosomes have the potential to become powerful platforms for genome engineering, to rewire yeast native gene networks as well as to act as versatile expression hubs for heterologous pathways and processes. The transplantation of the complete human purine pathway to a neochromosome (36) is an encouraging demonstration, however the potential of neochromosomes as orthogonal expression platforms hasn't been fully explored yet. Using the pathway swapping concept, we used neochromosomes as exclusive expression platform for the essential glycolytic pathway, and evaluated the impact of neochromosome-borne glycolysis expression on yeast physiology. The 50 kb (IMF2) and 100 kb (IMF6) 'empty' neochromosomes were assembled in the SwYG (16) ancestor strain IMX1338, in which the set of 13 genes coding for the glycolytic pathway has been relocated to a single genomic locus. Using CRISPR/Cas9 editing combined with *in vivo* assembly of 13 glycolytic transcriptional units (Figure 4), the 35 kb glycolytic cassette was introduced in the *mTurquoise2* locus of the 50 and 100 kb neochromosomes. In these strains harboring two glycolytic gene sets, the single locus glycolysis from chromosome IX was successfully removed by induction of double strand DNA breaks (DSBs) at the flanks of the glycolytic cassette and repaired with a 120 bp DNA fragment (16). The resulting strain with a single neochromosome-borne glycolytic gene set demonstrated that neochromosomes can serve as exclusive expression hubs for essential pathways (strain construction represented in Figure 7). The strains carrying 'glycolytic neochromosomes' IMF18 (glycolysis on 85 kb neochromosome) and IMF17 (glycolysis on 135 kb neochromosome) were grown in aerobic batch cultures in chemically defined medium with glucose as sole carbon source, and compared to IMX2109, a control strain carrying an identical genetic configuration of glycolysis on chromosome V. Both IMF17 and IMF18 grew significantly slower than the control strain (Figure 5), IMF17 growing 14% slower than IMF18. Several hypotheses could explain this decreased growth rate of strains with glycolytic neochromosomes, (i) the unfortunate occurrence of deleterious mutations in the native or synthetic chromosomes during transformation, (ii) the low expression of neochromosome-borne glycolytic genes and resulting decreased glycolytic flux, (iii) the loss of neochromosomes or (iv) stalled replication due to transcription-replication clashes at the clustered glycolysis. Both IMF17 and IMF18 carried the same mutation in the *GPM1* promoter, in addition IMF17 carried non-synonymous mutations in *TPI1* and *PFK2* ORFs (Supplementary Table S2). However, none of these mutations caused a decrease in the specific activity of the corresponding enzymes tested *in vitro* comparing IMF17 and the control strain



**Figure 7 - Strain construction overview.** The parental strain IMX1338 is framed by a filled grey box from which its direct decedents are indicated with black arrows. The schematic representations of linear or circular chromosomes indicate the relevant native or synthetic chromosomes that were modified in this study, and the significance of the symbols is explained in the top left box. Strains characterized in this study are framed by a dashed gray box.

IMX2109 (Figure 5), demonstrating that mutations in the glycolytic genes was most likely not responsible for the decreased growth rate of IMF17 and IMF18. A few mutations were observed in coding regions of other genes, and it cannot be excluded that they contributed to decreasing the growth rate of IMF17 and IMF18 (Supplementary Table S1). Secondly, the similar or slightly higher RNA expression (except *PFK1*) and specific activity in glycolytic enzymes between IMF17 and control strain IMX2109 expressing glycolysis from native chromosome V (Figure 5) indicated that low expression of neochromosome-borne glycolytic genes was not involved in the reduced growth rate of IMF17 (and IMF18). Furthermore, previous studies have shown that an increase in specific activity of glycolytic enzymes does not result in a decrease in growth rate (37,38). Alternatively, this decreased growth rate might result from neochromosome instability and potential loss in a fraction of the yeast

population. Because the loss of glycolytic neochromosomes leads to lethality in selective as well as in non-selective medium (*e.g.* YPD or YP ethanol), testing their stability is challenging. However, the viability of glycolytic neochromosome strains on YPD could be used as indicator of neochromosome loss when comparing with the viability on YPD of control strains and empty neochromosome strains IMF12 and IMF11. IMF12 and IMF11 can survive neochromosome loss on YPD medium, in which they display a viability similar to that of the control strain (85–92%). Conversely, the viability of IMF18 (85 kb glycolytic neochromosome) and IMF17 (135 kb glycolytic neochromosome), was significantly reduced on YPD to 73% and 68% respectively and on SMD medium to 77% and 70% respectively (Figure 5). Assuming that this decreased viability results from neochromosome loss, ca. 20% of the IMF17 and IMF18 population would lose their glycolytic neochromosome. The neochromosome loss would therefore be similar for neochromosomes with and without the glycolytic genes. The reduced growth rate of IMF17 and IMF18 does therefore most probably not result from a lower stability of glycolytic neochromosomes as compared to empty neochromosomes. Lastly, while stalled replication of the neochromosomes due to transcription-replication clashes might cause the decreased specific growth rate, transcriptome analysis did not support this hypothesis as no genetic markers of DNA replication stress and damage (*e.g.* *RNR1* to *RNR4* (39)) were differentially expressed between IMF17 and its control strain IMX2109 (Supplementary Table S3). Differential gene expression analysis by RNA sequencing of IMF17 and its control strain IMX2109 did not provide further leads on the cause of the reduced growth rate. Few transcriptional differences were observed between the two strains, most likely originating from technical differences in the shake flask grown cultures. This analysis additionally confirmed that *E. coli* DNA used as filler fragments for the construction of the neochromosomes was, as expected, non-coding. With the exception of a few small regions for which transcripts with low abundance could be detected, there was virtually no expression from the *E. coli* DNA (Supplementary Figure S17). It is highly unlikely that these transcripts from *E. coli* DNA lead to translation of (functional) peptides in *S. cerevisiae*.

## Discussion

The present study explores the potential of *de novo* assembled synthetic chromosomes to operate as large, modular expression islands in the popular workhorse and model strain *S. cerevisiae*. The design and construction workflow of neochromosomes developed in this study are highly versatile, enabling the facile construction of chromosomes with a wide variety of configurations.

Probing the limits of *in vivo* assembly, the present results demonstrate that a large number of DNA fragments of the size of expression cassettes can be efficiently assembled *de novo* in yeast. While other studies have reported this remarkable feature of *S. cerevisiae* (e.g. assembly of 38 fragments of 200 bp (13), and 25 fragments of 24 kb (14)), quantitative information on the transformation and assembly efficiencies and fidelity was so far lacking. The neochromosome assembly from 44 fragments, the record to date, showed a remarkably high efficiency, with 36% of faithfully assembled neochromosomes. This efficiency is substantially higher than previously reported, for instance by Annaluru *et al.* (40) where assembly of 27 fragments into chromosome III resulted in 0.5% of clones with the expected marker combination. Still, increasing the number of fragments was accompanied by a decrease in transformation efficiency and in the total number of clones on plates, which can become a limitation for the modular assembly of large neochromosomes. The construction of neochromosomes involves several cellular mechanisms, many of which are yet poorly understood but need to be considered to further enhance the neochromosomes construction efficiency. For instance, access of all required DNA fragments to the nucleus is essential for neochromosome assembly, but the mechanisms involved in DNA translocation during transformation are not fully resolved and are poorly investigated. Once all DNA fragments have reached the nucleus, the Homologous Recombination (HR) machinery has to be mobilized to repair a large number of DSBs. While the number of simultaneous DNA breaks that *S. cerevisiae* tolerates is yet undefined, hundreds of simultaneous repairs most probably present a formidable challenge. Additionally, while HR is the preferred mechanism for DSB repair in *S. cerevisiae*, other, non-homologous repair mechanisms exist (41,42), and might take over when the HR machinery is inactive or saturated, thereby promoting erroneous assemblies. Inactivation of non-homologous end joining (NHEJ) is unfortunately not an option in *S. cerevisiae* as, due to the pleiotropic role of the NHEJ proteins, it leads to deleterious consequences (43-45). Last but not least, assembly in a complete neochromosome is not a guarantee for cell survival, as they have to recover from the transformation stress, which is yet another poorly explored phenomenon. The combination of complex and often poorly understood events that have to take place for DNA to be assembled largely explains why typically only a small fraction of the transformed cell harbor the expected constructs (*ca.* one cell out of millions). One thing is however certain: the full potential of yeast *in vivo* assembly hasn't been reached yet. While intensive efforts have been made to turn *E. coli* into a universal molecular biology tool, the power of *S. cerevisiae* as DNA assembly platform has so far been largely untapped. Despite the pivotal role played by *S. cerevisiae* in the last decade in the assembly of genomes (14,46-49) there is so far relatively little effort invested in turning this yeast into a universal and powerful DNA assembly platform (50,51), a situation that we expect will change in the future.

Neochromosome stability and impact on the host physiology is a critical feature for the applicability of neochromosomes for synthetic biology. In the present study, introduction of four differently sized empty neochromosomes caused a decrease in specific growth rate and in neochromosome loss in 20% of the cell population. In-depth analysis by Illumina sequencing and flow cytometric analysis showed a lower average neochromosome copy number at population level and per cell for the smallest, 50 kb chromosome. For larger neochromosomes the average copy number was above one, considering that part of the population did not carry a neochromosome, it means that a substantial fraction of the cells carried more than one neochromosome. This phenotype is most likely explained by aberrant segregation of the neochromosomes. However this was not caused by strong expression of the gene located next to the centromere (52), possibly impairing the centromere function, nor the possible recombination between highly homologous DNA on the neochromosome. Another hypothesis that remains to be tested is that particular *E. coli* filler fragments cause chromosome destabilization. This would result from DNA architecture and not from transcription, since there was no or very little transcription from the *E. coli* DNA. The lower average copy number of the smallest neochromosome could also be caused by failed replication or loss of the neochromosome. Still, the average percentage of cells in a population (70–86%) containing at least one neochromosome copy was similar between chromosomes of 50–135 kb. Comparing the stability of the empty neochromosomes with published data is not straightforward as earlier work with YACs typically evaluates plasmid loss in non-selective media. However, the absence of size-dependency of neochromosome stability contrasts with earlier work showing that circular chromosomes increase in stability up to ~100 kb but decrease in stability above this size, a phenomenon tentatively attributed to the formation of dicentric dimeric circles (32,35,53). Considering future applications in synthetic biology, population homogeneity might be further enhanced by linear neochromosome design, either by linearization of the circular neochromosomes (for instance with the telomerator carried by the neochromosomes (25)) or direct, *de novo* assembly of linear neochromosomes. Alternatively, other strategies to improve segregation such as equipping chromosomes with synthetic kinetochores can be considered (54).

The present study demonstrates for the first time that synthetic chromosomes, fully *de novo* assembled from transcription-unit sized DNA parts can serve as exclusive, orthogonal expression platforms for essential metabolic pathways. There are few reports using synthetic and artificial chromosomes to express genes in yeast using heterologous genes or, in the case of the tRNA neochromosome, expressing a second, complete set of tRNAs, which understandably lead to a decrease in growth rate (36,53,55-58). In the present study, the native glycolytic and fermentative pathways were transplanted from a native chromosome to a neochromosome with an identical

genetic configuration (Figure 4), a unique experimental set-up that enables the direct comparison of native and neochromosome-borne expression of an entire pathway. This approach revealed that neochromosome-based expression of an essential pathway reduced the host specific growth rate by *ca.* 14–24% and suggested a negative correlation between neochromosome size and growth rate. Although this decrease in growth rate could be partly explained by aberrant segregation as observed for some of the empty neochromosomes strains, other factors most probably play a role. Reduced expression and capacity of the glycolytic and fermentative enzymes or population heterogeneity in term of neochromosome copy number are most likely not involved. One important aspect to consider is that the expression of the clustered glycolytic and fermentative genes is driven by some of the most highly transcriptionally active promoters, resulting in a transcriptional hotspot. An alternative explanation might therefore be a clash between the transcription and replication machineries, causing delayed replication (59). In IMX2109, carrying the same glycolytic cassette but on a native chromosome, such a conflict might be attenuated by the native genetic context (*e.g.* different chromatin structure, ARS sequences). Future neochromosome architecture should therefore take into account directionality of transcription with respect to neighboring ARSs and localization of highly transcribed genes.

Altogether this study highlights the potential of synthetic chromosomes to serve as platforms for modular assembly of native and heterologous pathways and paves the way towards modular genomes for synthetic biology.

## Materials and Methods

All tables and figures which are only referred to in the Materials and Methods are indicated with Supplementary Table SMM or Supplementary Figure SMM. These tables and figures can be found in the supplementary materials and methods data file at <https://doi-org.tudelft.idm.oclc.org/10.1093/nar/gkaa1167>.

### Strains, maintenance and growth media

All *S. cerevisiae* strains used in this study (Supplementary Table SMM1-SMM2) were derived from the CEN.PK family (60). For non-selective growth and propagation, the yeast strains were grown on Yeast extract-Peptone (YP) medium containing: 10 g L<sup>-1</sup> Bacto yeast extract and 20 g L<sup>-1</sup> Bacto peptone. For selective growth, Synthetic Medium (SM) was used, consisting of: 3 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 5 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 1 mL L<sup>-1</sup> of a trace element solution (61). YP and SM medium were set to pH 6.0 using 2 M KOH and sterilized by autoclaving at 110°C for 20 min and 121°C for 20 min, respectively. Subsequently, synthetic medium was supplemented with 1 mL L<sup>-1</sup> of a filter sterilized vitamin solution. As a carbon source 20 g L<sup>-1</sup> glucose, sterilized at 110°C for 20 min, was supplemented to SM or YP medium, resulting in SMD or YPD medium,

respectively. When auxotrophic strains were cultivated in SMD, it was supplemented with 125 mg L<sup>-1</sup> histidine and/or 150 mg L<sup>-1</sup> uracil. Screening of strains lacking the *URA3* gene was performed on SMD with 150 mg L<sup>-1</sup> uracil and 1 g L<sup>-1</sup> 5-FluoroOrotic Acid (SMD-5-FOA). For the selection based on the dominant markers *hphNT1* and *KanMX*, 200 mg L<sup>-1</sup> hygromycin (Hyg) and 200 mg L<sup>-1</sup> G418 were added to the medium, respectively. When G418 was used for selection in SMD medium, the 5 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> in the SM was replaced with either 1.0 g L<sup>-1</sup> of L-glutamate (monosodium: C<sub>5</sub>H<sub>8</sub>NO<sub>4</sub>Na; H<sub>2</sub>O content of 1 mol mol<sup>-1</sup>) or 2.3 g L<sup>-1</sup> urea. Moreover 6.6 g L<sup>-1</sup> K<sub>2</sub>SO<sub>4</sub> was added to compensate for sulfate supply. Solid medium was obtained by adding 20 g L<sup>-1</sup> Bacto agar to the YP or SM medium before autoclaving. Liquid yeast cultures were grown in 500 mL shake flask with 100 mL medium at 30°C and 200 rpm in an Innova incubator (New Brunswick Scientific, Edison, NJ), unless stated otherwise.

For the growth profiler experiment the strains IMX2059 and IMF23 were grown on a range of liquid media containing SM plus 2% glucose (SMD), SM plus 5% glucose (SMD 5% glu), SMD at pH 4 or 7.5, SMD plus 1 or 10mM H<sub>2</sub>O<sub>2</sub>, SMD plus 1 or 2M sorbitol, SMD plus 200 or 500 mM NaCl SM plus 2% galactose (SMGal), SM plus 2% maltose (SMMal), SM plus 2% sucrose (SMSuc). *Escherichia coli* XL1-blue was used for propagation and isolation of plasmids.

*E. coli* was grown in 5 mL Lysogeny Broth (10 g L<sup>-1</sup> Bacto tryptone, 5 g L<sup>-1</sup> Bacto yeast extract and 5 g L<sup>-1</sup> NaCl) supplemented with 100 mg L<sup>-1</sup> ampicillin when required. Cultivation was performed in 15 mL Greiner tubes or 25 mL shake flasks at 37°C and 200 rpm in an Innova 4000 shaker (New Brunswick Scientific).

For storage of *S. cerevisiae* and *E. coli* strains, the cultures were mixed with glycerol (30% v/v) and stored in 1 mL vials at -80°C.

### **Molecular biology techniques**

Genomic DNA from *S. cerevisiae* used for amplification of integrative cassettes or fragments to build synthetic chromosomes, was isolated with the YeaStar genomic DNA kit (Zymo Research, Irvine, CA) or the QIAGEN Blood & Cell Culture Kit with 100/G Genomic-tips (Qiagen, Hilden, Germany) following the manufacturer's recommendations. Genomic DNA of *E. coli*, to serve as template for fragments of the synthetic chromosomes, was isolated by pelleting 1.5 mL of overnight grown culture and resuspension of the cell pellet in 600 µL of Lysis buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.6% SDS (v/v), 0.12 g L<sup>-1</sup> proteinase K). This mixture was incubated for 1 hour at 37°C, whereafter the protocol I of the YeaStar genomic DNA kit (Zymo research) was followed from step 4 onwards (600 µL instead of 250 µL of chloroform was added). Plasmids were isolated from *E. coli* using the Sigma GenElute Plasmid DNA miniprep kit (Sigma-Aldrich, St. Louis, MO).

All DNA fragments for transformation in *S. cerevisiae* or *E. coli* of less than 10 kb were amplified by PCR using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA), all fragments of 10 kb or longer were amplified with LongRange PCR (Qiagen) according to the manufacturer's instructions. *E. coli* DNA was amplified with desalted primers while coding DNA from *S. cerevisiae* or plasmids (Supplementary Table SMM3-SMM4) was amplified with PAGE-purified oligonucleotides (all primers listed in Supplementary Tables SMM5-SMM18) (Sigma Aldrich). PCR products were purified by either one of three different methods. First, by separation using electrophoresis on a 1% (w/v) agarose gel (Thermo Fisher Scientific) in 1X TAE buffer (Thermo Fisher Scientific) with subsequent purification with the Zymoclean Gel DNA Recovery kit (Zymo Research). Second, by the GenElute PCR Clean-Up kit (Sigma-Aldrich). Third by size selection using the AMPure XP beads (Beckman Coulter, Brea, CA) according to the supplier's protocol. Purity of fragments was checked using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and concentration was measured either by the NanoDrop 2000 (Thermo Fisher Scientific) or by the Qubit dsDNA BR Assay kit (Thermo Fisher Scientific) using the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA).

Transformation in *S. cerevisiae* was executed using the lithium acetate/polyethylene glycol method (62). To verify correct transformants by diagnostic PCR, genomic DNA was isolated using an SDS/lithium acetate protocol (63) or using the YeaStar genomic DNA kit (Zymo Research). Unless stated otherwise, single-colony isolates were obtained by three consecutive re-streaks on selective solid medium and verification of the accurate genotype by diagnostic PCR before storage at -80°C. Plasmids were transformed in *E. coli* XL1-blue using chemical transformation (64) and verified by restriction analysis or diagnostic PCR after isolation. All diagnostic PCRs were performed using DreamTaq PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's instruction.

### **Plasmid construction**

Plasmids used in this study are described in Supplementary Table SMM3. Plasmids carrying the fluorescent markers *mRuby2* (pUDC191), *mTurquoise2* (pUDC192) and *Venus* (pUDC193) were constructed by Golden Gate BsaI assembly using the plasmid parts provided in the Yeast Toolkit (6). First the GFP dropout plasmid pUD538 was built using Golden Gate BsaI cloning, which consisted of ConLS (pYTK002), *sfGFP* (pYTK047), ConR1 (pYTK067), AmpR (pYTK083), *URA3* (pYTK074) and *CEN6/ARS4* (pYTK081). pUDC191 was assembled from the GFP dropout plasmid pUD538, *pCCW12* (pYTK010), *mRuby2* (pYTK034) and *tENO1* (pYTK051). pUDC192 was assembled from pUD538, *pTEF2* (pYTK014), *mTurquoise2* (pYTK032) and *tSSA1* (pYTK052). Finally, pUDC193 was assembled from pUD538, *pTEF1* (pYTK013), *Venus* (pYTK033) and *tTDH1* (pYTK056). A golden gate reaction mixture was made by using 20 fmol of each

part plasmid or dropout plasmid, 1  $\mu$ L T4 DNA ligase buffer (Thermo Fisher Scientific, Waltham, MA), 0.5  $\mu$ L T7 DNA ligase (NEB New England Biolabs, Ipswich, MA), 0.5  $\mu$ L FastDigest Eco31I (BsaI) (Thermo Fisher Scientific) and filled to a final volume of 10  $\mu$ L with MilliQ H<sub>2</sub>O. The reaction was performed in a thermocycler with 25 cycles of digestion at 42°C for 2 min followed by ligation at 16°C for 5 min. The reaction was ended by a final digestion at 60°C for 10 min followed by heat inactivation of the enzyme at 80°C for 10 min. 1  $\mu$ L of the assembly reaction was transformed in chemical competent *E. coli* XL1-blue cells as described in the supplementary materials. Correct transformants were screened based on the presence/absence of the *GFP* gene, which could be identified by colony color on the Safe Imager 2.0 Blue-Light Transilluminator (Thermo Fisher Scientific). The transformants were subsequently also confirmed by PCR.

The transcriptional units of the major glycolytic and fermentation enzymes *FBA1*, *TPI1*, *PGK1*, *ADH1*, *PYK1*, *TDH3*, *ENO2*, *HXK2*, *PGI1*, *PFK1*, *PFK2*, *GPM1* and *PDC1* were also cloned in plasmids by Golden Gate assembly to facilitate amplification by PCR (plasmids listed in Supplementary Table SMM3). The promoter (800bp)-gene-terminator (300bp) sequences were amplified from their own native location in the genome of CEN.PK113-7D using primers containing BsaI sites (Supplementary Table SMM5). Using BsaI Golden Gate assembly as described earlier, the entry plasmid pGGkd017 was constructed from pYTK002, pYTK047, pYTK072, pYTK074, pYTK082 and pYTK083 and used as backbone for assembly of the glycolysis plasmids.

pUDR400, a guide RNA (gRNA) plasmid for introducing Cas9-mediated double strand breaks in *mTurquoise2*, was constructed as described by Mans *et al.* (65) with pROS12 as backbone and gRNAcontaining primers listed in Supplementary Table SMM6. Similarly, gRNA plasmids pUDR514, pUDR547, pUDR557, pUDR289 and pUDR761 were constructed targeting the YPRcTau3 locus (primer 'YPRctau3\_targetRNA FW'), the intergenic *X2* locus (primer 'FW\_X-2\_gRNA'), the *sga1* SinLoG locus (primer 'CRISPR RNA Recycl fw'), the *mRuby2* locus (primer 'mRuby2\_targetRNA\_fw') and the *Venus* locus (primer 'Venus\_targetRNA\_fw') respectively (Supplementary Table SMM6).

### **Strain construction**

The strains constructed in this study are derived from SwYG described in Kuipers *et al.* (16) (Figure 7). SwYG is characterized by the genetic reduction and relocalization to the *SGA1* locus of the set of genes encoding the glycolytic and fermentative pathways. As host strain for synthetic chromosome assembly, IMX1338 was constructed by integration in SwYG's genome of an inducible expression cassette for the meganuclease I-SceI (66). To summarize, an I-SceI expression cassette consisting of *pGAL1-I-SceI-tCYC1* was amplified from pUDC073. Two fragments to assemble a gRNA

plasmid for a double strand break (DSB) in *SpHIS5* were amplified as described by Mans *et al.* (65). The expression cassette and two gRNA fragments were transformed in strain IMX589 and correct colonies were confirmed by diagnostic PCR. The assembled *URA3*-based gRNA plasmid was removed by counter selection on plates containing 5-FOA before the strain was stocked.

For the construction of strains carrying *in vivo* assembled synthetic chromosomes (IMF1, IMF2, IMF6), the amount of DNA fragments transformed was kept constant for all experiments with 200 fmol of each *E. coli* filler fragment, 200 fmol of each fluorescent marker and 100 fmol of the *CEN6/ARS4*, *ARS* and selectable markers. The total amount of transformed DNA therefore depended on the number and nature of DNA parts in the chromosome design (Supplementary Table SMM19-SMM20). DNA fragments for assembly of synthetic chromosomes were amplified from genomic DNA of *E. coli* XL1 blue/*E. coli* BL21, genomic DNA of *S. cerevisiae* CEN.PK113-7D or plasmids as specified in Supplementary Table SMM4. All primers for amplification of synthetic chromosome fragments consisted of a primer binding part of 18 bp to 25 bp, flanked by 60 bp Synthetic Homologous Recombination sequences (SHR) (10). IMF1, IMF2 (50 kb) and IMF6 (100 kb) were assembled from ~2.5 kb *E. coli* filler fragments, three fluorescent markers: *mRuby2*, *mTurquoise2*, *Venus*, Autonomously Replicating Sequences (*ARS*), *CEN6/ARS4*, a telomerase fragment and a *HIS3* selective marker (Figure 2, Supplementary Table SMM19-SMM20). After amplification and purification of these fragments they were concentrated by chromatography with Vivacon 500 PCR grade columns (Sartorius, Gottingen, Germany), according to the manufacturer's instruction. The fragments for each neochromosome were pooled using the amounts described above. This was supplemented with MilliQ H<sub>2</sub>O to a final volume of 50 µL, before transformation. The transformants were checked by fluorescent microscopy, FACS, CHEF and Whole Genome Sequencing (WGS) as described below. Moreover, strains IMF2 (50 kb) and IMF6 (100 kb) were verified by long-read nanopore sequencing.

The synthetic chromosomes were used as landing pads for *in vivo* assembly and integration of two types of DNA constructs, both 35 kb long. Both constructs were flanked by *ARS* sequences and carried the *KanMX* selectable marker, but differed by the presence of non-coding *E. coli* filler fragments for one, and 13 functional glycolytic and fermentative genes from yeast for the other (Supplementary Table SMM21-SMM24). Integration of these constructs in the 50 kb neochromosome of IMF2 and 100 kb neochromosome of IMF6 was facilitated by Cas9-mediated editing of the neochromosomes at the *mTurquoise2* locus. Integration of the *E. coli* and glycolytic constructs in IMF6 resulted in strains IMF11 and IMF13, respectively; and in IMF2 resulted in IMF12 and IMF14, respectively. As control, the same 35 kb glycolytic

construct was also integrated in IMX1338 genomic DNA at the *CAN1* locus using Cas9-mediated editing (strain IMX1959).

In more detail, for construction of IMF12 (85 kb) and IMF11 (135 kb), *E. coli* fragments adding up to ~35 kb were inserted at the *mTurquoise2* locus of IMF2 (50 kb) and IMF6 (100 kb). To this end, IMF2 and IMF6 were transformed with pUDR400 for introduction of a DSB at *mTurquoise2* and with sixteen DNA fragments flanked with 60 bp SHRs for *in vivo* assembly and insertion at that locus (Figure 4 and Supplementary Table SMM23-SMM24). The gRNA plasmid was removed by non-selective growth and correct transformants were verified by PCR, CHEF, WGS and long-read nanopore sequencing. For construction of IMF13 (135 kb) and IMF14 (85 kb), the major paralogs of glycolysis flanked by *ARS* sequences and the *KanMX* selective marker, were inserted at the *mTurquoise2* locus of IMF2 (50 kb) and IMF6 (100 kb). The glycolytic expression cassettes consisted of 800bp promoters, the gene and 300 bp terminators. The 16 DNA fragments (Figure 4 and Supplementary Table SMM21-SMM22), forming a total of ~35 kb, were *in vivo* assembled and inserted at the *mTurquoise2* locus, facilitated by the DSB induced by pUDR400. For construction of IMX1959, the same DNA fragments were also transformed in IMX1338 for insertion of the same configuration of the single locus glycolysis in the *can1* locus. To induce the DSB, the gRNA plasmid p426- SNR52p-gRNA.CAN1.Y-SUP4t was used. Again, gRNA plasmids were removed from all strains by nonselective growth and correct transformants were verified by PCR.

Strains IMF13, IMF14 and IMX1959 contained two genetic copies of the glycolytic and fermentative pathways. The copy harbored in the *SGA1* locus on chromosome IX was removed to obtain strains with a single locus glycolysis, either carried by a neochromosome (IMF17 and IMF18) or by chromosome V (*CAN1* locus, strain IMX2080). Removal of glycolysis from IMF13, IMF14 and IMX1959 resulted in strains IMF17, IMF18 and IMX2080 respectively. More precisely, strains IMF13, IMF14 and IMX1959 were transformed with pUDR557 and a repair fragment (synthesized by annealing the primers COUNTER SELECT oligo fw and COUNTER SELECT oligo rv (Supplementary Table SMM15)). Removal of the SinLoG at *sga1* was confirmed by PCR and gRNA plasmids were removed by non-selective growth. Finally strain IMX2080 was made prototrophic by inserting an expression cassette for *URA3* and *HIS3* in the *X2* locus. Two fragments were amplified *X2 flank-pURA3-URA3-tURA3-SHR DT* from the *S. cerevisiae* CEN.PK113-7D genome and *SHR DT-pHIS3-HIS3-tHIS3-X2flank* from pLM092 (Supplementary Table SMM16). These fragments were transformed together with gRNA plasmid pUDR547 in IMX2080. After validation of correct insertion at the *X2* locus by diagnostic PCR the gRNA plasmid was removed by non-selective growth and the strain was stocked as IMX2109. The original IMX1338 strain (single locus glycolysis on chromosome IX) was made prototrophic in the same way, resulting in

strain IMX2059. The strains IMF17, IMF18 and IMX2109 were confirmed by WGS and in addition strains IMF17 and IMF18 were confirmed by long-read nanopore sequencing. Strain IMX2059 was confirmed by diagnostic PCR.

Fluorescent control strains IMX2224 (*mRuby2*), IMX2225 (*Venus*) and IMX2226 (*mTurquoise2*) were constructed, with stable expression from the genome, with the same promoters and terminators as used for expression from the neochromosome and in the same genetic background as for the neochromosome strains. *pCCW12-mRuby2-tENO1* was amplified with primers YPRcTau3\_pCCW12\_fw and YPRcTau3\_tENO1\_rv from pUDC191, *pTEF1-Venus-tTDH1* was amplified with primers YPRcTau3\_pTEF1\_fwd and YPRcTau3\_tTDH1\_rev from pUDC193 and *pTEF2-mTurquoise2-tSSA1* was amplified with primers YPRcTau3\_pTEF2\_fw and YPRcTau3\_tSSA1\_rv from pUDC193 (Supplementary Table SMM17). The strains were made by transforming the individual fragments together with pUDR514 in IMX2059. gRNA plasmids were removed from all strains by non-selective growth and correct transformants were verified by PCR and FACS.

The prototrophic stability control strain IMC153 was made by transforming a centromeric 6.5 kb plasmid (pLM092) in IMX1338. Fluorescent control strains for *mRuby2* (IMC111), *mTurquoise2* (IMC112) and *Venus* (IMC113) were made by transforming centromeric plasmids with these markers in IMX581 (CEN.PK113-5D with Cas9).

Deletion of *mRuby2* from strain IMF6, resulting in strain IMF37, was accomplished by transformation of gRNA plasmid pUDR289 and 120bp repair DNA (made by annealing of primers *mRuby2\_repair oligo\_fw* and *mRuby2\_repair oligo\_rv* (Supplementary Table SMM17)) homologous to the 60 bp SHRs upstream (SHR AC) and downstream (SHR AD) of the *mRuby2* expression cassette. Similarly, *Venus* was deleted from IMF6 using pUDR761 and 120 bp repair DNA (primers *Venus\_repair oligo\_fw* and *Venus\_repair oligo\_rv* (Supplementary Table SMM17)), resulting in strain IMF38. Finally, both *mRuby2* and *Venus* were deleted from IMF6 by transformation of pUDR289, pUDR761 and 120 bp repair fragments for both *mRuby2* and *Venus*, resulting in strain IMF39. All three strains were grown on non-selective medium to remove the gRNA plasmid and verified by PCR and FACS before they were stocked.

All strains used in this study are listed in Supplementary Table SMM1-SMM2 and Figure 7 gives a graphical overview of the most important strains in this study.

### **Experimental quantification of neochromosome assembly efficiency**

To compare transformation efficiency of differently sized DNA fragments, 2.5 kb, 5 kb and 10 kb fragments were transformed in IMX1338 to build neochromosomes. The neochromosome design was the same as the 100 kb neochromosome of IMF6 and the

50 kb neochromosome of IMF2, except the telomerator fragment was omitted since it had only 30 bp homology in comparison to the 60 bp of the other fragments. Therefore, colonies were selected based on *HIS3* prototrophy alone instead of the *URA3* and *HIS3* prototrophy of IMF2 and IMF6. 2.5 kb *E. coli* filler fragments were amplified from *E. coli* XL1 blue/*E. coli* BL21 genome. 5 kb and 10 kb fragments were amplified from genomic DNA isolated from strain IMF1 (Supplementary Figure SMM1). The fragments were purified and concentrated before transformation. The fragments outlined in Supplementary Tables SMM25-SMM30 were transformed in the concentrations described for the construction of IMF1, IMF2 and IMF6 into IMX1338. Both a 50 kb chromosome as well as a 100 kb chromosome were assembled from all three fragment sizes. In the transformation the IMX1338 competent cells were also transformed with 50  $\mu$ L of MilliQ H<sub>2</sub>O and several dilutions ( $1.0 \times 10^5$  and  $1.0 \times 10^6$ ) were plated on YPD. For the transformation with the different size fragments also several dilutions of the transformation mix were plated (100x, 20x and 10x), which were considered as technical replicates when single colonies could be counted. Based on the colony forming units (CFU) counted on the plates, and the viability of the cells based on the YPD control, the efficiency (CFU/ $10^8$  cells) was calculated. The experiment was performed in biological duplicate, by performing the same transformations on different days.

NeoChr12, a 100 kb neochromosome with a different design, contained two fluorescent markers: *mRuby2*, *mTurquoise2* (to prevent homology between *mTurquoise2* and *Venus*), Autonomously Replicating Sequences (ARS), *CEN6/ARS4*, a telomerator fragment and a *HIS3* selective marker (Supplementary Table SMM31 and Supplementary Figure S7). After one restreak, the total number of correct transformants were assessed based on screening by FACS, CHEF and WGS (Supplementary Table SMM32). Four correct transformants were stocked as IMF23 to IMF26. In addition, strain IMF23 was verified by long-read nanopore sequencing.

### **Fluorescence detection by microscopy and flow cytometry**

To validate the expression of the fluorescent proteins *mRuby2*, *Venus* and *mTurquoise2*, from their genes on the neochromosome, the transformants were checked by fluorescent microscopy and/or by flow cytometry.

For validation based on fluorescent microscopy, single colony isolates were cultivated overnight in SMD. Subsequently, 5  $\mu$ L of culture were spotted on a 18x18 mm coverslip and placed on a microscopic slide. Fluorescence of the cells was observed by the ZEISS Axio Imager Z1 microscope (Carl Zeiss AG, Oberkochen, Germany), with three different filter sets. *mRuby2* was detected by Filter set 14 with excitation at 535 nm and emission at 590 nm. *Venus* was visualized using filter set 9, with excitation at 470 nm

and emission at 515 nm. Finally, filter set 47 was used to detect mTurquoise2 at an excitation of 436 nm and an emission of 480 nm.

The BDFACSAria™ II Cell Sorter in combination with the BDFACSDiva software (BD Biosciences, Franklin Lakes, NJ) was used for flow cytometry analysis. The machine was equipped with 355 nm, 445 nm, 488 nm, 561 nm and 640 nm lasers and a 70 µm nozzle, and operated with filtered FACSFlow™ (BD Biosciences). Prior to every experiment, a CS&T cycle with the corresponding CS&T Beads (BD Biosciences) was run to evaluate a correct cytometer performance. Furthermore, if the FACS was used for sorting an Auto Drop Delay cycle with the corresponding Accudrop Beads (BD Biosciences) was run to determine the drop delay. mTurquoise2 was excited by the 445 nm laser and the emission was detected using a 525 nm bandpass filter with a bandwidth of 50 nm. Venus was excited by the 488 nm laser and the emission was detected using a 545 nm bandpass filter with a bandwidth of 30 nm. mRuby2 was excited by the 561 nm laser and the emission was detected using a 582 nm bandpass filter with a bandwidth of 15 nm.

For FACS analysis, single colony isolates were grown overnight in SMD. In the morning the culture was transferred to new SMD at an OD<sub>660</sub> of 0.5. The culture was grown for 2 generations to an OD<sub>660</sub> of ~2.0. The level of fluorescence was analyzed by plotting the according fluorescence versus the area of the cell of 10,000 or 100,000 events. The sorting regions ('gates') of fluorescence or no fluorescence were based on the reference strain CEN.PK113-7D taken along in the experiments. FACS data were analyzed using FlowJo® software (version 3.05230, FlowJo, LLC, Ashland, OR, USA).

### **Contour-clamped homogeneous electric field (CHEF) electrophoresis**

Agarose plugs containing genomic DNA were made using the CHEF Yeast Genomic DNA Plug Kit (Bio-Rad laboratories, Hercules, CA) following the manufacturer's instructions. For in-plug linearization of the synthetic chromosomes, an I-SceI digestion (Thermo Fischer Scientific) was performed in the plugs. For this, the plugs were washed for 1 hour in 1 mL 0.1x wash buffer (Bio-Rad laboratories) per plug. Subsequently, the wash buffer was decanted and fresh wash buffer was added to sufficiently cover the plugs. The wash buffer was aspirated and the plugs were then equilibrated for 1 hour in 1 mL of 1x Tango buffer without Mg-acetate per plug. The buffer was aspirated again and 100 µL fresh Tango buffer without Mg-acetate and 2 µL of I-SceI enzyme were added to each plug. This was incubated for 2 hours on ice, after which 10 µL of 100 mM Mg-acetate was added to each plug. After 1h or overnight incubation, the buffer was removed and the plugs were kept in 1 mL 1x wash buffer each. One-half of each plug was used per well and 1 gradation (on the syringe) thin slice of the Lambda PFG Ladder (New England Biolabs) was used as standard in a 1% Certified™ Megabase Agarose (Bio-Rad laboratories) gel in 0.5x TBE Electrophoresis

Buffer (prepared from a 10x TBE Electrophoresis Buffer stock; Thermo Fischer Scientific).

The chromosomes were separated using a Bio-Rad Electrophoresis Cell in combination with a CHEF-DR® II Control Module and a CHEF-DR® II Drive module (Bio-Rad laboratories). The CHEF electrophoresis was run for 24 hours with the following parameters: an initial and final switch time of 35 seconds, a pulse angle of 120°, a voltage of 5.3 V/cm, a flow of 70 (about 0.75 L/min), a set temperature of 14°C and a buffer volume of 1.5 L. The CHEF gel was stained for 15 minutes with ethidium bromide (3 µg mL<sup>-1</sup>) in 0.5x TBE buffer and unstained for 15 minutes with 0.5x TBE buffer. The gel was visualized with an InGenius LHR gel Imaging System (Syngene, Bangalore, India).

## **Sequencing**

### Whole genome short-read sequencing

For whole genome sequencing, genomic DNA was isolated using the QIAGEN Blood & Cell Culture Kit with 100/G Genomic-tips (Qiagen), according to the manufacturer's instructions. The Miseq Reagent Kit v3 (Illumina, San Diego, CA, USA), was used to obtain 300 bp reads for paired-end sequencing. DNA was sheared to 550 bp with the M220 ultrasonicator (Covaris, Wolburn, MA) and subsequently the TruSeq DNA PCR-Free Library Preparation kit (Illumina) was employed for a six-strain library. The samples were quantified by qPCR on a Rotor-Gene Q PCR cycler (Qiagen) using the KAPA Library Quantification Kit (Kapa Biosystems, Wilmington, MA). Finally, the library was sequenced using an Illumina MiSeq sequencer (Illumina, San Diego, CA). Data were mapped to the CEN.PK113-7D (67) genome with corresponding *in silico* neochromosome design using the Burrows-Wheeler alignment tool (68) (version 0.7.15). Data was visualized using the Integrated Genomics Viewer (IGV) (69) (version 2.4.0). Chromosomal copy number variations (CNV) were quantified with the Magnolya algorithm (70) (version 0.15). Finally, an overview of single nucleotide polymorphisms (SNPs) was made using the SAMtools (68) (version 1.3.1) and Pilon (71) (with -vcf setting; version 1.18). Sequenced strains are indicated in Supplementary Table SMM1.

### Sanger short-read sequencing

Sanger sequencing (Baseclear, Leiden, The Netherlands) was performed on the original purified *E. coli* chunks used for assembly of the neochromosome (primers in Supplementary Table SMM18).

### Whole genome long-read sequencing

Genomic DNA of all samples were in-house prepared using the QIAGEN Blood & Cell Culture Kit with 100/G Genomic-tips (Qiagen) and sequenced on a MinION (Oxford

Nanopore, sequencing kit SQKLSK109, expansion kit EXP NBD104) in three runs. Samples IMF2, IMF11, IMF12, IMF17, IMF18 on a R9 flowcell, IMF6 and IMF23 on a R10 flowcell combined with samples IMF6-1, IMF6-2, IMF6-3, IMF23-1, IMF23-2 and IMF23-3 (genetic stability experiment) while these were also sequenced on a R9 flowcell. Guppy (Oxford Nanopore, version 4.0.11\_GPU) was used for basecalling and demultiplexing. The resulting fastq files were filtered on length (>1 kb) and *de novo* assembled with Flye (version 2.7.1) (72). Sample IMF17 was annotated with MAKER2 (version 2.31.8) (17), in addition the *E. coli* DNA on the synthetic chromosome was annotated with Prokka (version 1.12) (73).

### RNA-Sequencing

Strains IMF17 and IMX2109 were inoculated from a freezer stock in a wake-up culture of 100 mL SMD in a 500 mL shake flask. The strains were transferred to a pre-culture and grown until mid-exponential phase. These cultures were transferred in biological triplicate to fresh medium and grown until mid-exponential phase at an OD<sub>660</sub> of 8.65 mL of cell culture was harvested in liquid nitrogen and processed as previously described by Piper *et al.* (74). RNA was extracted with the hot-phenol method as described by Schmitt *et al.* (75). MacroGen (MacroGen Europe B.V., Amsterdam, The Netherlands) performed the cDNA library preparation using the TruSeq stranded mRNA kit (Illumina) and RNA sequencing. Sequencing was done with Illumina paired end 150 bp sequencing read system (PE150) using a 250 bp insert strand specific library. Sequencing reads of triplicates of samples IMF17 and IMX2109 were aligned to a *de novo* assembled IMF17 (this study) genome using STAR (76). Expression was quantified by featureCounts (version 1.6.0) (77). Normalized FPKM counts were obtained by applying rpkms function from the edgeR package (78). Genes were filtered (mean<sub>IMF17</sub>>10 or mean<sub>IMX2109</sub>>10) on low expression. The raw expressed genes that were kept were used in the differential expression analysis by applying DESeq2 (79).

### **Determination of genetic neochromosome stability during strain propagation**

From a -80°C freezer stock, the strains IMF6 and IMF23 were plated on selective SMD medium for single colony isolation. One single colony of each strain was inoculated in a 500 mL shakeflask containing 100 mL SMD medium. When sufficient cell density was reached, genomic DNA was isolated and sequenced with long-read nanopore technology as described earlier. For each strain, from these starter cultures, three independent propagation lines were started by transferring 200 µL of culture in 20 mL fresh SMD medium in a 100 mL shakeflask. In this manner each propagation line was transferred fourteen times. Before transfer the OD<sub>660</sub> was measured to ensure the culture was in stationary phase. For the last transfer a 500 mL shakeflask containing 100 mL of SMD medium was used such that when sufficient cell density was reached, genomic DNA could again be isolated and sequenced with long-read nanopore technology as described earlier. The neochromosomes contigs from the long-read

sequencing assemblies were compared with the *in silico* neochromosome design using the global alignment tool from Clone manager (CMSuite9, Sci Ed Software LLC, Westminster, CO).

## **Physiological characterization**

### Physiological characterization in shake flasks

A 1 mL aliquot stored at -80°C was used to inoculate a wake-up culture of 100 mL SMD in a 500 mL shake flask. The culture was grown until late exponential phase and used to inoculate a pre-culture. The pre-culture was grown until mid-exponential phase and used to inoculate a measuring culture in biological duplicate at an initial OD<sub>660</sub> of 0.3. Optical densities at 660 nm of the cultures were monitored by a Jenway 7200 spectrophotometer in technical duplicate (Cole-Parmer, Vernon Hills, IL). A maximum specific growth rate ( $\mu_{\max}$ ) was calculated from at least 5 data points in the exponential phase.

### Physiological characterization in the growth profiler

Growth studies in microtiter plate were performed at 30 °C or 37 °C and 250 rpm using a Growth Profiler 960 (EnzyScreen BV, Heemstede, The Netherlands). Strains from glycerol freezer stocks were inoculated and grown overnight in 10 mL SMD medium in a 50 mL shake flask. This culture was used to inoculate a pre-culture in 10 mL SMD medium in a 50 mL shake flask, which was cultivated until mid-exponential growth. From this culture the growth study was started in a 96-wells microtiter plate (EnzyScreen, type CR1496dl), with final working volumes of 250  $\mu$ L and approximate starting OD<sub>660</sub> of 0.3. Microtiter plates were closed with a sandwich cover (EnzyScreen, type CR1296). Images of cultures were made at 30 min intervals. Green-values for each well were corrected for plate position using measurements of a culture of OD<sub>660</sub> 5.05 of CEN.PK113-7D. Corrected green values were converted to OD-values based on a 15-point calibration, resulting in the following equation: OD-equivalent =  $a \times Gv(t) + b \times Gv(t)^c - d$  in which  $Gv(t)$  is the corrected green-value measured in a well at time point 't'. The constants used for calculation are position dependent. For this study two microtiter plates were used: MTP07 (a:0.0738, b:2.06\*10<sup>-7</sup>, c: 3.5737, d:-1.2655) and MPT08 (a:0.0774, b:1.66\*10<sup>-7</sup>, c: 3.6235, d:-1.6150)

Growth rates were calculated in the time frame where the calculated OD was between 2 and 10. Each experimental condition was performed in biological triplicates.

## **Neochromosome stability determination**

Neochromosome stability was assessed based on plating on selective and non-selective growth medium with respect to the neochromosome borne *URA3* and *HIS3* markers. Strains were inoculated from a freezer stock in 20 mL of selective SMD medium in a

100 mL shake flask and grown overnight. The strains were then inoculated in biological duplicate to fresh 20 mL medium in a 100 mL shake flask at an OD<sub>660</sub> of 0.5 and were grown for 2 generations until an approximate OD<sub>660</sub> of 2.0. From the culture 3 mL were taken, washed in sterile COULTER® ISOTON® II Diluent (Beckman Coulter, Brea, CA) and subsequently resuspended in 3 mL sterile COULTER® ISOTON® II Diluent (Beckman Coulter) to be sorted by the FACS. 96 cells of the entire population were sorted on both solid non-selective YPD and solid selective SMD medium in a single well OmniTray (Thermo Fischer Scientific).

The viability was determined by:

$$\text{Viability (\%)} = \frac{\text{number of colonies grown on YPD plate}}{96} \times 100\%$$

The stability was determined by:

$$\text{Stability (\%)} = \frac{\text{number of colonies grown on SMD plate}}{\text{number of colonies on YPD plate}} \times 100\%$$

After the samples for the FACS were taken the cultures were placed back in the incubator and grown until the morning, after which the biological duplicates of each strain were transferred to new medium at an OD<sub>660</sub> of 0.5 and the abovementioned protocol was repeated. In total, the strains were analyzed by FACS for 4 consecutive days, except for IMF23 which was analyzed on the first and the fourth day.

### ***In vitro* enzyme activities**

For determining the glycolytic enzyme activities, strains IMX2109 and IMF17 was inoculated from a freezer stock in 100 mL selective SMD medium in a 500 mL shake flask. They were transferred in biological duplicate to fresh 100 mL SMD medium and grown until mid-exponential phase at an OD<sub>660</sub> of 8. 50 mL of cell culture were harvested from each biological duplicate by centrifuging for 10 minutes at 5000 rpm at 4°C. Cell extracts were prepared as described by Postma *et al.* (80). Assays were prepared in at least technical duplicates for HXK, PGI, FBA, TPI, GAPDH (16.5 U of enzyme instead of 25U), PGK, GPM, ENO, PYK, PDC and ADH as described by Jansen *et al.* (81) and PFK according to Cruz *et al.* (82). Cell extracts were diluted with 100 mM KH<sub>2</sub>PO<sub>4</sub> buffer with 2 mM MgCl<sub>2</sub> (pH 7.5) when necessary, except for the assay for TPI, for which the cell extract was diluted with demineralized water. The spectrophotometric assays were carried out in a U-3010 spectrophotometer (Hitachi, Tokyo, Japan) at 30 °C and 340 nm. The protocol by Lowry *et al.* (83) was used to determine protein concentrations in the cell extract, using bovine serum albumin (fatty acid-free) as standard.

## **Statistical analysis**

If replicates were performed, data are presented as mean  $\pm$  s.d. and the number of replicates is indicated. Statistical significance was determined either by using a t-test (paired or unpaired is indicated) for comparison of two samples using excel or for comparison of multiple samples by ANOVA with Post-Hoc Tukey-Kramer using GraphPad Prism 4 (Graphpad, San Diego, CA). For both tests, differences were considered significant when  $P < 0.05$ .

## **Data availability**

All genomic data for this chapter have been deposited in the NCBI database (<https://www.ncbi.nlm.nih.gov.tudelft.idm.oclc.org/>) under the BioProjectID PRJNA596648. All RNA-seq sequences have been deposited at NCBI GEO (<https://www.ncbi.nlm.nih.gov.tudelft.idm.oclc.org/geo/>) under GEO accession number: 'GSE159524'.

## **Acknowledgements**

We thank Marijke Luttkik for assistance with assaying the glycolytic enzymes activity, Mark Bisschops for technical assistance with FACS analysis, Melanie Wijsman for assistance with the neochromosome stability assays, Jordi Geelhoed for performing the growth profiler experiment and Pilar de la Torre for (RNA) sequencing of strains. We thank Anna Wronska and Thomas Perli for sharing the plasmids pGGkd017 and pUDR514, respectively. This project was funded by the AdLibYeast European Research Council (ERC) consolidator 648141 grant awarded to Pascale Daran-Lapujade.

## Supplementary data

All tables and figures which are only referred to in the Supplementary data or the materials and methods are indicated with Supplementary Table SMM or Supplementary Figure SMM. These tables and figures can be found in the supplementary materials and methods data file at

<https://doi-org.tudelft.idm.oclc.org/10.1093/nar/gkaa1167>.

**Table S1 Amino acid substitutions identified in the genome of the constructed strains as compared to most relevant parental strain.**

Systematic name	Name	Type	Amino acid change
<b>IMF2 compared to IMX1338</b>			
No mutations observed			
<b>IMF6 compared to IMX1338</b>			
YEL074W	-	Non-synonymous	Pro-34-His
<b>IMF24 (genome of NeoChr12.1) compared to IMX1338</b>			
YER116C	<i>SLX8</i>	Non-synonymous	Pro-150-Leu
<b>IMF25 (genome NeoChr12.3) compared to IMX1338</b>			
No mutations observed			
<b>IMF26 (genome NeoChr12.4) compared to IMX1338</b>			
YCL052C	<i>PBN1</i>	Non-synonymous	Ile-109-Thr
YOL076W	<i>MDM20</i>	Non-synonymous	Leu-249-Ile
<b>IMF23 (genome NeoChr12.8) compared to IMX1338</b>			
No mutations observed			
<b>IMF12 compared to IMF2</b>			
YMR323W	<i>ERR3</i>	Non-synonymous	Leu-447-Phe
<b>IMF11 compared to IMF6</b>			
YKL205W	<i>LOS1</i>	Synonymous	Val-860-Val
<b>IMF18 compared to IMF2</b>			
YBR084W	<i>MIS1</i>	Non-synonymous	Phe-547-Leu
YDL159W	<i>STE7</i>	Non-synonymous	Cys-447-Ser
YDR295C	<i>HDA2</i>	Non-synonymous	Ala-556-Glu
YGL150C	<i>INO80</i>	Non-synonymous	Ala-882-Val
YBL113C	-	Intron	-
YPL006W	<i>NCR1</i>	Synonymous	Gly-119-Gly
YPR147C	-	Synonymous	Leu-52-Leu
<b>IMF17 compared to IMF6</b>			
YFL025C	<i>BST1</i>	Non-synonymous	Asn-791-His
YIL169C	<i>CSS1</i>	Non-synonymous	Thr-906-Ser
YKL176C	<i>LST4</i>	Non-synonymous	Ser-705-Tyr
<b>IMX2109 compared to IMX1338</b>			
YJR162C	-	Synonymous	Thr-33-Thr
YIL130W	<i>ASG1</i>	Non-synonymous	Asp-617-Asn
YKR004C	<i>ECM9</i>	Non-synonymous	Ser-47-Tyr
YPL283W-A	-	Intron	

YLR172C	<i>DPH5</i>	Non-synonymous	Ala-79-Thr
YNR050C	<i>LYS9</i>	Synonymous	Phe-150-Phe

**Table S2 Mutation identified in Synthetic Chromosomes as compared with *in silico* design and with the most relevant parental strain**

For the neochromosomes containing *E. coli* DNA it was discovered that the template DNA used for PCR amplification and subsequent transformation was a mix of two *E. coli* strains: XL1-Blue and BL21. Therefore, all mutations identified from the *in silico* neochromosome design (based on *E. coli* BL21) that were identical to the sequence of *E. coli* XL1-Blue were discarded. To verify that indeed these variations were already present in PCR fragments used for transformation and did not occur during neochromosome assembly/propagation, a part of the PCR fragments used for transformation were sanger sequenced (Supplementary Figure SMM2). For IMF11, IMF12, IMF17 and IMF18, five SNPs were observed for ARS418 and 2 for ARS1211 with respect to the *in silico* design. However, these were verified not to be real SNPs, since for the *in silico* template *S. cerevisiae* S228C was used while ARS418 and ARS1211 were amplified from *S. cerevisiae* CEN.PK113-7D. These strains indeed differ in the 7 SNPs observed for the two ARSs.

\*The SNP in *TPI1* for IMF17 was a non-synonymous Val-7-Phe mutation.

#The SNP in *PFK2* for IMF17 was a non-synonymous Met-390-Ile mutation

Position	Fragment	Mutation type
<b>NeoChr2 (IMF2)</b>		
12242	<i>HIS3</i> promoter	C to CA
18451	SHR BD	C to CT
24938	<i>TDH1</i> terminator ( <i>Venus</i> )	34 bp deletion
42085	SHR BK	C to CA
48959	SHR AD	G to A
<b>NeoChr1 (IMF6)</b>		
1703	Chunk 1A	A to T
22970	Chunk 3A	C to T
30653	SHR AH	AC to A
32769	<i>HIS3</i> promoter	C to CA
38952	SHR BD	T to TG
46433	SHR BF	CA to C
54835	Chunk 5A	A to T
56692	SHR BG	T to TC
59309	SHR BH	AT to A
59315	SHR BH	AC to A
65869	<i>TDH1</i> terminator ( <i>Venus</i> )	34 bp deletion
71313	SHR BQ	C to CT
77341	Telomerator ( <i>URA3</i> CDS)	G to GA (in stretch of A's)
77997	SHR AO	C to CA
87376	Chunk 7D	G to A
100069	SHR AD	A to AG
<b>NeoChr12.1 (IMF24)</b>		
17843	SHR BO	G to GC
39753	Chunk 15C	Insertion of 12 bp
97403	<i>SSA1</i> terminator ( <i>mTurquoise2</i> )	CAT to C

<b>NeoChr12.3 (IMF25)</b>		
64232	<i>HIS3</i> promoter	C to CA
<b>NeoChr12.4 (IMF26)</b>		
7596	SHR BL	A to AT
20696	<i>CCW12</i> promoter ( <i>mRuby2</i> )	CA to C
22174	<i>ARS1</i>	A to AC
24751	Chunk 18A	GGT to G
37589	SHR DH	C to T
43110	<i>CEN6/ARS4</i>	AT to A
51015	SHR DM	GA to G
53542	SHR DN	C to T
64232	<i>HIS3</i> promoter	C to CA
65169	<i>HIS3</i> terminator	C to T
70141	Chunk 4B	C to T
75441	SHR AK	T to TC
97403	<i>SSA1</i> terminator ( <i>mTurquoise2</i> )	CAT to C
<b>NeoChr12.8 (IMF23)</b>		
43	SHR AO	AT to A
20696	<i>CCW12</i> promoter ( <i>mRuby2</i> )	C to CA
40152	SHR DI	A to AC
40171	SHR DI	AG to A
41708	Chunk 15D	G to T
56968	Chunk 17B	C to T
64232	<i>HIS3</i> promoter	C to CA
77907	SHR BF	A to AG
77911	SHR BF	AC to A
80507	SHR BS	Deletion of 12 bp
93363	SHR BI	GA to G
97595	SHR DS	T to TC
<b>NeoChr20 (IMF12) as compared to NeoChr2 (IMF2)</b>		
10248	SHR AH	C to CG
14530	SHR DF	G to C
23462	Chunk 16A	G to A
33079	Chunk 17A	A to G
40176	Chunk 17D	TA to T
<b>NeoChr19 (IMF11) as compared to NeoChr1 (IMF6)</b>		
40004	SHR DI	GC to G
62440	Chunk 17D	C to T
63119	SHR DR	AG to A
<b>NeoChr16 (IMF18) as compared to NeoChr2 (IMF2)</b>		
23389	SHR 0	CT to C
30653	SHR C	A to AG
43179	<i>GPM1</i> promoter	Deletion of 18 bp (AT's)
<b>NeoChr15 (IMF17) as compared to NeoChr1 (IMF6)</b>		
32334	SHR AA	AT to A
35439	<i>TP11*</i>	G to T
41106	SHR N	T to TG
60046	<i>PFK2 #</i>	C to G
62020	SHR L	AT to A
63582	<i>GPM1</i> promoter	Deletion of 18 bp (AT's)

63970	SHR M	T to TG
105411	Chunk 6B	Insertion of 59 bp
105412	Chunk 6B	T to C
105415	Chunk 6B	C to T
105417	Chunk 6B	G to A
105421	Chunk 6B	T to C
105430	Chunk 6B	G to A
8590	<i>tADH1</i>	A to T
20368	<i>tHXK2</i>	AT to A
30676	<i>pPFK2</i>	GA to G
32946	<i>pGPM1</i>	Deletion of 18 bp (AT's)

**Table S3 RNA sequencing of IMX2109 and IMF17: Analysis of differentially expressed genes**

From mid-exponential aerobic SMD cultures of IMX2109 and IMF17, RNA was isolated, processed and sequenced. Non-expressed genes were filtered out of the differential expression analysis (Supplementary Figure SMM3). In total the differential expression analysis (padj<0.05 and FC>2) revealed 67 differentially expressed genes in the native genome. Five (**A**) from the 67 genes were differentially expressed because they were absent in IMX2109 since the 2 $\mu$  plasmid was lost from this strain and the glycolysis was inserted at *CAN1*. In addition, 11 genes showed for both strain IMX2109 and IMF17, fpkm values which were close to the detection level (normalized fpkm count<20). This results in 51 meaningful differentially expressed genes on the native genome (**B**), which indicates a very small transcriptional difference between strains IMX2109 and IMF17. In addition, most of the genes (65 %) show only a mild transcriptional response (FC<3). FUNSPEC (funspec.med.utoronto.ca, Bonferroni correction and P-value cutoff=0.05) was used to evaluate which classes and categories of genes are enriched based on the GO Database (C) and MIPS Database (D).

**A) Non-expressed genes in IMX2109**

Systematic name	Name	Fold change
<b>R0040C</b>	<i>REP2</i>	-
<b>R0020C</b>	<i>REP1</i>	-
<b>R0030W</b>	<i>RAF1</i>	-
<b>R0010W</b>	<i>FLP1</i>	-
<b>YEL063C</b>	<i>CAN1</i>	-

## B) Differentially expressed genes

Systematic name	Name	Fold change
YDR077W	<i>SED1</i>	2,5
YBR054W	<i>YRO2</i>	4,2
YFL014W	<i>HSP12</i>	4,8
YBR072W	<i>HSP26</i>	4,0
YER062C	<i>HOR2</i>	2,9
YDL022W	<i>GPD1</i>	2,1
YNL160W	<i>YGP1</i>	9,6
YLR327C	<i>TMA10</i>	2,4
YML054C	<i>CYB2</i>	2,3
YKL096W	<i>CWP1</i>	2,4
YGR144W	<i>THI4</i>	13,7
YBR093C	<i>PHO5</i>	5,7
YOL052C-A	<i>DDR2</i>	2,8
YER053C	<i>PIC2</i>	2,5
YML123C	<i>PHO84</i>	2,2
YGL121C	<i>GPG1</i>	3,4
YHR087W	<i>RTC3</i>	3,1
YOR374W	<i>ALD4</i>	2,1
YGR289C	<i>MAL11</i>	9,3
YNR034W-A	Uncharacterized	4,6
YGR292W	<i>MAL12</i>	15,7
YER150W	<i>SPI1</i>	2,7
YDR516C	<i>EMI2</i>	2,3
YER067W	<i>RGI1</i>	2,6
YDR119W-	Uncharacterized	2,1

YIL136W	<i>OM45</i>	2,1
YOR161C	<i>PNS1</i>	2,1
YBR092C	<i>PHO3</i>	3,3
YDR070C	<i>FMP16</i>	3
YOL055C	<i>THI20</i>	2,1
YML100W	<i>TSL1</i>	2,3
YMR105C	<i>PGM2</i>	2,1
YBL049W	<i>MOH1</i>	2,8
YBR169C	<i>SSE2</i>	2,5
YER037W	<i>PHM8</i>	3
YFL030W	<i>AGX1</i>	2,6
YOR173W	<i>DCS2</i>	2,5
YGR248W	<i>SOL4</i>	2,5
YJL108C	<i>PRM10</i>	2,4
YGR088W	<i>CTT1</i>	2,6
YFL059W	<i>SNZ3</i>	2,4
YNL333W	<i>SNZ2</i>	2
YEL039C	<i>CYC7</i>	5,1
YGR043C	<i>NQM1</i>	3,9
YHR136C	<i>SPL2</i>	2,1
YPR160W	<i>GPH1</i>	4,8
YAL061W	<i>BDH2</i>	3
YJL161W	<i>FMP33</i>	2,7
YMR169C	<i>ALD3</i>	2

## C) GO Biological Process

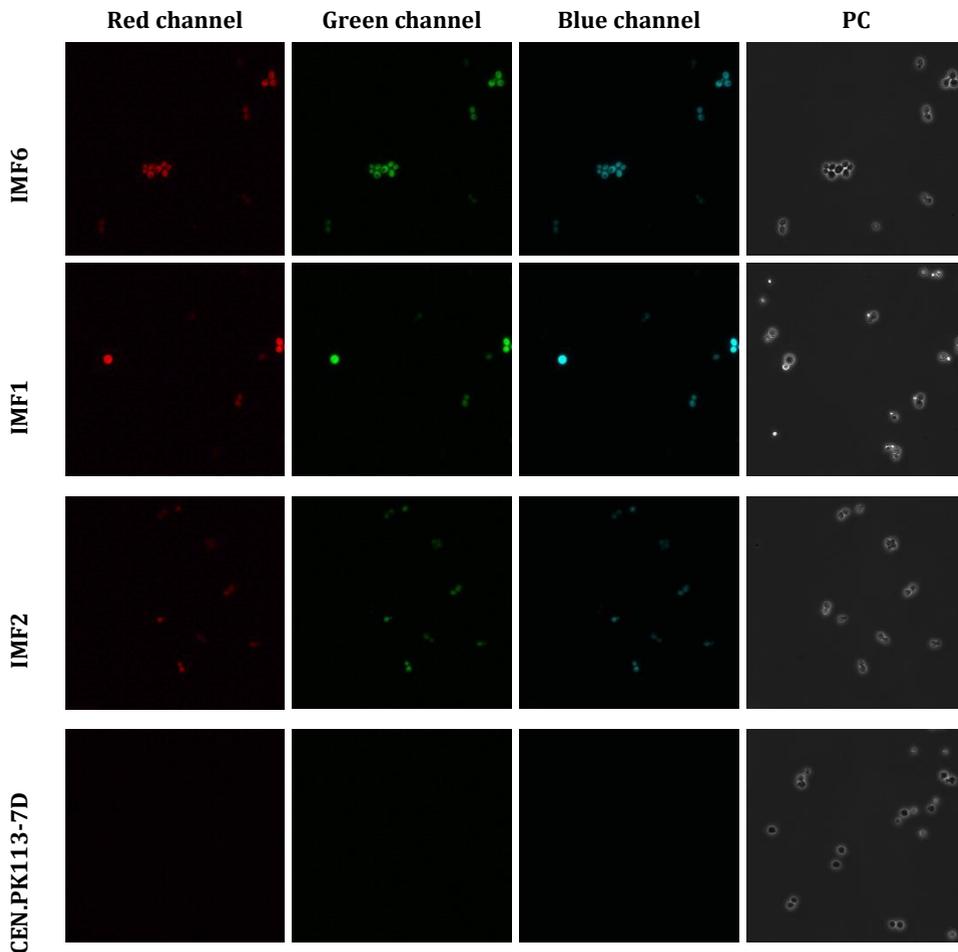
Category	P-value	In Category from Cluster	k	f
response to stress [GO:0006950]	1.908e-06	<i>HSP26 SSE2 HSP30 GPD1 HOR2 HSP12 CTT1 TSL1 DDR2</i>	9	152
carbohydrate metabolic process [GO:0005975]	6.574e-06	<i>GPD1 EMI2 NQM1 SOL4 MAL12 PGM2 GPH1</i>	7	94
thiamine biosynthetic process [GO:0009228]	6.978e-06	<i>SNZ3 THI4 SNZ2 THI20</i>	4	17

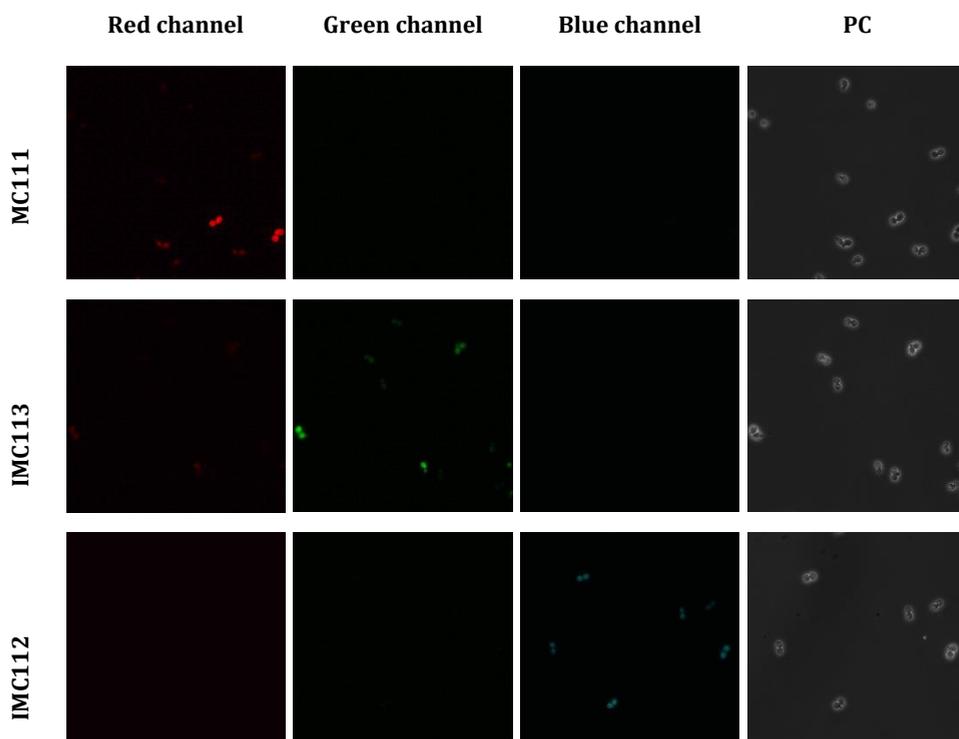
#### D) MIPS Functional Classification

Category	P-value	In Category from Cluster	k	f
stress response [32.01]	2.195e-09	<i>YRO2 SSE2 HSP30 SED1 CYC7 SNZ3 PIR3 TSL1 ALD3 YGP1 SNZ2 DDR2</i>	12	162
C-compound and carbohydrate metabolism [01.05]	4.296e-05	<i>BDH2 PHO3 PHO5 GPD1 EMI2 HOR2 HSP12 CYB2 ALD3</i>	9	223

#### Figure S1 Microscope imaging of neochromosomes

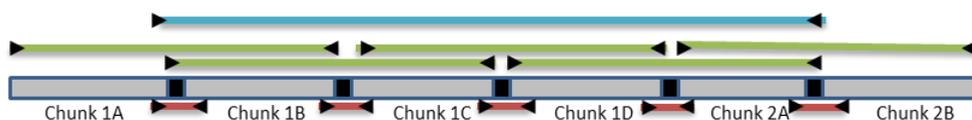
Sample images captured with fluorescence microscopy (40x). IMF6, IMF1 and IMF2 contained a putative correct assembly of NeoChr1 (for IMF6 and IMF1) and NeoChr2 respectively, based on fluorescence for the red, green and blue channels. mRuby2 control: IMC111, Venus control: IMC113, mTurquoise2 control: IMC112, Negative control: CEN.PK113-7D. PC: Phase Contrast. For visibility red channel images were adjusted by 60 % in brightness. The red bar in the bottom right of the fluorescence microscopy pictures (40x) represents 10  $\mu$ m.





**Figure S2 Strategy to test correct assembly of a neochromosome by diagnostic PCR.**

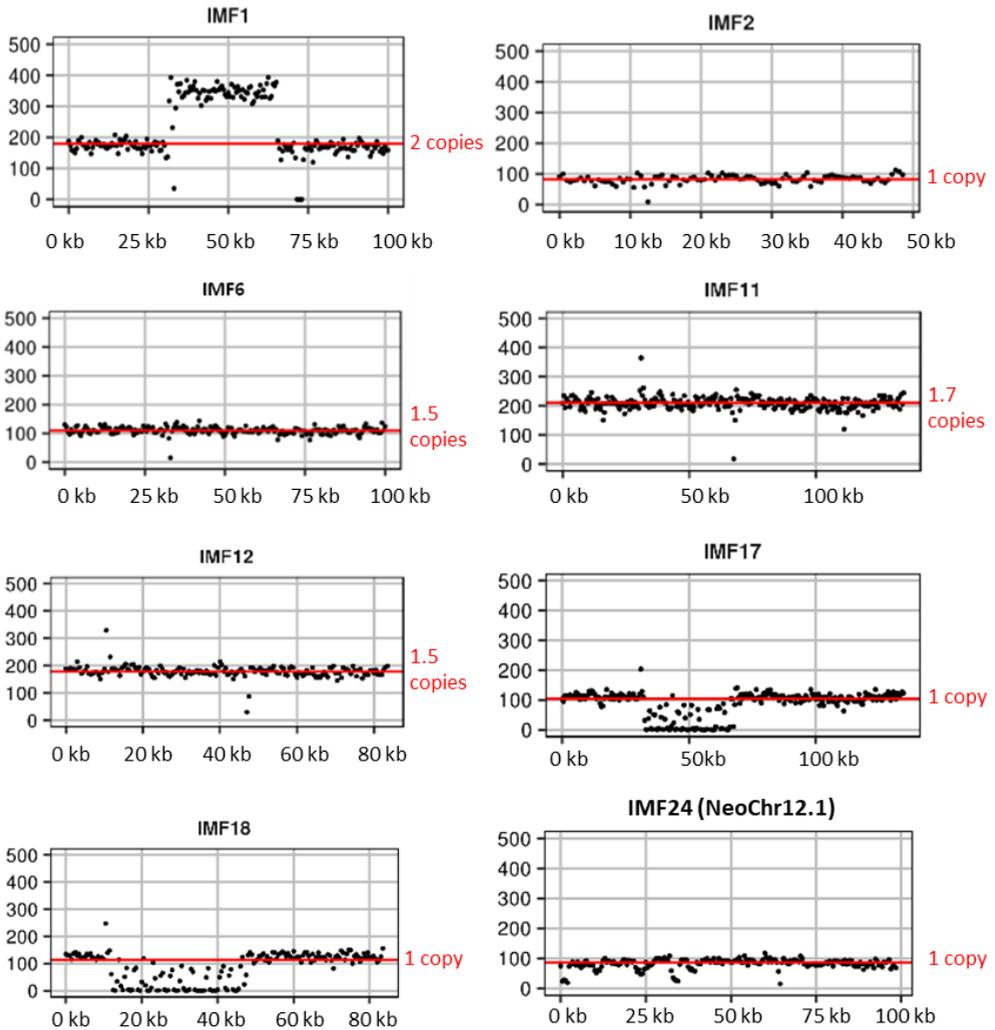
The neochromosome in strains IMF1, IMF2 and IMF6 were initially checked by diagnostic PCR. Correct assembly of adjacent fragments was checked in three possible ways. Most often amplification of a 5 kb fragment confirming two adjacent *E. coli* chunks (green). Otherwise an amplification of 10 kb fragment spanning 4 *E. coli* chunks (blue). Lastly, an amplification of a short PCR fragment spanning over an SHR (red).

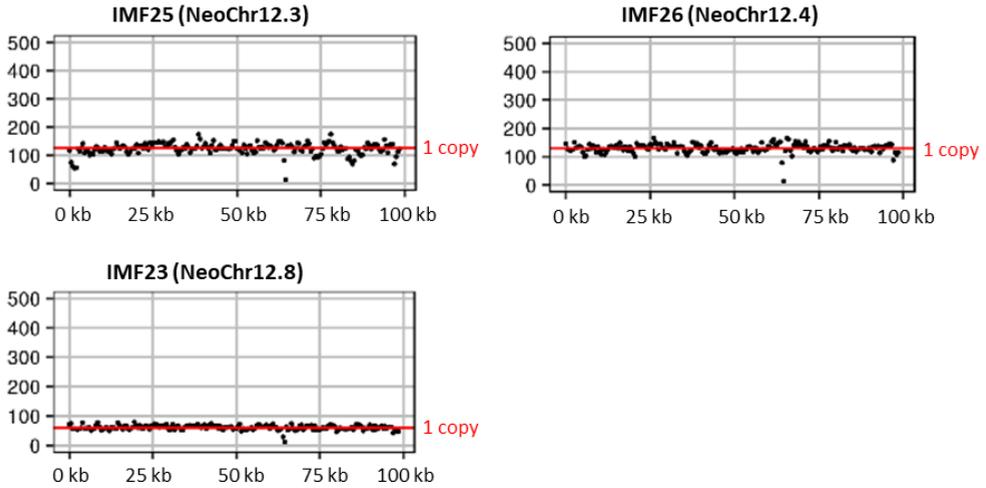


**Figure S3 Sequencing: coverage and Magnolya plots of neochromosomes**

WGS data from strains carrying neochromosome. Both the coverage plot (left) and the Magnolya plot (right) for each neochromosome are represented. The y-axis of the coverage plots represents the read coverage (a.u.). For IMF1, a whole segment in the middle of the chromosome from the middle of *mTurquoise2* to the middle of *Venus* is duplicated. All other strains harbor all fragments of the neochromosomes in the correct order. For parts in neochromosome the

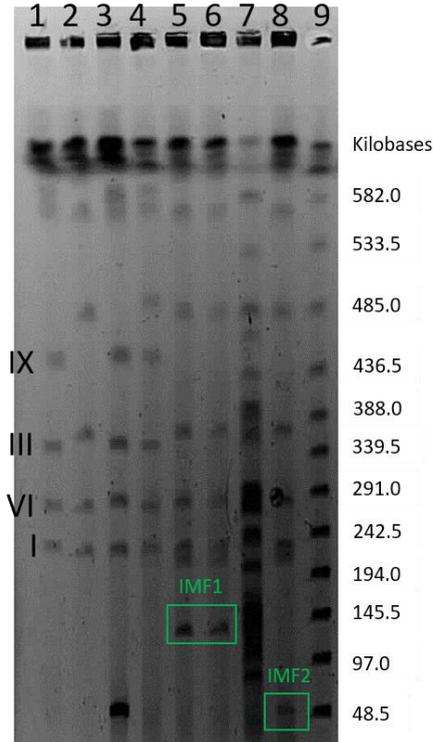
coverage is 0 or 0.5 (ex. IMF18 and IMF17). This is because CEN.PK113-7D was used as reference strain to align to the sequencing reads to. This reference contains promoters/genes/terminators which are absent in the actual strain (IMX1338) in which the neochromosomes were constructed. Therefore, reads will (also) map to this native location instead of to the neochromosome. All these sites were verified to be present in the particular neochromosome. From the Magnolya plots it can be observed that the neochromosome in IMF2 is present on average in 0.6 copies/cell, for IMF12 and IMF6 on average in 1.5 copies/cell, for IMF11 on average in 1.7 copies/cell and for IMF1 on average in 2.0 copies/cell. The neochromosomes: NeoChr12.1 (IMF24), NeoChr12.3 (IMF25), NeoChr12.4 (IMF26) and NeoChr12.8 (IMF23) are on average in 1 copy/cell, as well as the neochromosome in IMF18 and IMF17.





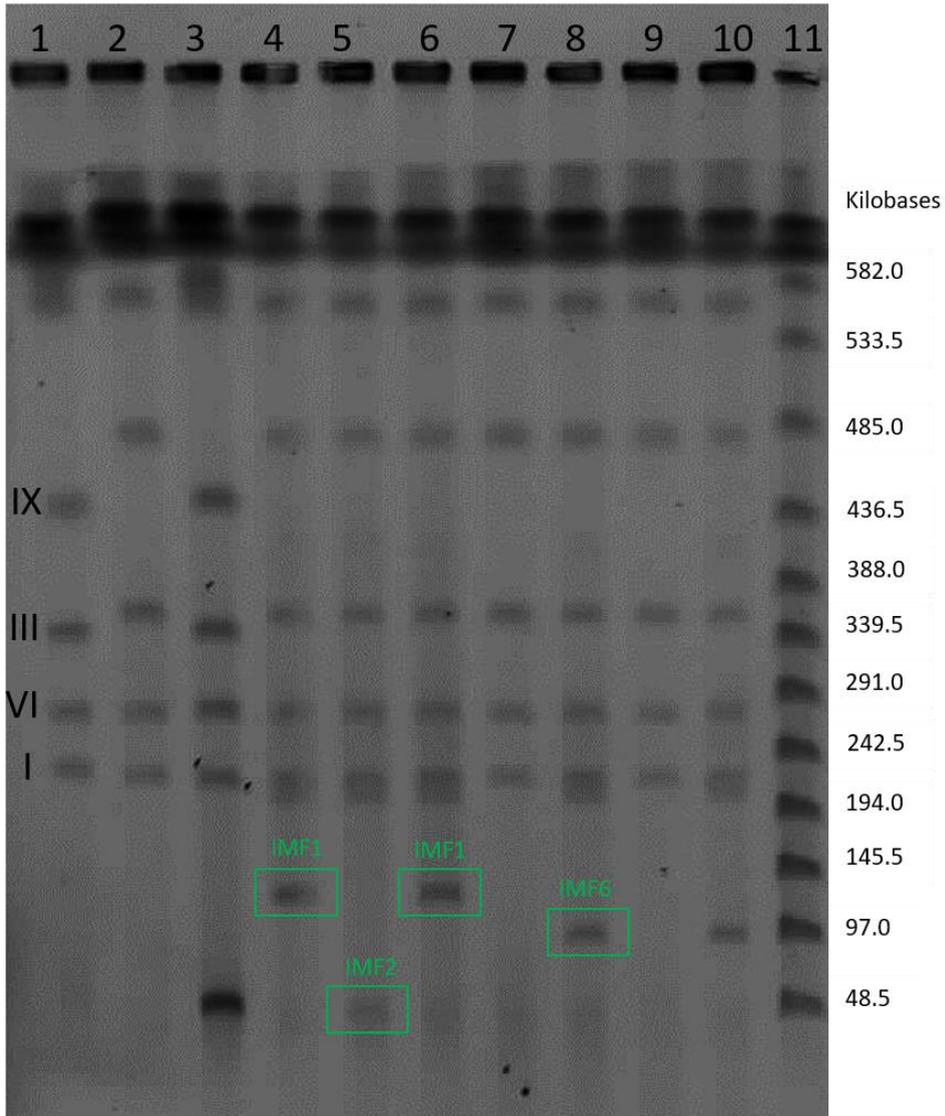
**Figure S4 CHEF analysis of IMF1 and IMF2**

CHEF analysis of IMF1 and IMF2. 1) CEN.PK113-5D (- control), 2) IMX1338 (- control), 3) IMS0480 (+ control), 4) in plug I-SceI digested IMX212, 5) in plug I-SceI digested (1h) IMF1, 6) in plug I-SceI digested (O/N) IMF1, 7) In plug NotI digested IMF1, 8) In plug I-SceI digested IMF2, 9) Lambda PFG ladder. NeoChr1 in IMF1, shows it is approximately 135 kb, which is 35 kb larger than the *in silico* design. NeoChr2 in IMF2 has the size of 50 kb as *in silico* designed.



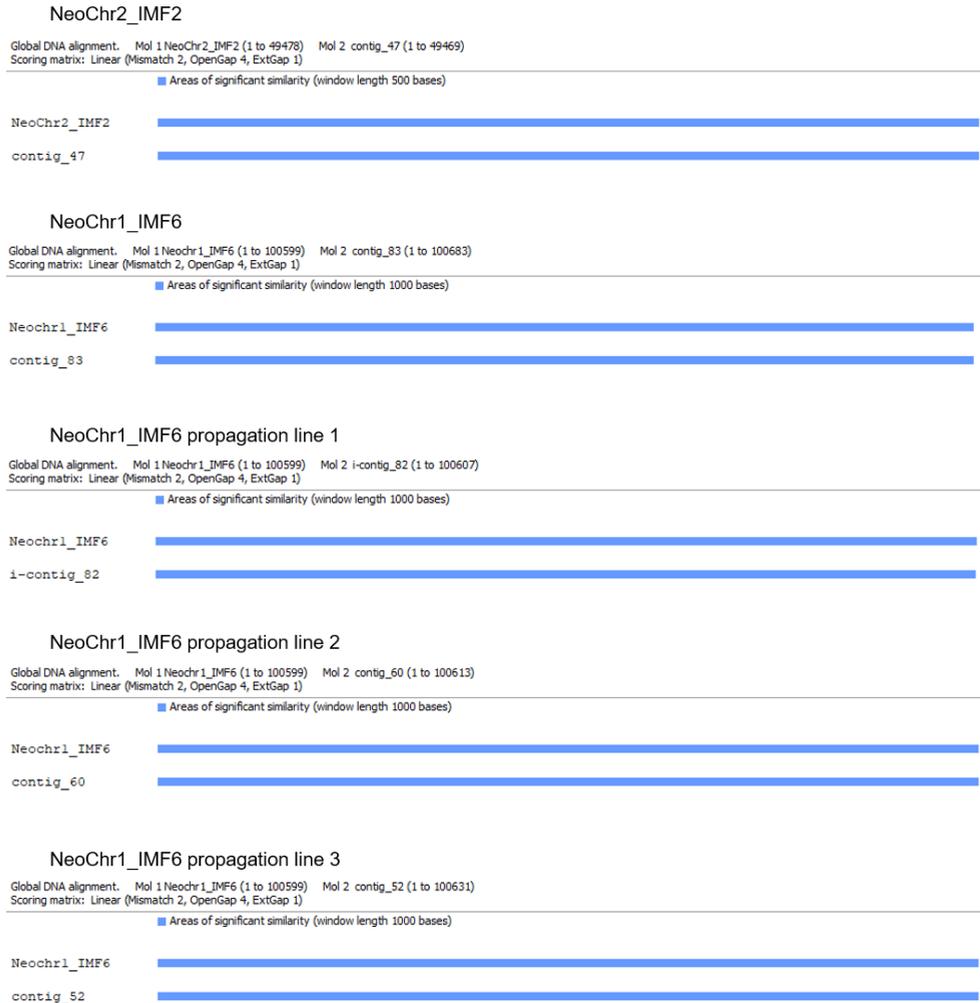
**Figure S5 CHEF IMF1, IMF2 and IMF6**

1) CEN.PK113-5D (- control), 2) IMX1338 (- control), 3) IMS0480 (+ control), 4) in plug I-SceI digested IMF1, 5) in plug I-SceI digested IMF2, 6) in plug I-SceI digested IMF1, 7) undigested transformant 3 (IMF6), 8) in plug I-SceI digested transformant 3 (IMF6), 9) undigested transformant 4, 10) in plug I-SceI digested transformant 4., 11) Lambda PFG ladder. NeoChr1 in IMF1, shows it is approximately 135 kb, which is 35 kb larger than the *in silico* design. NeoChr2 in IMF2 has the size of 50 kb as *in silico* designed. NeoChr1 in transformant 3 and 4 both show a correct size of approximately 100 kb.



### Figure S6 Long-read nanopore sequencing of Neochromosomes

The genomic DNA of strains IMF2, IMF6, IMF11, IMF12, IMF17, IMF18 and IMF23 carrying neochromosomes was sequenced by long-read nanopore technology. The reads were assembled into contigs using flye. Using clone manager, a global alignment of the neochromosome contig against the *in silico* neochromosome design was performed. Blue represents areas of significant similarity while a black line represents areas without significant similarities. All strains sequenced with long-read sequencing revealed neochromosomes corresponding to their *in silico* design without duplications or rearrangements.



### NeoChr19\_IMF11

Global DNA alignment. Mol 1 NeoChr19\_IMF11 (1 to 135208) Mol 2 contig\_49 (1 to 135199)  
Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

■ Areas of significant similarity (window length 2000 bases)



### NeoChr20\_IMF12

Global DNA alignment. Mol 1 NeoChr20\_IMF12 (1 to 84211) Mol 2 contig\_53 (1 to 84200)  
Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

■ Areas of significant similarity (window length 1000 bases)



### NeoChr15\_IMF17

Global DNA alignment. Mol 1 NeoChr15\_IMF17 (1 to 135208) Mol 2 contig\_53 (1 to 135152)  
Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

■ Areas of significant similarity (window length 2000 bases)



### NeoChr16\_IMF18

Global DNA alignment. Mol 1 NeoChr16\_IMF18 (1 to 84211) Mol 2 i-contig\_62 (1 to 84145)  
Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

■ Areas of significant similarity (window length 1000 bases)



### NeoChr12\_IMF23

Global DNA alignment. Mol 1 NeoChr12\_IMF23 (1 to 99228) Mol 2 contig\_87 (1 to 99673)  
Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

■ Areas of significant similarity (window length 1000 bases)



### NeoChr12\_IMF23 propagation line 1

Global DNA alignment. Mol 1 NeoChr12\_IMF23 (1 to 99228) Mol 2 contig\_50 (1 to 99224)  
Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

■ Areas of significant similarity (window length 1000 bases)



### NeoChr12\_IMF23 propagation line 2

Global DNA alignment: Mol 1 NeoChr12\_IMF23 (1 to 99228) Mol 2 i-contig\_55 (1 to 99222)  
 Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

■ Areas of significant similarity (window length 1000 bases)



### NeoChr12\_IMF23 propagation line 3

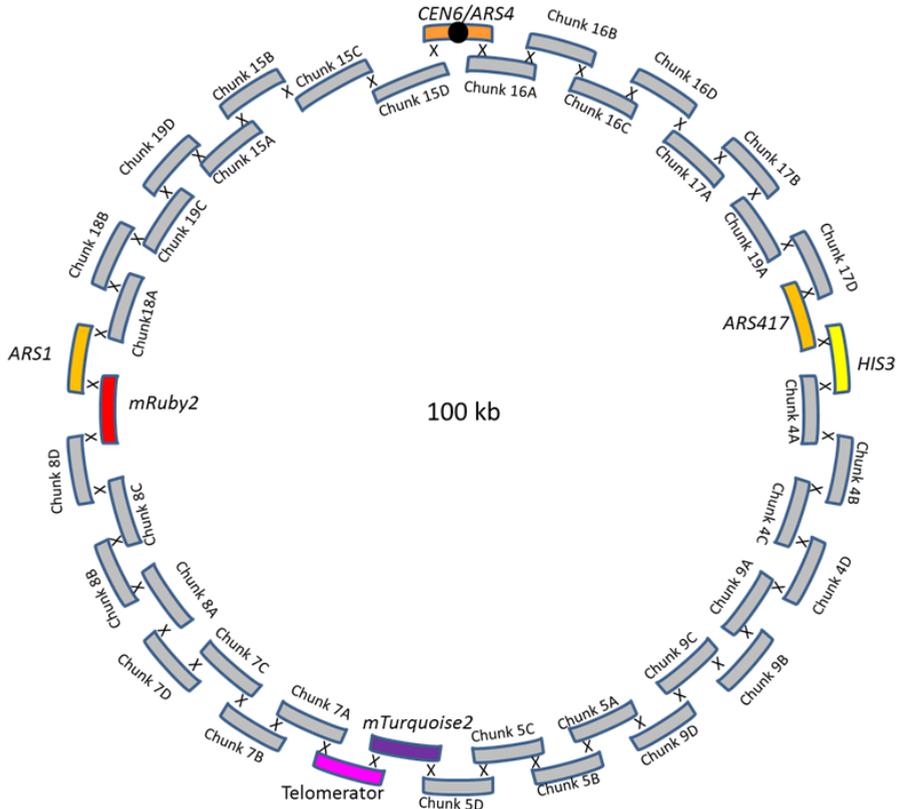
Global DNA alignment: Mol 1 NeoChr12\_IMF23 (1 to 99228) Mol 2 i-contig\_66 (1 to 99224)  
 Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

■ Areas of significant similarity (window length 1000 bases)



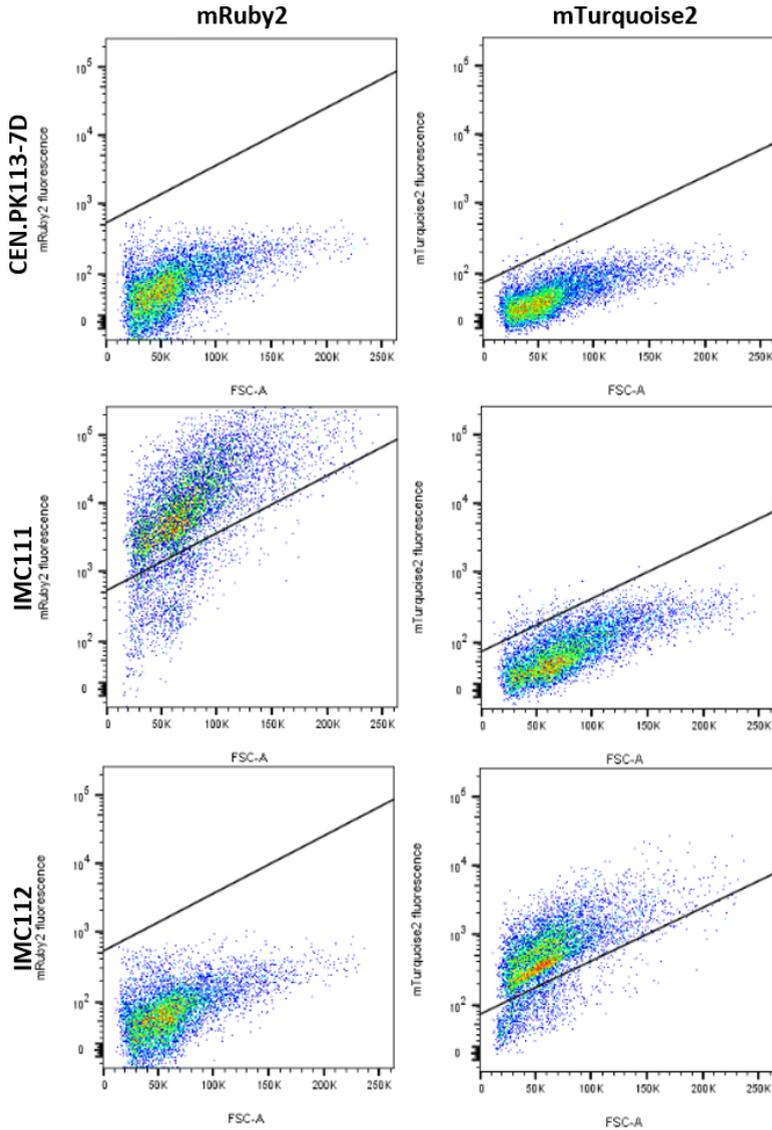
## Figure S7 NeoChr12 assembly

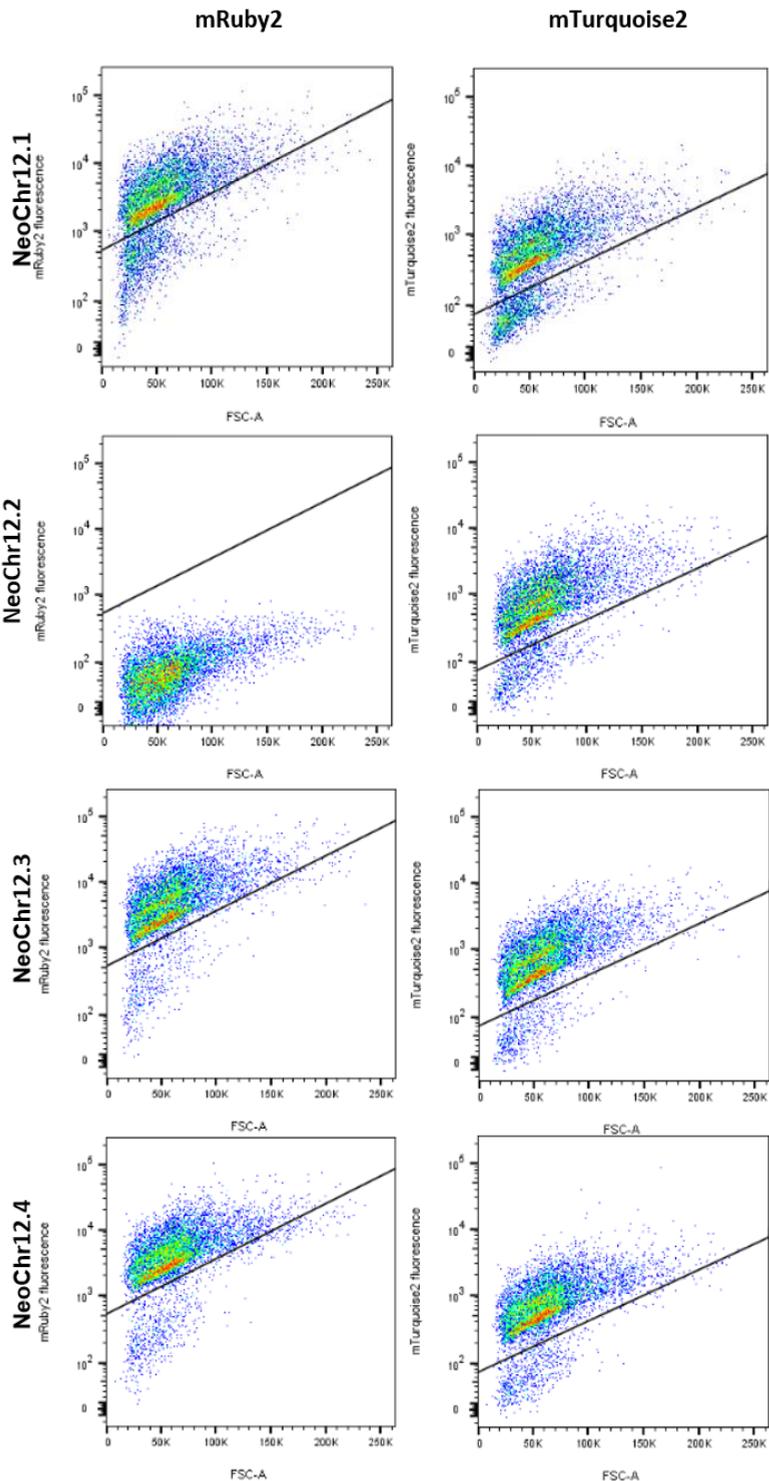
The assembly of 43 fragments into a 100 kb NeoChr12 chromosome. Chunks represent 2.5 kb non-coding *E. coli* DNA. *HIS3*, *URA3* (in telomerator), *mRuby2* and *mTurquoise2* were used as markers. *CEN6/ARS4* was included for segregation of the neochromosome and *ARS1* and *ARS417* were included to ensure correct replication.

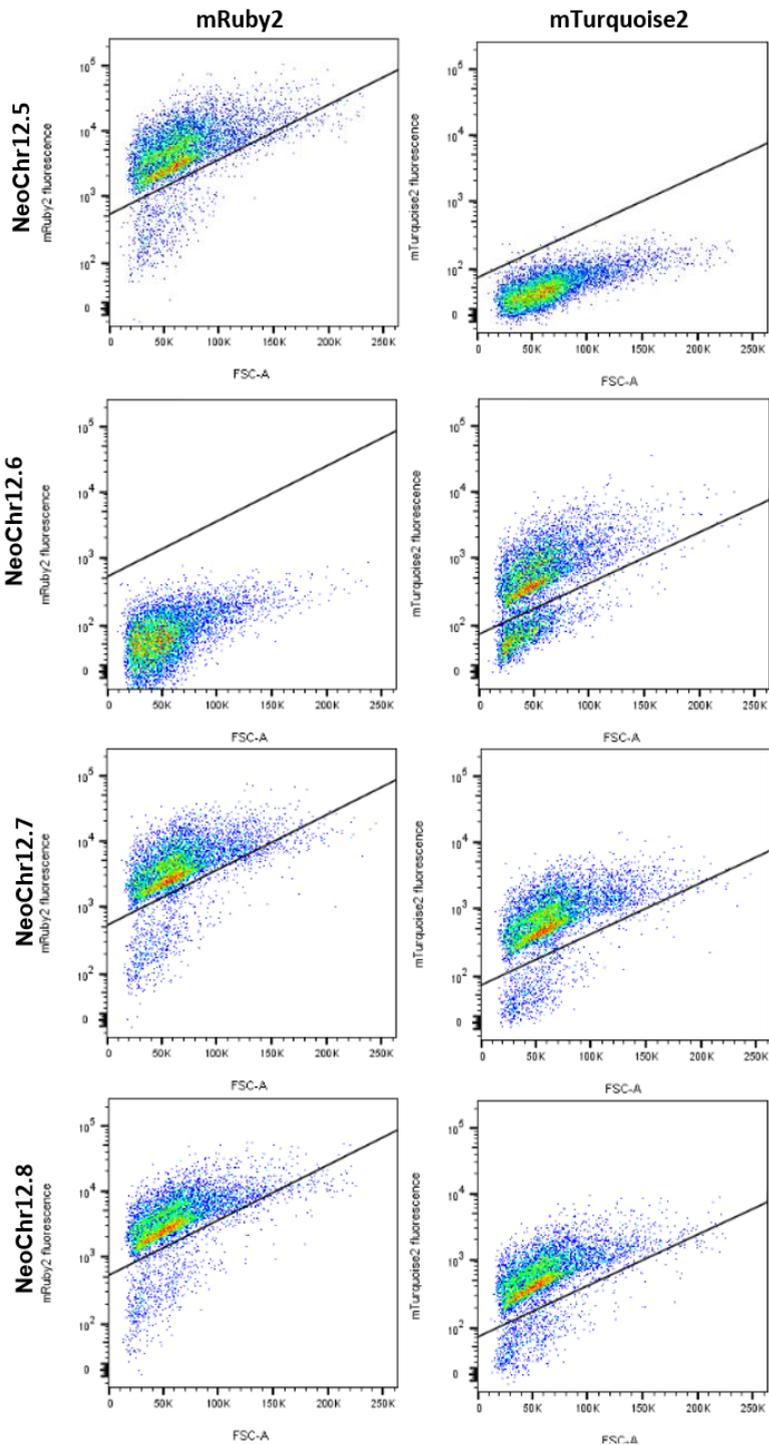


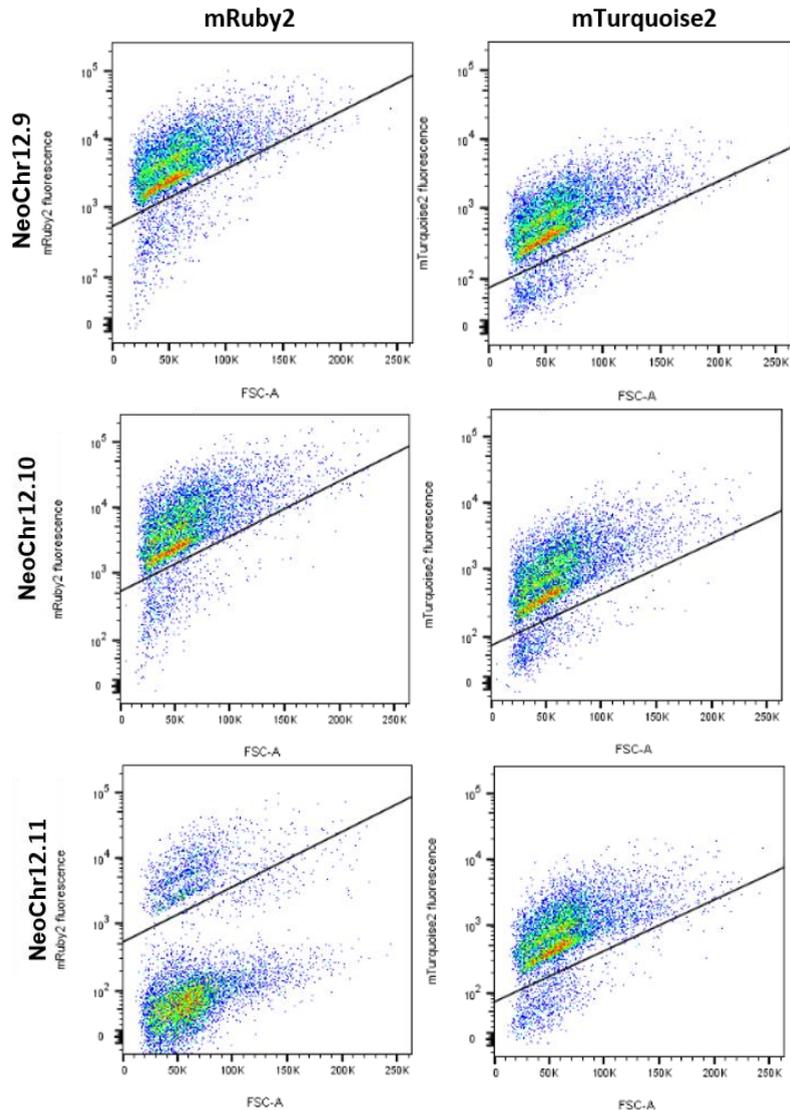
### Figure S8 FACS imaging of NeoChr12 transformants

FACS analysis of neochromosomes. The fluorescence is plotted on the y-axis and the FSC-A on the x-axis. Negative control: CEN.PK113-7D. Positive controls: IMC111(mRuby2), IMC112 (mTurquoise2). Gates for fluorescence of the two different fluorescent proteins were drawn based on the IMC111 and IMC112 controls. Double fluorescence: NeoChr12.1, NeoChr12.3, NeoChr12.4, NeoChr12.7, NeoChr12.8, NeoChr12.9, NeoChr12.10 and NeoChr12.11. mRuby2 fluorescence only: NeoChr12.5. mTurquoise2 fluorescence only: NeoChr12.2 and NeoChr12.6. Approximately 10000 events are shown.



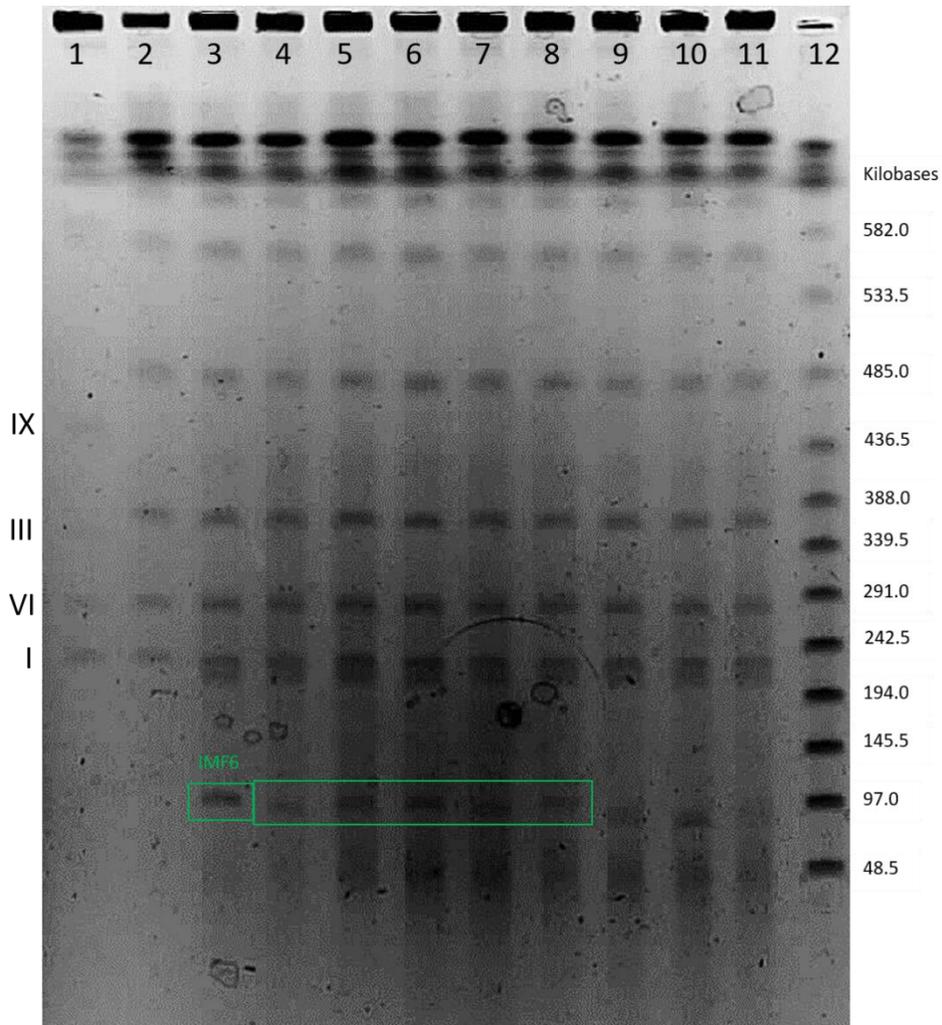






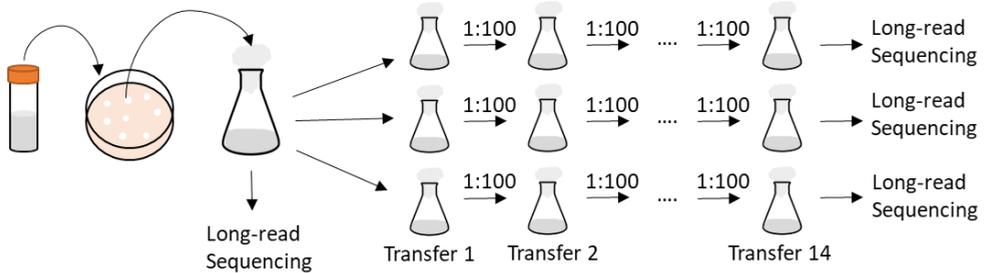
**Figure S9 CHEF analysis of NeoChr12 transformants.**

CHEF analysis of NeoChr12 transformants. 1) CEN.PK113-7D (- control), 2) IMX1338 (- control), 3) in plug I-SceI digested IMF6 (+ control) 4) in plug I-SceI digested NeoChr12.1 (IMF24), 5) in plug I-SceI digested NeoChr12.3 (IMF25), 6) in plug I-SceI digested NeoChr12.4 (IMF26), 7) in plug I-SceI digested NeoChr12.7, 8) in plug I-SceI digested NeoChr12.8 (IMF23), 9) in plug I-SceI digested NeoChr12.9, 10) in plug I-SceI digested NeoChr12.10, 11) in plug I-SceI digested NeoChr12.11, 12) Lambda PFG ladder. Transformants 1, 3, 4, 7 and 8 appear to have a neochromosome of the correct size of 100 kb. Transformants 9 and 10 have a smaller neochromosome than 100 kb. There seems to be no band for a neochromosome for transformant 11.



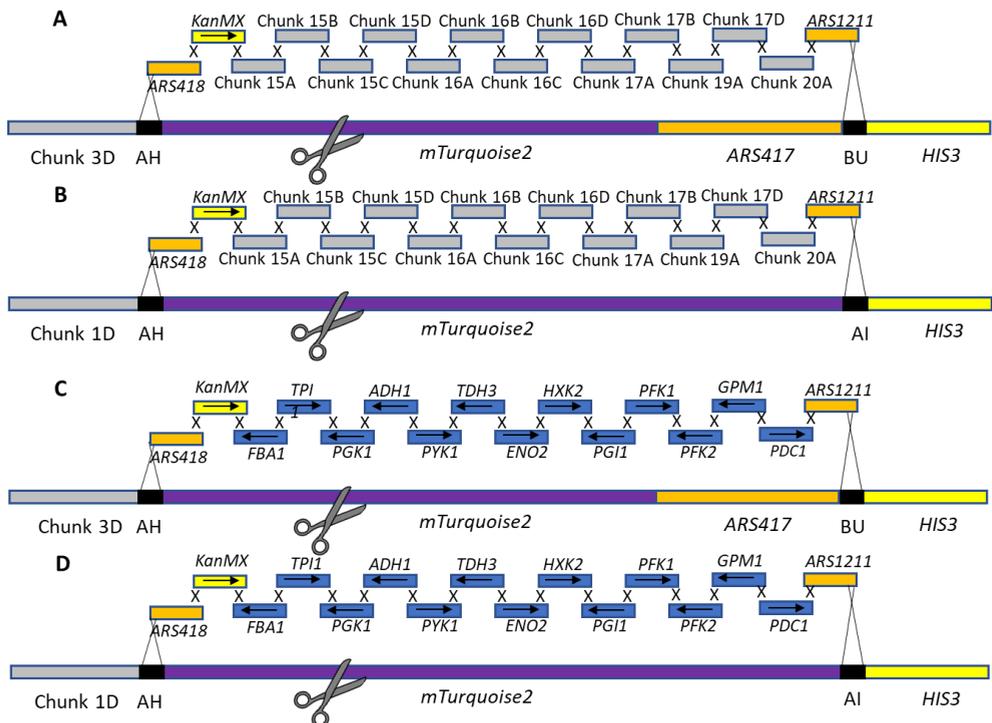
**Figure S10 Experiment to test genetic stability of neochromosomes**

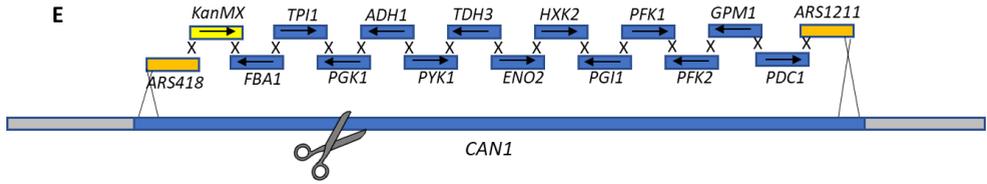
Illustration of the experiment to investigate the genetic stability of neochromosomes. From a -80°C freezer stock, single colonies were isolated on an SMD agar plate. A single colony was inoculated in liquid SMD medium. From this starter culture, genomic DNA was isolated and sequenced by long-read sequencing. The starter culture was also used to inoculate 3 individual propagation lines. Each line was transferred to fresh medium every day, for fourteen days. The last liquid cultures were used to isolate genomic DNA and perform long-read sequencing. The sequencing results of the end and starter cultures were compared to the *in silico* neochromosome designs of the strains.



**Figure S11 Genetic design to test the potential of neochromosomes as landing pads**

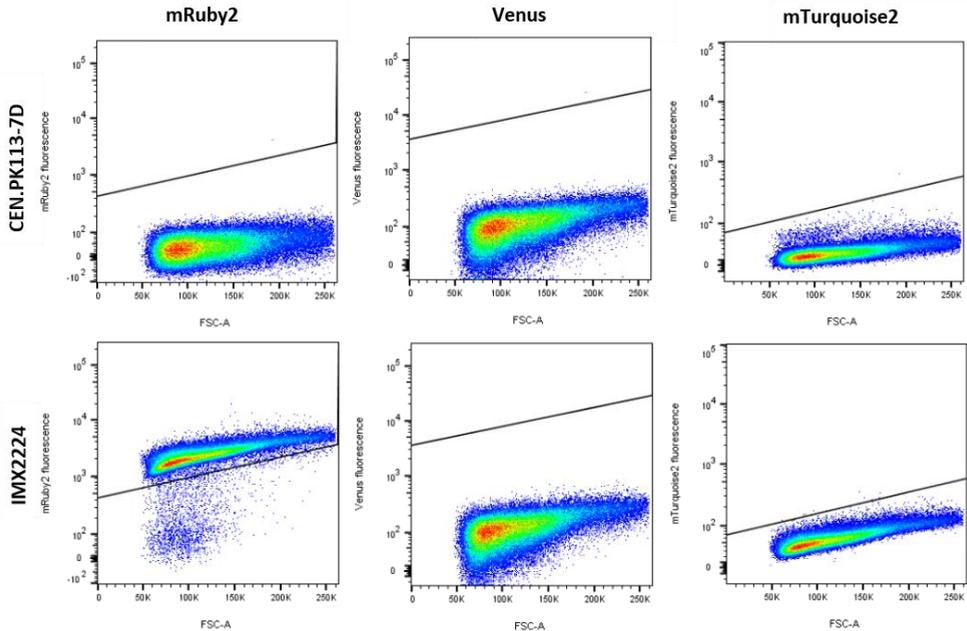
35 kb of non-coding DNA was inserted by CRISPR/Cas9 at the *mTurquoise2* locus of the 100 kb neochromosome in IMF6 (A) and of the 50 kb neochromosome in IMF2 (B) resulting in a 135 kb (IMF11) and an 85 kb (IMF12) neochromosome, respectively. 35 kb of glycolytic genes were inserted using CRISPR/Cas9 at the *mTurquoise2* locus of the 100 kb neochromosome in IMF6 (C) and of the 50 kb neochromosome of IMF2 (D) resulting in a 135 kb (IMF11) and an 85 kb (IMF12) neochromosome, respectively. As control, the same glycolytic cassettes were integrated at the native *CAN1* locus (E) resulting in strain IMX1959.

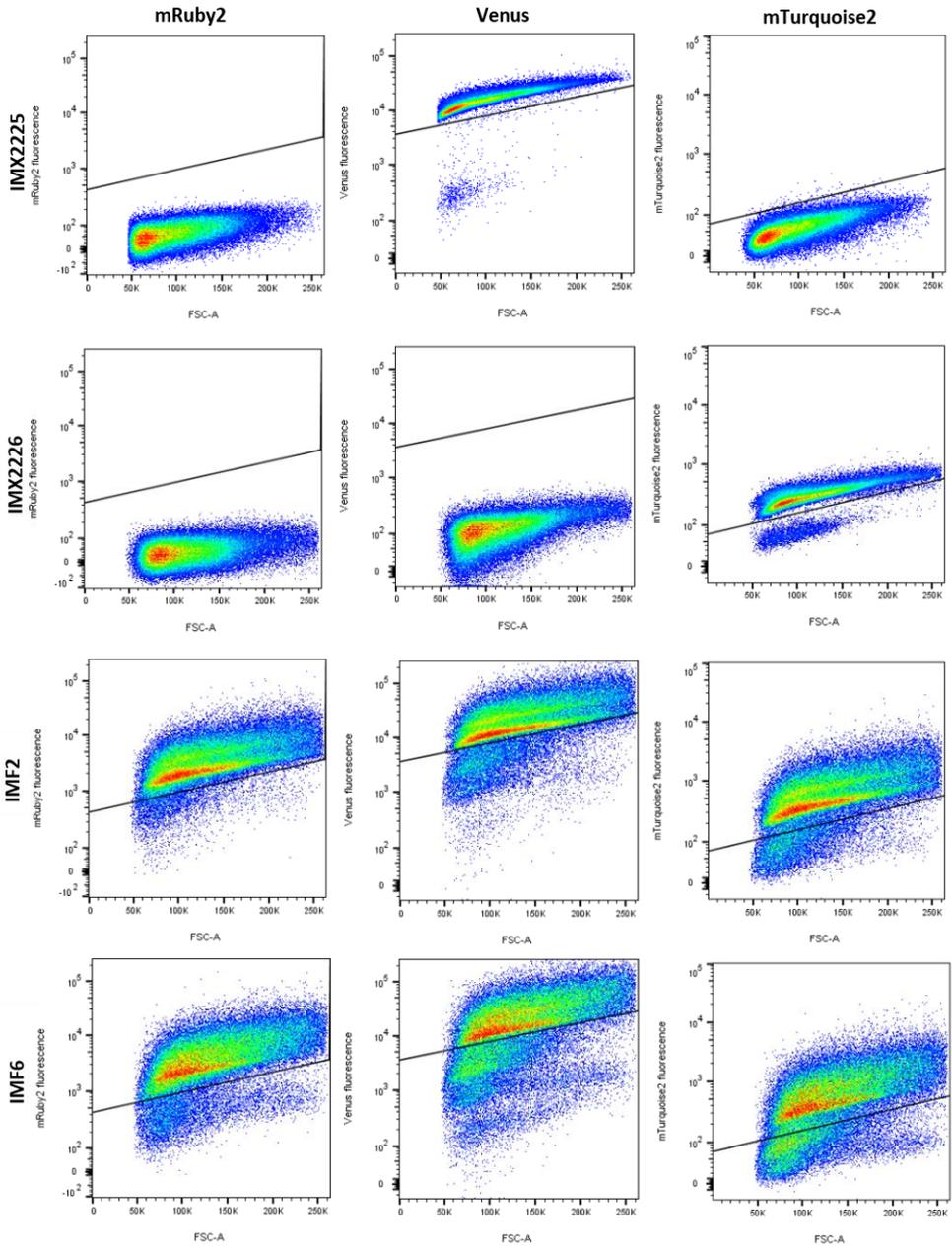


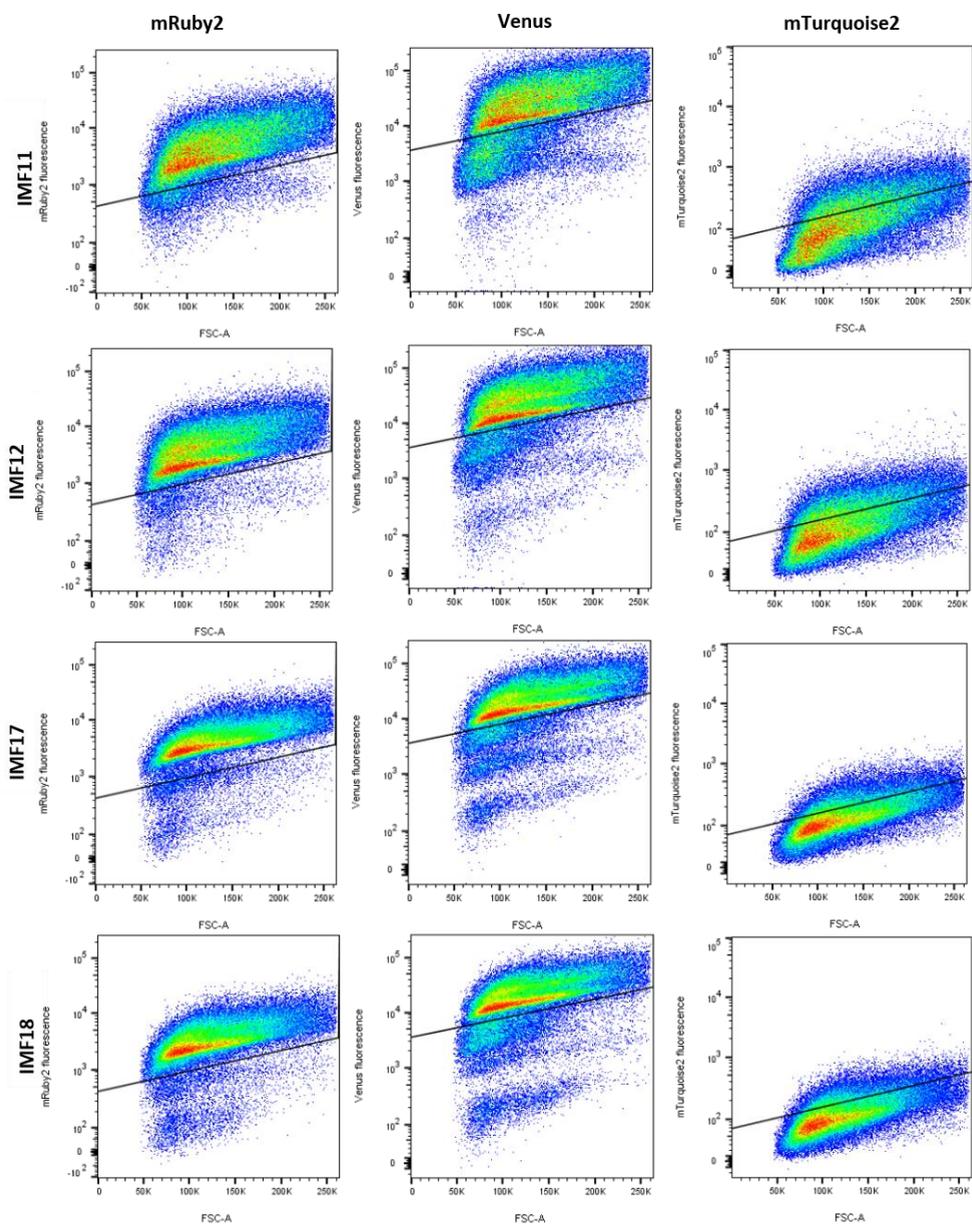


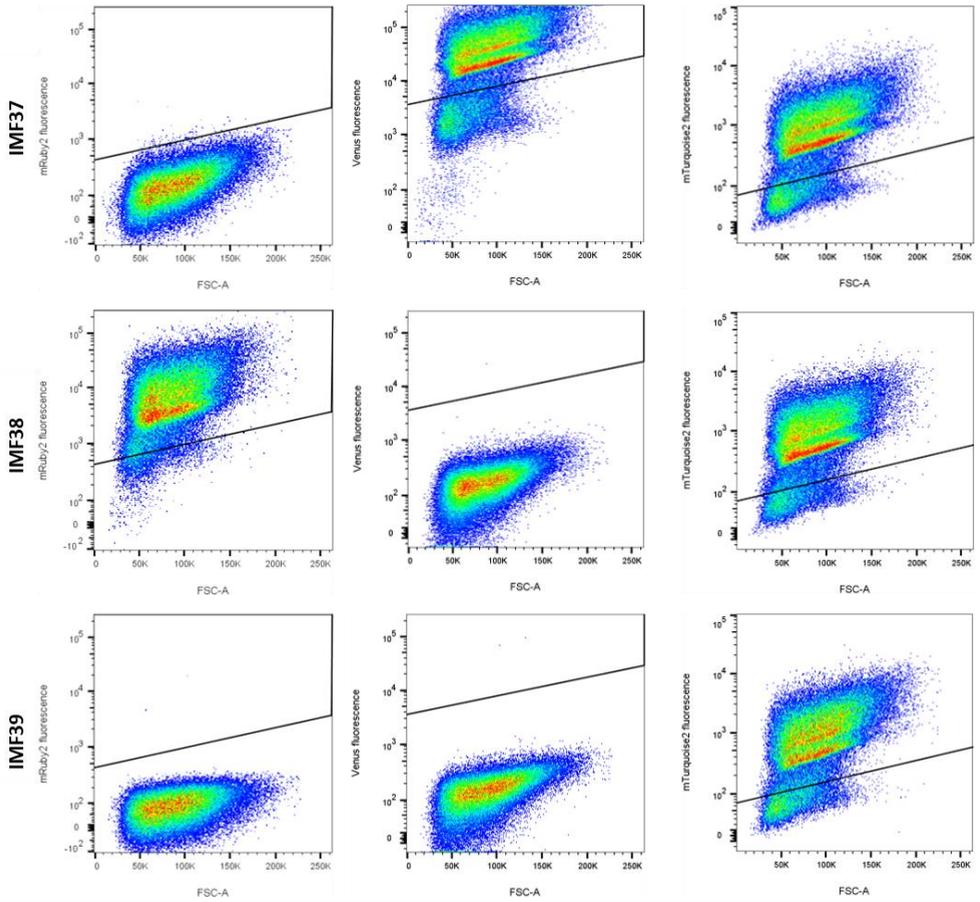
**Figure S12 FACS imaging of neochromosomes**

The fluorescence is plotted on the y-axis and the FSC-A on the x-axis. The negative control: CEN.PK113-7D. Positive controls: IMX2224 (mRuby2), IMX2225 (Venus), IMX2226 (mTurquoise2). The Venus protein showed bleed through in the mTurquoise2 channel, this was compensated based on the compensation matrix in FlowJo V10 with IMX2226 as positive signal and IMX2225 as negative signal. Gates for fluorescence of the three different fluorescent proteins were drawn based on the IMX2224, IMX2225 and IMX2226 controls. IMF2 (50 kb, empty) showed all three fluorescence. IMF6 (100 kb, empty) showed all three fluorescence. IMF11 (135 kb, empty) showed mRuby2 fluorescence and Venus fluorescence. IMF12 (85 kb, empty) showed mRuby2 fluorescence and Venus fluorescence. IMF17 (135 kb, glycolysis) showed mRuby2 fluorescence and Venus fluorescence. IMF18 (85 kb, glycolysis) showed mRuby2 fluorescence and Venus fluorescence. IMF37 showed Venus fluorescence and mTurquoise2 fluorescence. IMF38 showed mRuby2 fluorescence and mTurquoise2 fluorescence. IMF39 showed mTurquoise2 fluorescence. Approximately 100000 events are shown for each plot.



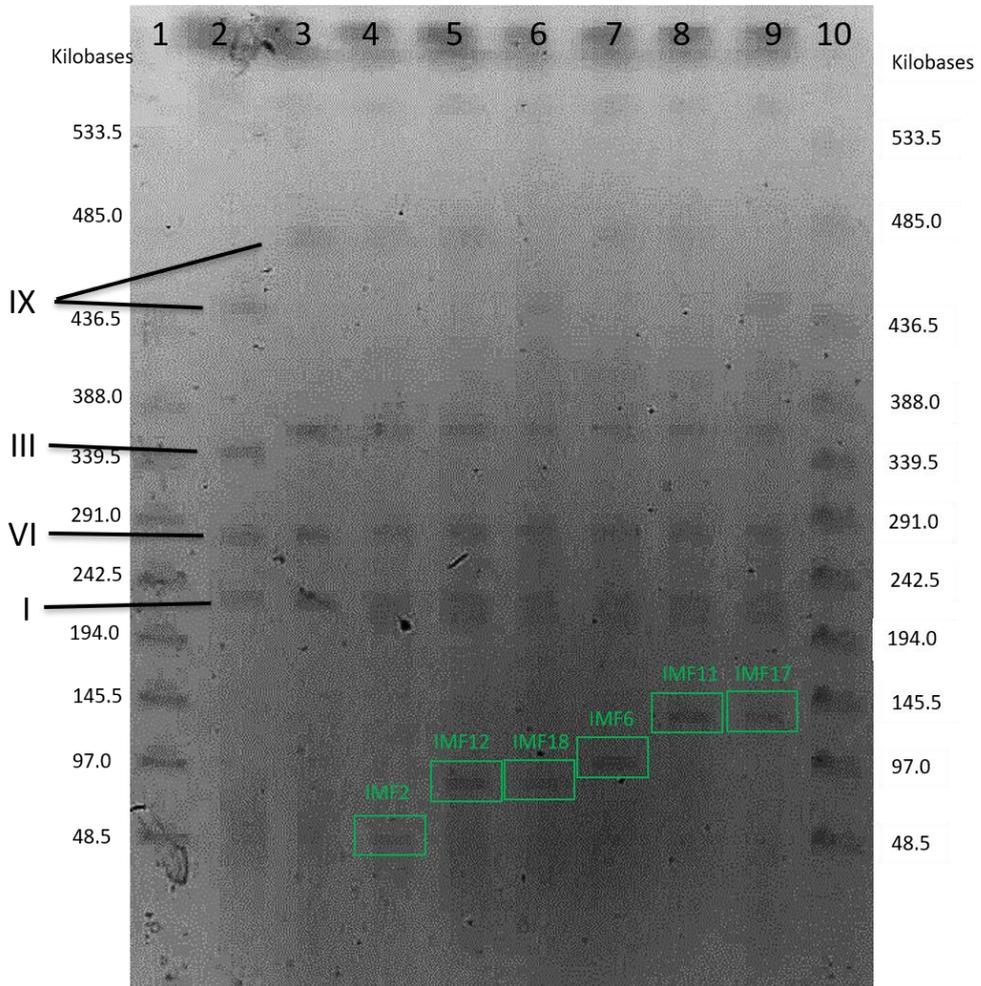






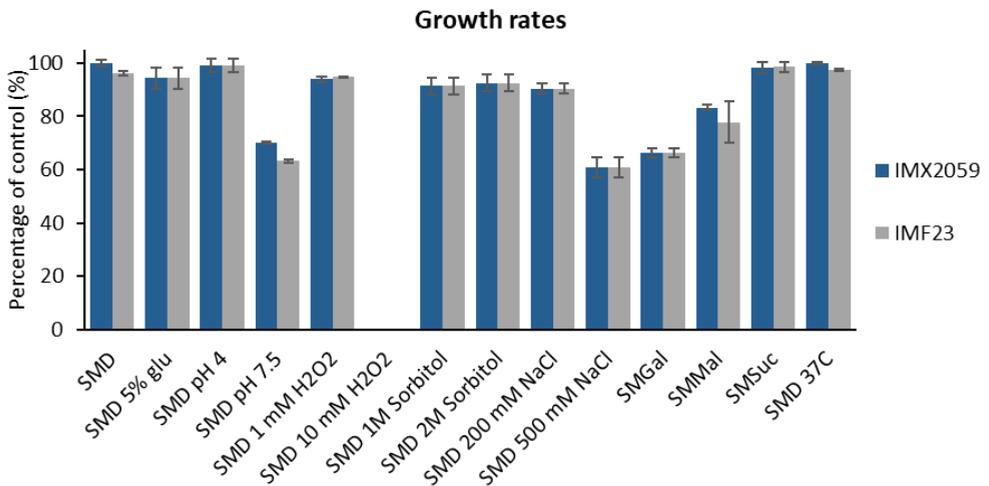
**Figure S13 CHEF analysis of IMF2, IMF12, IMF18, IMF6, IMF11, IMF17**

1) Lambda PFG ladder, 2) CEN.PK113-7D (- control), 3) IMX1338 (- control), 4) in plug I-SceI digested IMF2 (50 kb neochromosome), 5) in plug I-SceI digested IMF12 (85 kb neochromosome), 6) in plug I-SceI digested IMF18 (85 kb neochromosome containing SinLoG), 7) in plug I-SceI digested IMF6 (100 kb neochromosome), 8) in plug I-SceI digested IMF11 (135 kb neochromosome), 9) in plug I-SceI digested IMF17 (135 kb neochromosome containing SinLoG), 10) Lambda PFG ladder.



### Figure S14 Impact of neochromosomes on the physiology of the host

The growth rate of the prototrophic control strain IMX2059 without neochromosome and strain IMF23 carrying the 100 kb neochromosome with improved design was compared on different growth media and stress conditions using a growth profiler. To account for variation in growth rate among growth profiler plates, the growth rates are represented as the percentage of the growth rate of control strain IMX2059 on SMD (taken along in each plate). Bars represent the average and standard deviation of biological triplicates. All significant differences between IMX2059 and IMF23 are indicated with an asterisk (2-tailed student t-test,  $P < 0.05$ ). The strains were tested on different carbon sources (SMD (2% glucose), SMD 5% glu (5% glucose), SMGal (2% galactose), SMMal (2% maltose), SMSuc (2% sucrose)), pH stress (SMD pH4, SMD pH7.5), oxidative stress (SMD 1 mM H<sub>2</sub>O<sub>2</sub>, SMD 10 mM H<sub>2</sub>O<sub>2</sub>), osmotic stress (SMD 1M sorbitol, SMD 2M sorbitol), salt stress (SMD 200 mM NaCl, SMD 500 mM NaCl) and heat stress (SMD 37°C). IMF23 showed little or no difference in physiology compared to IMX2059.



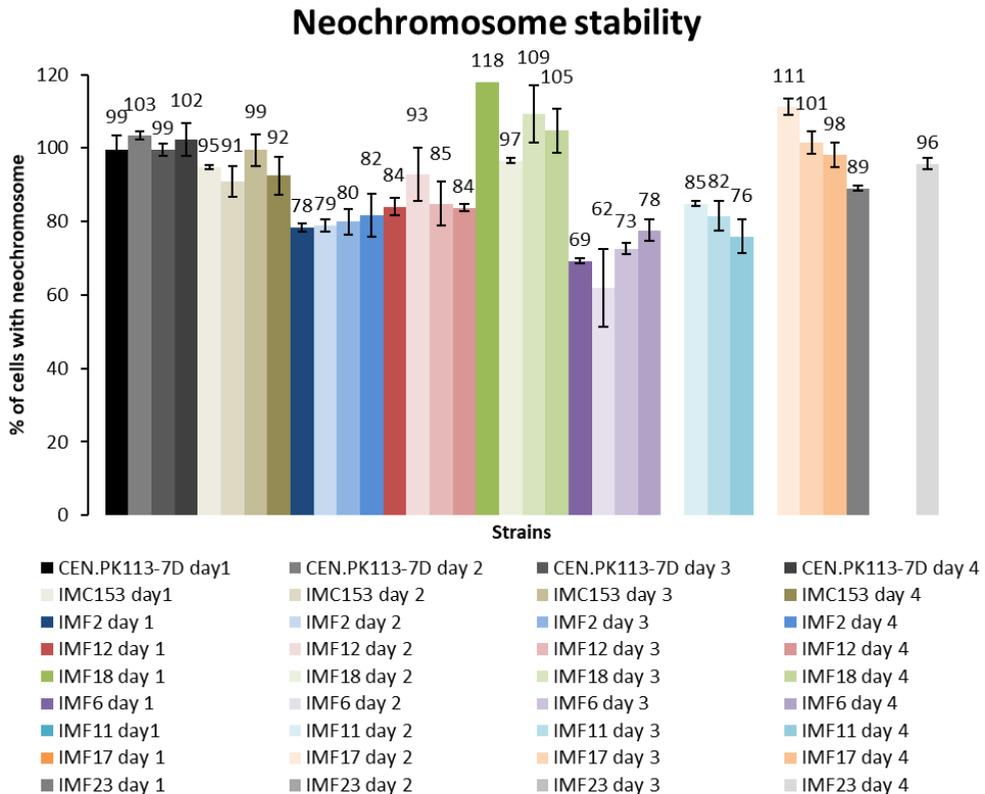
**Figure S15 Neochromosome stability upon transfer.**

The stability of the neochromosome calculated by:

$$\% \text{ of cells with NeoChr: } \frac{\text{number of colonies on SMD}}{\text{number of colonies of YPD}} \times 100\%$$

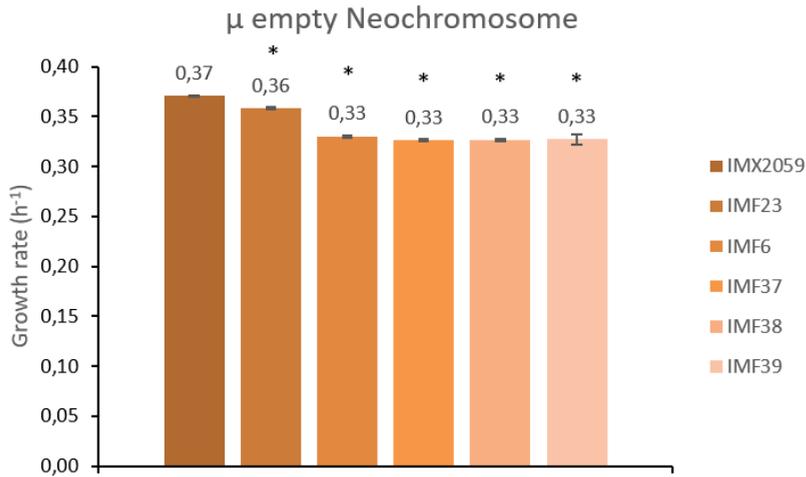
- IMF18 day 1: only in single measurements
- IMF11 day 1: no measurements
- IMF17 day 1: no measurements
- IMF23: only measurements on day 1 and 4

For CEN.PK113-7D, IMF18 (50 kb, glycolysis), IMF17 (50 kb, glycolysis) the percentage is approximately 100% for all measurements. This is due to the fact that SMD and YPD medium are both selective media with respect to the (synthetic) chromosome(s) that they carry. For IMF23 (100 kb empty, improved design) and IMC153 (6.5 kb plasmid) the percentage of cells with the neochromosome/plasmid is high on all days meaning it is very stable. For IMF2 (50 kb, empty), IMF12 (85 kb, empty), IMF6 (100 kb, empty) and IMF11 (135 kb, empty) the percentage of cells with neochromosome was lower on all days tested (around 80 %), revealing reduced stability of the respective neochromosomes.



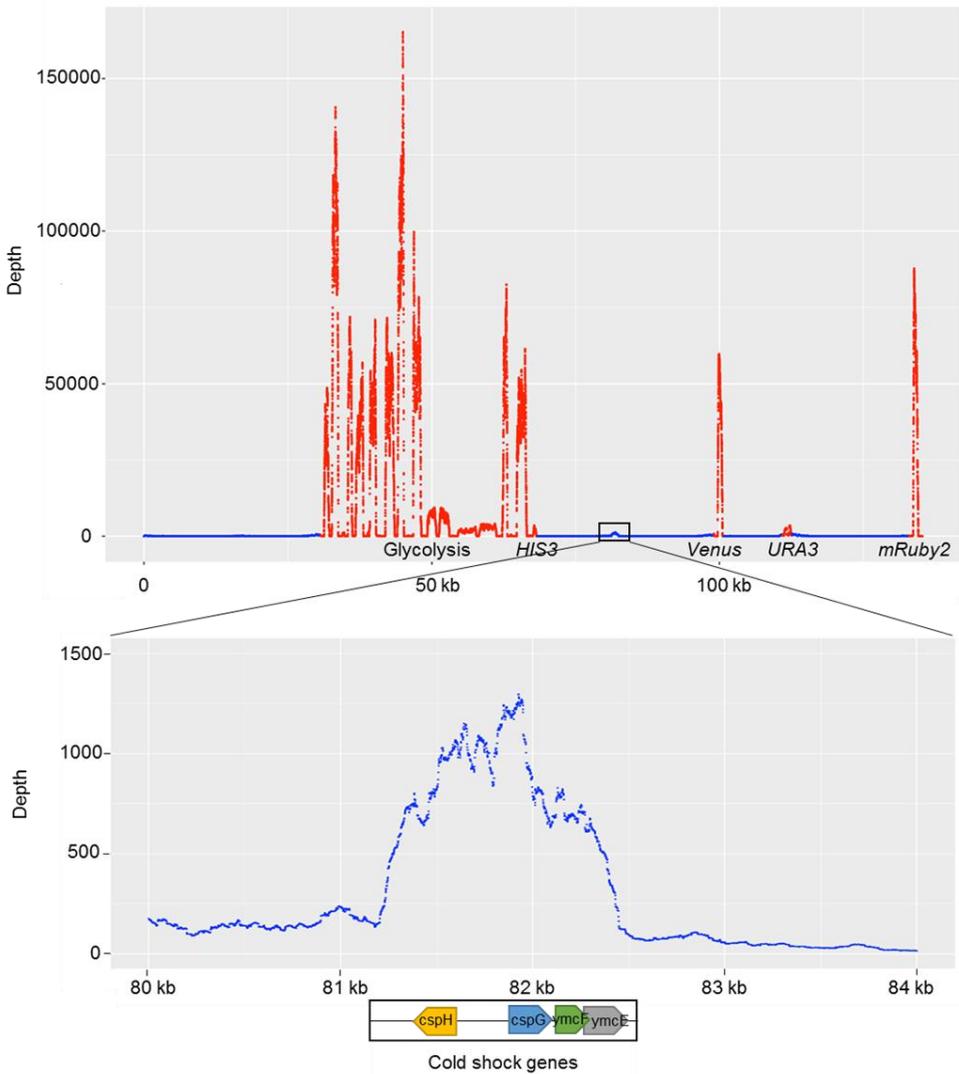
### Figure S16 Growth rate of IMF37, IMF38 and IMF39

Growth rate of strains carrying 100 kb synthetic neochromosomes: IMF23 (100 kb improved design), IMF6 (100 kb initial design), IMF37 (100 kb initial design, *mRuby2* deleted), IMF38 (100 kb initial design, *Venus* deleted), IMF39 (100 kb, *mRuby2* and *Venus* deleted). Growth rates represent the average and standard deviation of biological duplicates. All significant differences with respect to the control strain IMX2059 are indicated with an asterisk (one-way ANOVA with Post-Hoc Tukey-Kramer,  $P < 0.05$ ).



**Figure S17 RNA sequencing: evaluation of transcription from *E. coli* DNA**

In the top graph the expression along the 135 kb neochromosome of IMF17 is represented. Each data point represents the average raw coverage of the three biological triplicates at that particular point on the neochromosome. Blue depicts *E. coli* template, while red represent *S. cerevisiae* template or genes that should express in *S. cerevisiae* (*Venus* and *mRuby2*). There are four small defined areas where there is above threshold expression from *E. coli* DNA template. The bottom graph is a zoom-in of the most significant expression from *E. coli* DNA template and the *E. coli* genes which are located in this area. Apart from this few defined areas with low expression, there is no expression from *E. coli* DNA.



## References

1. Walker, G.M. and Walker, R.S.K. (2018) Chapter Three - Enhancing Yeast Alcoholic Fermentations. In: Gadd, G. M. and Sariaslani, S. (eds.), *Advances in Applied Microbiology*. Academic Press, Vol. **105**, pp. 87-129.
2. Meehl, M.A. and Stadheim, T.A. (2014) Biopharmaceutical discovery and production in yeast. *Curr. Opin. Biotechnol.*, **30**, 120-127.
3. Chubukov, V., Mukhopadhyay, A., Petzold, C.J., Keasling, J.D. and Martín, H.G. (2016) Synthetic and systems biology for microbial production of commodity chemicals. *NPJ Syst. Biol. Appl.*, **2**, 16009.
4. Li, Y., Li, S., Thodey, K., Trenchard, I., Cravens, A. and Smolke, C.D. (2018) Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E3922-E3931.
5. Paddon, C.J. and Keasling, J.D. (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.*, **12**, 355-367.
6. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L. *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253.
7. Veneziano, R., Shepherd, T.R., Ratanalert, S., Bellou, L., Tao, C. and Bathe, M. (2018) *In vitro* synthesis of gene-length single-stranded DNA. *Sci. Rep.*, **8**, 6548.
8. Hughes, R.A. and Ellington, A.D. (2017) Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.*, **9**, a023812.
9. Kosuri, S. and Church, G.M. (2014) Large-scale *de novo* DNA synthesis: technologies and applications. *Nat. Methods*, **11**, 499-507.
10. Casini, A., Storch, M., Baldwin, G.S. and Ellis, T. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.*, **16**, 568-576.
11. Kok, S.d., Stanton, L.H., Slaby, T., Durot, M., Holmes, V.F., Patel, K.G., Platt, D., Shapland, E.B., Serber, Z., Dean, J. *et al.* (2014) Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. *ACS Synth. Biol.*, **3**, 97-106.
12. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215-1220.
13. Gibson, D.G. (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.*, **37**, 6984-6990.
14. Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O. and Hutchison, C.A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 20404-20409.
15. Bruschi, C., Gjuracic, K. and Tosato, V. (2006) Yeast Artificial Chromosomes. In: *Encyclopedia of Life Sciences*. John Wiley & Sons Ltd, Chichester.
16. Kuijpers, N.G., Solis-Escalante, D., Luttki, M.A., Bisschops, M.M., Boonekamp, F.J., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2016) Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 15060-15065.
17. Solis-Escalante, D., Kuijpers, N.G., Barrajon-Simancas, N., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2015) A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell*, **14**, 804-816.
18. Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y. *et al.* (2017) Design of a synthetic yeast genome. *Science*, **355**, 1040-1044.

19. Rideau, F., Le Roy, C., Descamps, E.C.T., Renaudin, H., Lartigue, C. and Bébéar, C. (2017) Cloning, Stability, and Modification of *Mycoplasma hominis* Genome in Yeast. *ACS Synth. Biol.*, **6**, 891-901.
20. Kuijpers, N.G., Solis-Escalante, D., Bosman, L., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb. Cell Fact.*, **12**, 47.
21. Dhar, M.K., Sehgal, S. and Kaul, S. (2012) Structure, replication efficiency and fragility of yeast ARS elements. *Res. Microbiol.*, **163**, 243-253.
22. Kanhere, A. and Bansal, M. (2005) Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. *Nucleic Acids Res.*, **33**, 3165-3175.
23. Griesenbeck, J., Tschochner, H. and Grohmann, D. (2017) Structure and Function of RNA Polymerases and the Transcription Machineries. In: Harris, J. R. and Marles-Wright, J. (eds.), *Macromolecular Protein Complexes: Structure and Function*. Springer International Publishing, Cham, pp. 225-270.
24. 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.
25. Mitchell, L.A. and Boeke, J.D. (2014) Circular permutation of a synthetic eukaryotic chromosome with the telomerase. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 17003-17010.
26. Lambert, T.J. (2019) FPbase: a community-editable fluorescent protein database. *Nat. Methods*, **16**, 277-278.
27. Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K. and Miyawaki, A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biol.*, **20**, 87.
28. Goedhart, J., von Stetten, D., Noirclerc-Savoye, M., Lelimosin, M., Joosen, L., Hink, M.A., van Weeren, L., Gadella Jr, T.W.J. and Royant, A. (2012) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.*, **3**, 751.
29. Shao, Y., Lu, N., Wu, Z., Cai, C., Wang, S., Zhang, L.-L., Zhou, F., Xiao, S., Liu, L., Zeng, X. et al. (2018) Creating a functional single-chromosome yeast. *Nature*, **560**, 331-335.
30. Luo, J., Sun, X., Cormack, B.P. and Boeke, J.D. (2018) Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature*, **560**, 392-396.
31. Gorter de Vries, A.R., Pronk, J.T. and Daran, J.G. (2017) Industrial relevance of chromosomal copy number variation in *Saccharomyces* Yeasts. *Appl. Environ. Microbiol.*, **83**, 1-15.
32. Murray, A.W., Schultes, N.P. and Szostak, J.W. (1986) Chromosome length controls mitotic chromosome segregation in yeast. *Cell*, **45**, 529-536.
33. Surosky, R.T., Newlon, C.S. and Tye, B.-K. (1986) The mitotic stability of deletion derivatives of chromosome III in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 414-418.
34. Sleister, H.M., Mills, K.A., Blackwell, S.E., Killary, A.M., Murray, J.C. and Malone, R.E. (1992) Construction of a human chromosome 4 YAC pool and analysis of artificial chromosome stability. *Nucleic Acids Res.*, **20**, 3419-3425.
35. Hieter, P., Mann, C., Snyder, M. and Davis, R.W. (1985) Mitotic stability of yeast chromosomes: a colony color assay that measures nondisjunction and chromosome loss. *Cell*, **40**, 381-392.
36. Agmon, N., Temple, J., Tang, Z., Schraink, T., Baron, M., Chen, J., Mita, P., Martin, J.A., Tu, B.P., Yanai, I. et al. (2019) Phylogenetic debugging of a complete human biosynthetic pathway transplanted into yeast. *Nucleic Acids Res.*, **48**, 486-499.
37. Hauf, J., Zimmermann, F.K. and Muller, S. (2000) Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast *Saccharomyces cerevisiae*. *Enzyme Microb. Technol.*, **26**, 688-698.
38. Schaaff, I., Heinisch, J. and Zimmermann, F.K. (1989) Overproduction of glycolytic enzymes in yeast. *Yeast*, **5**, 285-290.

39. Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J.A., Ou, J., Moffat, J., Boone, C. *et al.* (2012) Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nat. Cell Biol.*, **14**, 966-976.
40. Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., Richardson, S.M., Dymond, J.S., Kuang, Z., Scheifele, L.Z., Cooper, E.M. *et al.* (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science*, **344**, 55-58.
41. Aylon, Y. and Kupiec, M. (2004) DSB repair: the yeast paradigm. *DNA Repair (Amst.)*, **3**, 797-815.
42. Heyer, W.D., Ehmsen, K.T. and Liu, J. (2010) Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.*, **44**, 113-139.
43. Barnes, G. and Rio, D. (1997) DNA double-strand-break sensitivity, DNA replication, and cell cycle arrest phenotypes of Ku-deficient *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 867-872.
44. Laroche, T., Martin, S.G., Gotta, M., Gorham, H.C., Pryde, F.E., Louis, E.J. and Gasser, S.M. (1998) Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Curr. Biol.*, **8**, 653-656.
45. Gravel, S., Larrivee, M., Labrecque, P. and Wellinger, R.J. (1998) Yeast Ku as a regulator of chromosomal DNA end structure. *Science*, **280**, 741-744.
46. Benders, G.A., Noskov, V.N., Denisova, E.A., Lartigue, C., Gibson, D.G., Assad-Garcia, N., Chuang, R.-Y., Carrera, W., Moodie, M., Algire, M.A. *et al.* (2010) Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.*, **38**, 2558-2569.
47. Tagwerker, C., Dupont, C.L., Karas, B.J., Ma, L., Chuang, R.-Y., Benders, G.A., Ramon, A., Novotny, M., Montague, M.G., Venepally, P. *et al.* (2012) Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res.*, **40**, 10375-10383.
48. Karas, B.J., Molparia, B., Jablanovic, J., Hermann, W.J., Lin, Y.C., Dupont, C.L., Tagwerker, C., Yonemoto, I.T., Noskov, V.N., Chuang, R.Y. *et al.* (2013) Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng.*, **7**, 30.
49. Venetz, J.E., Del Medico, L., Wölflé, A., Schächle, P., Bucher, Y., Appert, D., Tschan, F., Flores-Tinoco, C.E., van Kooten, M., Guennoun, R. *et al.* (2019) Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8070-8079.
50. DiCarlo, J.E., Conley, A.J., Penttila, M., Jantti, J., Wang, H.H. and Church, G.M. (2013) Yeast oligo-mediated genome engineering (YOGE). *ACS Synth. Biol.*, **2**, 741-749.
51. Yu, S.C., Kuemmel, F., Skoufou-Papoutsaki, M.N. and Spanu, P.D. (2019) Yeast transformation efficiency is enhanced by TORC1- and eisosome-dependent signaling. *MicrobiologyOpen*, **8**, e00730.
52. Hill, A. and Bloom, K. (1987) Genetic manipulation of centromere function. *Mol. Cell Biol.*, **7**, 2397-2405.
53. Walker, R.S.K. (2017) *De novo* biological engineering of a tRNA neochromosome in yeast. <https://era.ed.ac.uk/handle/1842/28921>.
54. Lacefield, S., Lau, D.T. and Murray, A.W. (2009) Recruiting a microtubule-binding complex to DNA directs chromosome segregation in budding yeast. *Nat. Cell Biol.*, **11**, 1116-1120.
55. Naesby, M., Nielsen, S.V., Nielsen, C.A., Green, T., Tange, T.O., Simon, E., Knechtle, P., Hansson, A., Schwab, M.S., Titiz, O. *et al.* (2009) Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. *Microb. Cell Fact.*, **8**, 45.
56. Klein, J., Heal, J.R., Hamilton, W.D., Bousseghoune, T., Tange, T.O., Delegrange, F., Jaeschke, G., Hatsch, A. and Heim, J. (2014) Yeast synthetic biology platform generates novel chemical structures as scaffolds for drug discovery. *ACS Synth. Biol.*, **3**, 314-323.

57. Essani, K., Glieder, A. and Geier, M. (2015) Combinatorial pathway assembly in yeast. Vol. 2, pp. 423-436.
58. Hughes, S.R., Cox, E.J., Bang, S.S., Pinkelman, R.J., López-Núñez, J.C., Saha, B.C., Qureshi, N., Gibbons, W.R., Fry, M.R. and Moser, B.R. (2015) Process for assembly and transformation into *Saccharomyces cerevisiae* of a synthetic yeast artificial chromosome containing a multigene cassette to express enzymes that enhance xylose utilization designed for an automated platform. *J. Lab. Autom.*, **20**, 621-635.
59. Hamperl, S. and Cimprich, K.A. (2016) Conflict resolution in the genome: how transcription and replication make it work. *Cell*, **167**, 1455-1467.
60. Entian, K.-D. and Kötter, P. (2007) 25 Yeast genetic strain and plasmid collections. In: Stansfield, I. and Stark, M. J. R. (eds.), *Methods in Microbiology*. Academic Press, Vol. **36**, pp. 629-666.
61. Verduyn, C., Postma, E., Scheffers, W.A. and Van Dijken, J.P. (1992) Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, **8**, 501-517.
62. Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.*, **350**, 87-96.
63. Looke, M., Kristjuhan, K. and Kristjuhan, A. (2011) Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques*, **50**, 325-328.
64. Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23-28.
65. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J. and Daran, J.M. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **15**, 1-15.
66. Boonekamp, F.J., Dashko, S., Duiker, D., Gehrman, T., van den Broek, M., den Ridder, M., Pabst, M., Robert, V., Abeel, T., Postma, E.D. et al. (2020) Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences. *ACS Synth. Biol.*, **9**, 1361-1375.
67. Salazar, A.N., Gorter de Vries, A.R., van den Broek, M., Wijsman, M., de la Torre Cortes, P., Brickwedde, A., Brouwers, N., Daran, J.G. and Abeel, T. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.*, **17**.
68. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.
69. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178-192.
70. Nijkamp, J.F., van den Broek, M.A., Geertman, J.M., Reinders, M.J., Daran, J.M. and de Ridder, D. (2012) *De novo* detection of copy number variation by co-assembly. *Bioinformatics*, **28**, 3195-3202.
71. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K. et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS one*, **9**, e112963.
72. Kolmogorov, M., Yuan, J., Lin, Y. and Pevzner, P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.*, **37**, 540-546.
73. Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068-2069.
74. Piper, M.D., Daran-Lapujade, P., Bro, C., Regenber, B., Knudsen, S., Nielsen, J. and Pronk, J.T. (2002) Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **277**, 37001-37008.

75. Schmitt, M.E., Brown, T.A. and Trumppower, B.L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **18**, 3091-3092.
76. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
77. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923-930.
78. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139-140.
79. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
80. Postma, E., Verduyn, C., Scheffers, W.A. and Van Dijken, J.P. (1989) Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **55**, 468-477.
81. Jansen, M.L., Diderich, J.A., Mashego, M., Hassane, A., de Winde, J.H., Daran-Lapujade, P. and Pronk, J.T. (2005) Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity. *Microbiology*, **151**, 1657-1669.
82. Cruz, L.A., Hebly, M., Duong, G.H., Wahl, S.A., Pronk, J.T., Heijnen, J.J., Daran-Lapujade, P. and van Gulik, W.M. (2012) Similar temperature dependencies of glycolytic enzymes: an evolutionary adaptation to temperature dynamics? *BMC Syst. Biol.*, **6**, 151.
83. Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951) Protein measurement with the Folin phenol reagent. *J. Biol. Chem.*, **193**, 265-275.



# Chapter 5

## Modular, synthetic chromosomes as new tools for large scale engineering of metabolism

Eline D. Postma#  
Else-Jasmijn Hassing#  
Venda Mangkusaputra  
Jordi Geelhoed  
Pilar de la Torre  
Christiaan Mooiman  
Marcel van den Broek  
Martin Pabst  
Jean-Marc Daran  
Pascale Daran-Lapujade

#Eline D. Postma and Else-Jasmijn Hassing contributed equally to this chapter and  
should be considered co-first authors.

This chapter has been submitted for publication in *Metabolic Engineering*

## Abbreviations

### Metabolites

2PE: 2-phenylethanol  
CIN: cinnamic acid  
COCOA: coumaroyl-CoA  
COUM: coumaric acid  
DHK: dihydrokaempferol  
K3G: kaempferol 3-O-glucoside  
KAE: kaempferol  
LPE: leucopelargonidin  
NAR: naringenin  
NARCC: naringenin chalcone  
P3G: pelargonidin 3-O-glucoside  
PAA: phenylacetic acid  
PEL: pelargonidin  
PHE: L-phenylalanine  
PHLOR: phloretic acid  
*p*OHPPY: *p*-hydroxyphenylpyruvate  
*p*OH2PE: *p*-hydroxyphenylethanol  
*p*OH PAA: *p*-hydroxyphenylacetic acid  
PPY: phenylpyruvate  
TYR: tyrosine

### Enzymatic reaction

3GT: anthocyanin 3-O-glucosyltransferase  
ANS: anthocyanidin synthase  
C4H: cinnamate 4-hydroxylase  
4CL: 4-coumarate CoA ligase  
CHI: chalcone isomerase  
CHS: chalcone synthase  
CPR: cytochrome P450 reductase  
DFR: dihydroflavonol 4-reductase  
F3H: flavanone 3-hydroxylase  
PAL: phenylalanine ammonia lyase  
TAL: tyrosine ammonia lyase

## Abstract

The construction of powerful cell factories requires intensive genetic engineering for the addition of new functionalities and the remodeling of native pathways and processes. The present study, using *Saccharomyces cerevisiae*, explores the potential of modular, specialized *de novo*-assembled neochromosomes (named NeoChrs) to facilitate extensive genome reprogramming. Linear and circular NeoChrs carrying 20 native and 21 heterologous genes were designed and constructed by *in vivo* assembly in yeast in two transformation steps leading to *de novo* production of anthocyanins, native to plants. Turned into exclusive expression platforms for heterologous and essential metabolic routes, the NeoChrs behaved like native chromosomes regarding mitotic and genetic stability, copy number, harmlessness for the host and editability by CRISPR-Cas9. This study paves the way for future microbial cell factories with modular genomes in which core metabolic networks, localized on satellite, specialized NeoChrs can be swapped for alternative configurations and serve as landing pads for addition of functionalities.

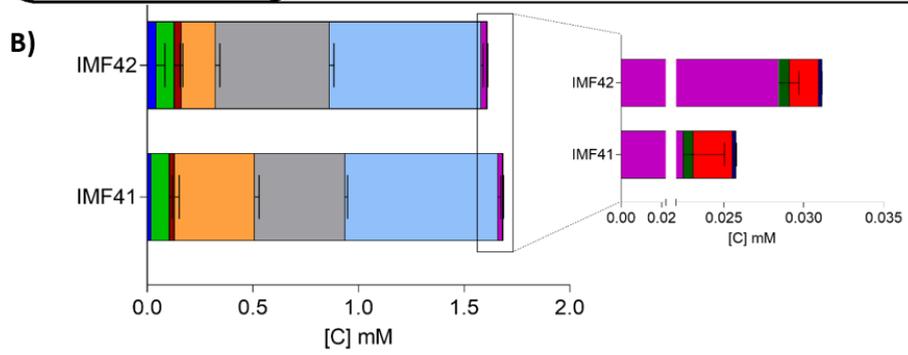
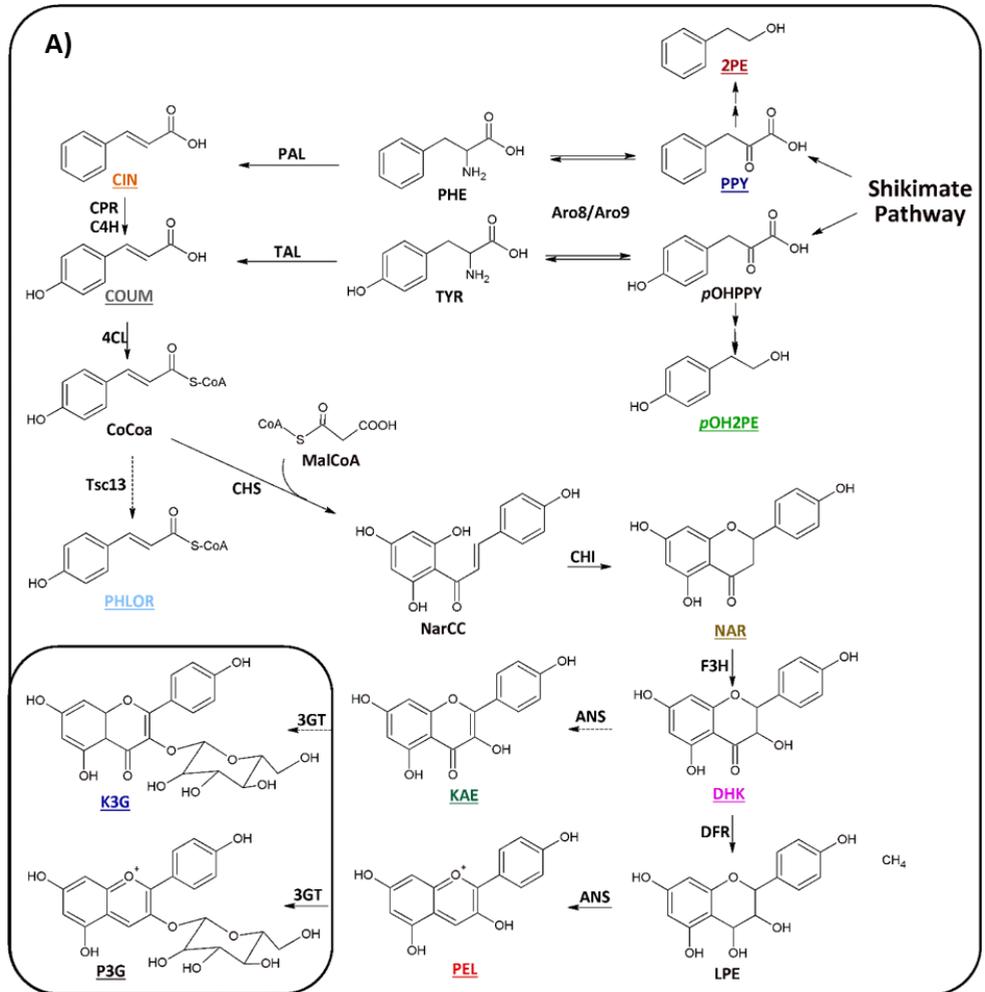
## Introduction

While microbial cell factories have a great potential for the sustainable production of fuels, chemicals and therapeutics, microbial processes are often less economically attractive than chemical, oil-based processes (1,2). Increasing the cost efficiency of microbial cell factories requires high product titer, rate and yield, features only attainable by intensive genetic engineering of the microbial host via costly strain construction programs. These programs focus on the transplantation of new functionalities, as well as on the reprogramming of the microbial host metabolic networks in which the new functionalities are plugged. These networks of biochemical reactions are encoded by several hundreds of genes scattered over large mosaic microbial genomes. Even considering the CRISPR/Cas9 revolution, rewiring these biochemical networks remains a daunting challenge, in particular for eukaryotic cell factories characterized by a high degree of genetic redundancy (3). Microbial platforms in which core biochemical networks can be remodeled at will have a great role to play in the construction of powerful cell factories, but also to reach a deep fundamental understanding of these biochemical networks and their regulation.

The pathway swapping concept was developed to address this persistent challenge (4). This modular concept is based on the genetic reduction and clustering of genes belonging to a metabolic pathway, so the pathway can be easily swapped by any other design. This concept was demonstrated in the model, and industrial yeast *Saccharomyces cerevisiae*, using glycolysis and alcoholic fermentation, a 12-step pathway catalyzed by a set of 26 enzymes, as an example. In the SwYG (Switchable Yeast Glycolysis) strain this set of 26 genes was reduced to 13 and relocalized to a single chromosomal locus, enabling the facile remodeling of the entire pathway (4). Scaling up pathway swapping to the set of core biochemical reactions required for most biotechnological applications (*i.e.* central carbon metabolism) would offer an unprecedented ability to deeply reprogram cell factories metabolism. However, it would require the integration of hundreds of genes on existing chromosomes, which is a potential source of genetic instability. This problem can be addressed by the implementation of supernumerary, *de novo* assembled synthetic chromosomes (named NeoChrs) as orthogonal expression platforms for genome remodeling (5).

The present study explores the potential of combining pathway swapping with NeoChrs to simultaneously equip *S. cerevisiae* with new heterologous routes and remodel native metabolic networks. *De novo* production from glucose of the anthocyanin pelargonidin 3-O-glucoside (P3G), a food and industrial dye, was attempted in *S. cerevisiae* as proof of principle (6). Although already demonstrated in *S. cerevisiae*, the synthesis of anthocyanins is highly inefficient in all microbial

platforms tested hitherto (6-8). Requiring the channeling of carbon through 27 core metabolic catalytic steps in the glycolytic, pentose phosphate, shikimate and aromatic amino acids biosynthesis pathway, as well as 10 reactions native to plants (Figure 1 and Supplementary Figure S1), P3G is a perfect paradigm to test the potential of



**Figure 1 - A)** Schematic overview of the anthocyanin production pathway. Compounds that can be measured by HPLC or GC are underlined, coumaroyl-CoA, narigenin-chalcone, and leucopelargonidin were not measured. The end products of the pathway are indicated in a frame. **B)** Extracellular production of aromatic intermediates of the anthocyanin pathway by engineered *S. cerevisiae* strains IMF41 (Circular NeoChr) and IMF42 (Linear NeoChr). The yeast strains were grown in shake-flask cultures with chemically defined minimal medium with 20 g L<sup>-1</sup> glucose as carbon source and urea as nitrogen source for 72 hours. The color of the bars corresponds to the color of the metabolites in panel A. Phenylpyruvate (blue), *p*-hydroxyphenylethanol (green), 2-phenylethanol (dark red), cinnamic acid (orange), coumaric acid (grey), phloretic acid (sky blue), dihydrokaempferol (purple), kaempferol (dark green), pelargonidin (red) and kaempferol 3-O-glucoside (dark blue). Naringenin and pelargonidin 3-O-glucoside were not detected. The right panel shows a magnification of the data of the metabolites downstream naringenin. The data represents the average  $\pm$  standard deviation of independent biological triplicates.

NeoChrs for metabolic engineering. Modular NeoChrs in linear and circular form harboring yeast native, bacterial and plant genes required for P3G *de novo* synthesis were designed and constructed (Figure 1 and Supplementary Figure S1). The resulting strains were tested by in-depth genetic and physiological characterization, and compared to strains carrying test NeoChrs of equivalent size but mostly composed of non-coding DNA. Ultimately, the ability of NeoChrs to serve as landing pad for large pathways and to be edited by CRISPR-Cas9 for metabolic engineering purposes was evaluated.

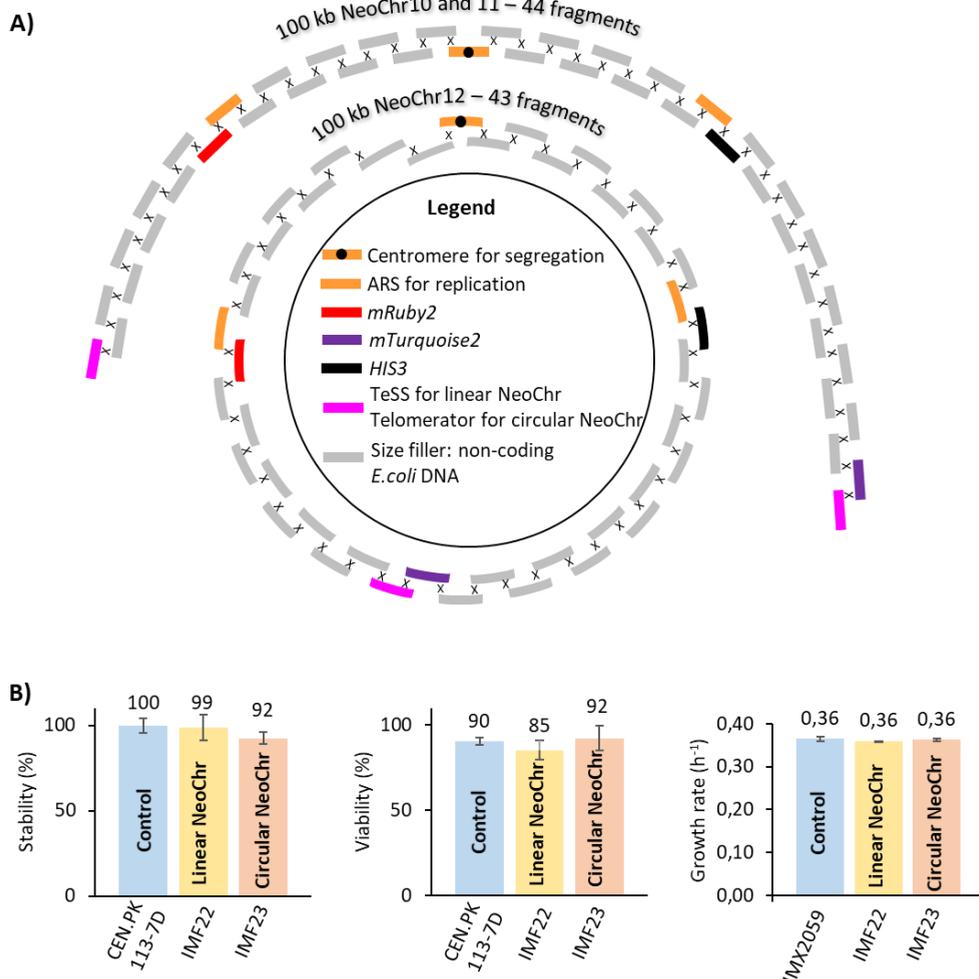
## Results

### Neochromosome genetic design: does circular or linear configuration matter?

In *S. cerevisiae* the highly efficient homology directed repair machinery can be exploited for the modular assembly of tailored synthetic chromosomes (5). The application of these *de novo* assembled NeoChrs as orthogonal expression platform for pathway engineering requires the fulfilment of several core, size-independent, properties: efficiency and fidelity of assembly, stability in terms of sequence fidelity as well as copy number during replication and segregation, and finally absence of toxicity towards the microbial host. These properties were well met in one particular NeoChr design in strain IMF23 constructed by Postma *et al.* (5), offering a promising proof of principle. The chromosomes in this study however were circular, and mimicking the linear configuration of yeast native chromosomes might further improve these core properties (9,10). To test this hypothesis, a linear, 100 kb NeoChr was assembled *in vivo* from transcription-unit sized DNA parts and compared to the previously published, circular NeoChr.

The design of the linear test NeoChr in the present study was identical to the 100 kb circular NeoChr design in strain IMF23, consisting of 43 DNA fragments of 2.5 kb (5). To summarize, 36 of the DNA fragments used were non-coding in yeast and originated

from *Escherichia coli*, the remaining 7 fragments encompassed selection markers, fluorescent reporters and the elements required for chromosome replication and segregation (Figure 2). Direct assembly of linear chromosomes of this scale has never



**Figure 2 - Construction and physiological characterization of test Neochromosomes**

**A)** Schematic representation of the *in vivo* assembly design for linear and circular chromosomes. **B)** Physiological characterization of strains with a 100 kb *de novo* assembled linear (IMF22) and circular (IMF23) test NeoChr. **Left graph**, stability of NeoChrs calculated as ratio of the number of colonies on selective medium (with respect to the NeoChr) divided by the number of colonies on non-selective medium. Data represent the average and standard deviation of two days of measurements (day 1 and day 4) of culture duplicates. **Middle graph**, viability of strains, counted as the ratio between the number of colonies growing on non-selective YPD medium and the total number of individually plated cells. Bars represent the average and standard deviation of two days of measurements (day 1 and day 4) of culture duplicates. **Right graph**, specific growth rate of strains grown on selective SMD medium. Growth rates represent the average and standard deviation of six biological replicates for IMX2059 and IMF23 and two biological replicates for IMF22. None of the measurements show significant differences between strains (one-way ANOVA with Post-Hoc Tukey-Kramer,  $p < 0.05$ ).

been attempted before, therefore particular attention was given to the DNA parts framing the chromosomes, containing the telomeric regions. The telomerator containing short telomere seed regions (TeSS) flanking an I-SceI recognition sequence (embedded as intron in a functional *URA3* selection marker), was previously shown to lead to stable linear chromosomes upon *in vivo* digestion (11,12). The same TeSS were used for the direct assembly of linear chromosomes. This design, composed of 44 fragments and identical TeSS regions on the right and left arm of the NeoChr was named NeoChr10. To evaluate the risk of unwanted recombination events between the two identical terminal TeSS fragments, which might cause circularization during *in vivo* assembly of all the DNA parts, a second design was tested. In this second design, one of the TeSS was mutagenized to prevent homologous recombination with the original TeSS, leading to NeoChr11. With the exception of the telomerator module (carrying the *URA3* selection marker), the design of the linear chromosomes (NeoChr10 and 11) and circular chromosome (NeoChr12) was identical, which enabled the direct comparison of the core properties between linear and circular chromosomes. The identification of correctly assembled chromosomes was performed by screening for expression of fluorescent markers by FACS, for chromosome size by CHEF and for DNA sequence by whole genome, long-read sequencing (Table 1 and Supplementary Figure S2-S3).

As reported by Postma *et al.* (5), 36% of the transformants with circular chromosomes were true to the *in silico* design. Conversely, irrespective of the telomere sequence used, no linear NeoChrs faithful to the original design were found. Seven transformants of the linear NeoChr10 and one transformant of the NeoChr11, with correct size according to CHEF analysis (Table 1) were sequenced by Nanopore technology. Presence of intact telomeres for all NeoChrs and absence of circularization for both NeoChr10 and 11, demonstrated that the short TeSS supplied were sufficient for the formation of functional telomeres and that homology between the two telomeric fragments was not a hurdle for the direct assembly of linear chromosomes. A wide variety of configurations was observed for these assembled linear NeoChrs (Supplementary Figure S4), ranging from the absence of a single fragment (NeoChr10.13 and 10.47), a combination of missing and duplicated fragments (*e.g.* NeoChr10.62) and more complex configurations with missing, duplicated or inverted fragments and swapped regions (NeoChr10.54 and 11.19). These misassemblies revealed the difficulty encountered by yeast cells to assemble all supplied DNA parts, and demonstrated the intervention of non-homologous end joining, while homology directed repair is typically the preferred mode of DNA double strand break repair in *S. cerevisiae* (13-15). The large impact of linear versus circular configuration, despite otherwise identical design, on *in vivo* chromosome assembly possibly revealed differences in the accessibility of telomerator and TeSS fragments in the nucleus for repair and assembly with the other fragments. Nevertheless, two

NeoChrs, NeoChr10.13 and 10.47, displayed a remarkably high degree of fidelity with the *in silico* design, with a single fragment (7A) missing while the remaining 43 fragments were correctly assembled. Overall, keeping in mind that circular chromosomes carried an additional auxotrophic marker as compared to linear chromosomes, these results demonstrated that circular chromosomes are superior in terms of assembly efficiency and fidelity as compared to linear chromosomes.

**Table 1 - NeoChrs assembly efficiency and fidelity.** NeoChrs assembly fidelity tested by the screening pipeline. Correct sequence means that all fragments are present in the correct configuration with respect to the *in silico* design.

<sup>a</sup> excluding telomeres.

<sup>b</sup> due to a bacterial infection on plates, the total number of colonies is not reliable. Yeast colonies were subsequently isolated by microscopy based on fluorescence (88 for NeoChr25 and 15 for NeoChr26), the data in the rows below are therefore reliable.

<sup>c</sup> For NeoChr25 all 12 colonies were screened on CHEF, 6 colonies showed correct size while no bands were observed for the other colonies. ND: not determined.

	Test NeoChrs			NeoChrs	
NeoChr	NeoChr10	NeoChr11	NeoChr12	NeoChr25	NeoChr26
<b>Configuration</b>	Linear	Linear	Circular	Linear	Circular
<b>Chromosome size<sup>a</sup> (bp)</b>	99228	99228	99228	99832	99832
<b>Number of fragments</b>	44	44	43	43	42
<b>Auxotrophic markers</b>	<i>HIS3</i>	<i>HIS3</i>	<i>HIS3</i> and <i>URA3</i>	<i>HIS3</i>	<i>HIS3</i> and <i>URA3</i>
<b>Number of colonies</b>	72	31	11	391 <sup>b</sup>	94 <sup>b</sup>
<b>Correct fluorescence</b>	9 (=13%)	7 (=20%)	8 (=72%)	12 (14%)	12 (80%)
<b>Correct size</b>	7 (=10%)	1 (=3%)	5 (=45%)	6 (=7%) <sup>c</sup>	ND <sup>c</sup>
<b>Correct sequence</b>	0 out of 7 screened	0 out of 1 screened	4 (=36%) out of 5 screened	1 out of 4 screened	3 out of 5 screened

The physiology of strains IMF22, carrying the linear NeoChr10.13, and IMF23, harboring a correct circular chromosome NeoChr12, was compared (Figure 2B). Both strains grew as fast as the control strain on minimal, chemically defined medium and

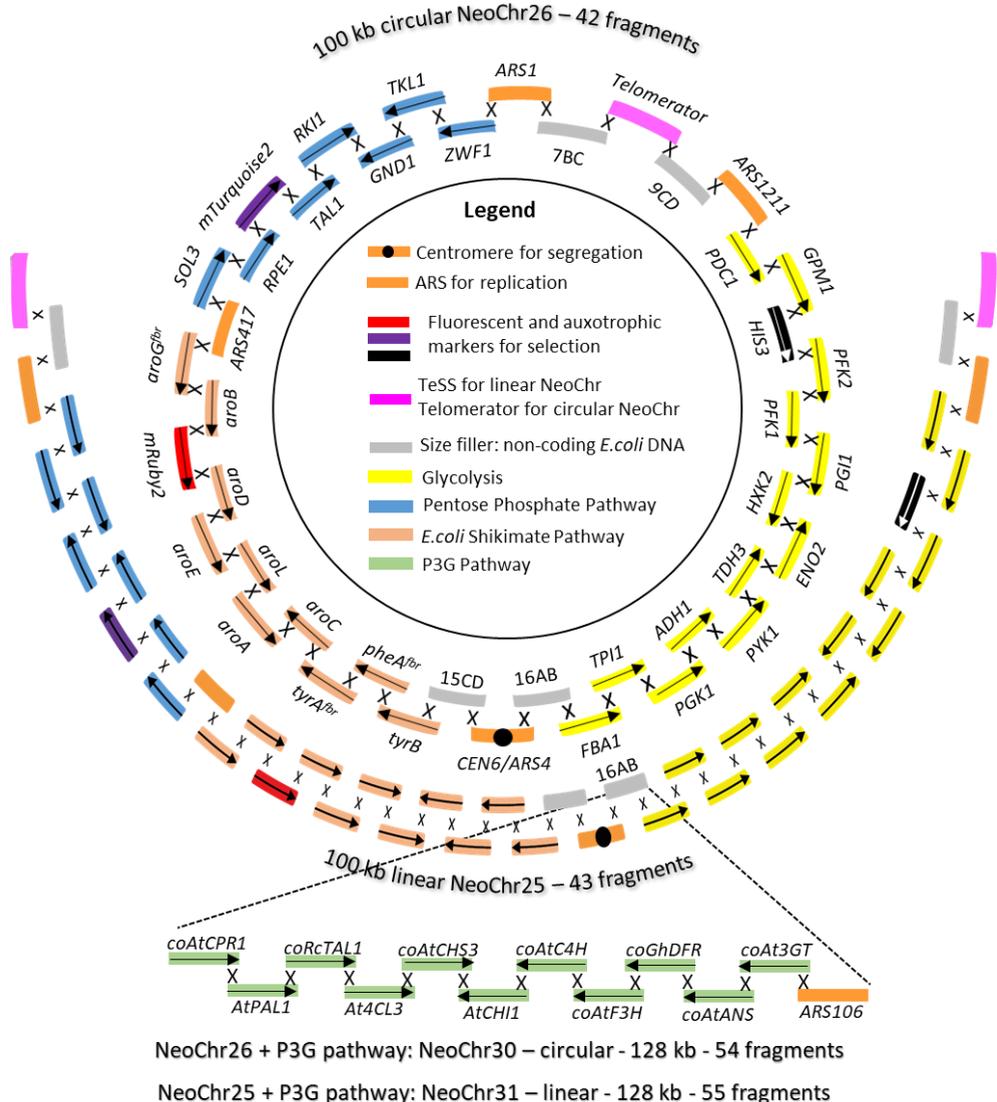
displayed the same viability on complex medium. During propagation of the strains over four days (ca. 25 generations), the fraction of the population containing both linear and circular chromosomes (measured as the ratio of colonies on selective over non-selective medium) remained unaltered and similar to the control strain, demonstrating the stability of the NeoChrs during cell division. Moreover, fluorescence analysis and sequencing showed that both NeoChr-configurations were present in a single copy per cell (Supplementary Figure S5-S6). To conclude, both linear and circular *de novo* assembled 100 kb NeoChrs were present in one copy per cell, stable and innocuous to their host. As the linear or circular nature of the test NeoChrs did not visibly affect the phenotype of the yeast strains, both configurations were further tested as platforms for metabolic engineering.

### ***In vivo, de novo* modular assembly of specialized NeoChrs for anthocyanin synthesis**

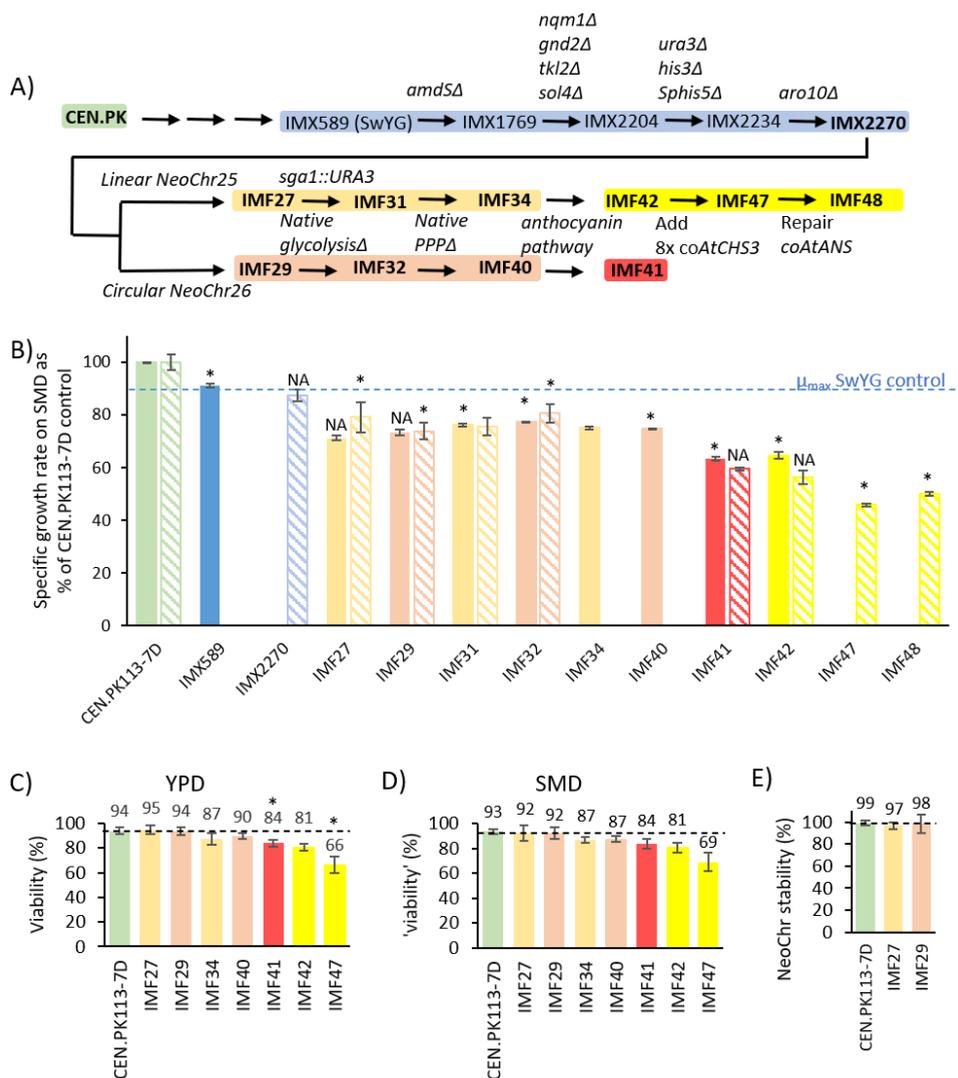
Following the construction strategy as described above, linear and circular NeoChrs, which were expressing the genes native to yeast, bacteria and plants required for P3G production, were designed (Figure 3). The precursors for P3G synthesis natively produced in yeast are L-tyrosine and L-phenylalanine (Figure 1). For *de novo* P3G synthesis from glucose, carbon flows through glycolysis, the pentose phosphate pathway (PPP), cytosolic acetyl-CoA and malonyl-CoA synthesis, the shikimate pathway and aromatic amino acid biosynthesis (Supplementary Figure S1). The genes encoding enzymes in these pathways are scattered over the 16 yeast chromosomes and many have a high degree of genetic redundancy, making the remodeling of these metabolic routes extremely challenging. Expanding on the pathway swapping concept, the strain construction strategy was built on SwYG, a strain harboring a single locus, minimized glycolysis and fermentation pathway (13 genes involved in the conversion of glucose to ethanol (4)). This strain was further engineered by the genetic reduction of the pentose phosphate pathway (removal of the four minor paralogs *NQM1*, *GND2*, *TKL2*, *SOLA* (16)) with the goal of relocating the genes encoding glycolysis, ethanol fermentation and the PPP from native chromosomes to the specialized NeoChrs (Figure 3-4 and Supplementary Figure S1). The biosynthesis of amino acids is tightly regulated in *S. cerevisiae*, particularly via feedback inhibition (17). Therefore, to improve the supply of tyrosine and phenylalanine for P3G synthesis, the entire pathways for aromatic amino acids synthesis from *E. coli* (10 genes) including key feedback resistant alleles (*coEcaroG<sup>br</sup>* (18), *coEtyrA<sup>br</sup>* (19), *coEpheA<sup>br</sup>* (20)) were integrated in the design of the specialized NeoChrs. Finally, as aromatic fusel alcohols and acids are undesired by-products during flavonoid production (21), the 2-oxo acid decarboxylases responsible for their production (Pdc5, Pdc6 and Aro10) had to be removed (22,23). In the SwYG strain, *PDC5* and *PDC6*, homologues to the major pyruvate decarboxylase encoded by *PDC1*, were already deleted. *ARO10*, that encodes

a 2-oxo acid decarboxylase with broad substrate specificity, was therefore deleted as well from the SwYG strain (Figure 4A).

To summarize the NeoChr design for P3G synthesis, the specialized synthetic chromosomes were intended as exclusive expression platforms for the glycolytic, fermentative, pentose phosphate pathways (20 genes), the bacterial shikimate and



**Figure 3 - Genetic design of NeoChr for anthocyanin production.** *In silico* design of NeoChr25, NeoChr26, NeoChr30 and NeoChr31. The arrows indicate the directionality of transcription. The names of the genes and auxiliary parts are indicated on the circular chromosome. All genes from *E. coli* were codon-optimized. The linear NeoChr differs from the circular NeoChr by the presence of two TeSS ends instead of the telomerator (carrying the *URA3* selection marker, pink parts). The eleven plant genes required for P3G biosynthesis were integrated into chunk 16AB from NeoChr25 and NoeChr26, resulting in linear NeoChr30 and circular NeoChr31 respectively.



**Figure 4 - Construction and physiological characterization of strains with NeoChrs engineered for anthocyanin production.**

**A)** Overview of the strains constructed in this study for pathway engineering. **B)** Specific growth rate on SMD medium of engineered strains expressed as % of the control CEN.PK113-7D. The same color-coding has been used as for panel A. Solid-filled bars: average and standard deviations of biological duplicates grown in shake flasks. Pattern-filled bars: average and standard deviations of three to six biological replicates grown in plate reader. For strain IMF27 and IMX589, uracil was supplemented to the medium. For IMX2270 uracil and histidine were supplemented to the medium. **C** and **D**, viability calculated as the number of colonies on YPD (**C**) or SMD (**D**) divided by the total number of plated cells (384 FACS-sorted cells). **E)** Stability of the NeoChrs measured as number of transformants on selective medium (SMD (supplemented with uracil)) divided by the number of colonies on non-selective medium (YPD). For panels **C**, **D** and **E**, bars represent the average and standard deviation of biological duplicates measured on two days (day 1 and day 4). Significant differences with respect to the first ancestor in the same graph are indicated with an asterisk (two-tailed paired homoscedastic t-test  $P < 0.05$ ). In Panel **B** NA indicates data for which statistics could not be calculated because the ancestor was grown in a different set-up than the tested strain (shake flask or plate reader).

amino acid biosynthesis pathway (10 genes), and the 11 plant genes involved in the synthesis of P3G from tyrosine and phenylalanine.

The genetic design of the specialized NeoChrs was particularly challenging. This study presents the very first attempt of pathway construction of this magnitude, and information to rationally design an optimal configuration for NeoChr assembly with high efficiency, stability and expression is scarce. The first design consideration was high fidelity assembly and maintenance of the NeoChrs. While the present study embraces the remarkable ability of *S. cerevisiae* to recombine homologous sequences, homology directed repair (HDR) might cause unwanted recombinations within the NeoChr during assembly and propagation. To prevent unwanted recombinations, homology within the NeoChr was kept to a minimum. However, maintaining a low homology for promoter and terminators regions for 41 transcription units was challenging. The 20 yeast genes were framed by their native promoter and terminator, however the 21 bacterial and plant genes also required terminators and strong constitutive promoters with unique sequences. Such promoters are not abundant in the *S. cerevisiae* molecular toolbox, and most often of glycolytic origin. The heterologous genes were therefore expressed from a selection of previously characterized *S. cerevisiae* promoters (24,25) which also showed high and condition-independent expression in *S. cerevisiae* from a large transcriptome compendium (26), and promoters from *Saccharomyces eubayanus* and *Saccharomyces kudriavzevii* that have little sequence similarity but share functionality with their *S. cerevisiae* relatives (27) (Supplementary Table S1). The promoter regions of the transcription units on the NeoChr design did not display homologous sequences longer than 43 bp (*pScADH1/pSkADH1* and *pScTDH3/pSkTDH3*). Terminators, that generally have a milder impact on gene expression than promoters, especially when paired with strong promoters (28-30), were selected from a set of in-house central carbon metabolism terminators (Supplementary Table S1).

The second consideration for the genetic design of the NeoChrs (Figure 3) was the spatial organization of transcriptional units and other elements along the chromosomes, which might impact both maintenance and gene expression of NeoChrs. To facilitate future modular pathway remodeling, genes encoding glycolysis, fermentation, PPP, shikimate and aromatic amino acids pathways were clustered per pathway. To prevent gene silencing resulting from chromatin structure near centromere and telomeres, transcriptional units were separated from these elements by 5 kb non-coding DNA fragments originating from *E. coli*. To facilitate the screening of correctly assembled chromosomes, fluorescent reporters (*mRuby2*, *mTurquoise2*) and auxotrophic markers (*HIS3* and *URA3* for circular and *HIS3* only for linear NeoChrs) were evenly spaced along the NeoChrs. Finally, while highly expressed genes are scattered across native chromosomes (31), the designed NeoChrs are

transcriptional hotspots with high density of heavily transcribed genes. To prevent potential clashes between the replication and transcription machineries, the directionality of transcription units was chosen to follow the same orientation as replication (11).

To test the suitability of NeoChrs for both *de novo* assembly of pathways and as landing pads for pathways a two-steps construction approach was followed. In a first step, 100 kb circular and linear chromosomes carrying all yeast and bacterial genes were assembled *de novo* in yeast (NeoChr26 and 25, respectively) (Figure 4A). In a second step, the genes for anthocyanin synthesis from plant and from *Rhodobacter capsulatus* (*coRcTAL1* only) were simultaneously *de novo* assembled and integrated as a module in the NeoChrs (Figure 3).

*De novo* assembly of fully coding NeoChrs verified by fluorescence, CHEF and sequencing, confirmed the higher assembly efficiency of circular over linear NeoChrs observed with the test NeoChrs (Table 1, Supplementary Figure S7-S8 and Supplementary Table S2-S3). Remarkably, one linear NeoChr was faithful to the *in silico* design. Genome re-sequencing of a transformant with this NeoChr (NeoChr25, strain IMF27) and sequencing of a transformant with a correctly assembled circular NeoChr (NeoChr26, strain IMF29), identified few mutations, with a single non-synonymous mutation in a coding region of NeoChr26 (*RK11*, Supplementary Table S2). Summarizing, the assembly of the linear and circular NeoChrs was successful and few mutations in the native and synthetic chromosomes were caused by the addition of the NeoChrs (Supplementary Table S2-S3). Next, exclusive expression of glycolysis, fermentation and PPP from the specialized NeoChrs was achieved by Cas9-mediated excision of the single locus glycolysis from ChrIX and multiplex deletion of the seven genes of the minimized PPP (Figure 4A). The targeted deletion of the seven PPP genes scattered over the native chromosomes was facilitated by the insertion of watermarks on their near-identical copy carried by the NeoChrs (32).

As previously observed, genetic relocation of the glycolytic and fermentative pathways in SwYG caused a mild (IMX589, 10%) reduction in growth rate as compared to the control strain CEN.PK113-7D (Figure 4B, (4)). Additional genetic reduction of the PPP and deletion of *ARO10* (strain IMX2270) had no visible additional phenotypic effect (Figure 4B). Remarkably, NeoChrs with 30 yeast and bacterial genes were as stable during cell division as native chromosomes in control strain CEN.PK113-7D, and strains IMF27 and IMF29 were as viable as the control strain. Their growth rate was slightly decreased (10-16%) with respect to their parental strain IMX2270 (Figure 4B). The linear or circular topology affected neither NeoChr stability and viability nor specific growth rate. Finally, exclusive expression of the yeast glycolytic, fermentative and pentose phosphate pathway from the NeoChrs (strains IMF34 (linear) and IMF40

(circular)) did not further decrease viability nor the specific growth rate, which remained high at ca. 0.30 h<sup>-1</sup> (Figure 4B and C).

### **Expression of the anthocyanin biosynthetic pathway from a NeoChr**

Inspired by earlier work (7), synthesis of P3G was attempted by integration of 11 anthocyanin genes from *Arabidopsis thaliana* (*AtPAL1*, *coAtCPR1*, *coAtC4H*, *At4CL3*, *coAtCHS3*, *AtCHI1*, *coAtF3H*, *coAt3GT*, *coAtANS*), *Gerbera hybrida* (*coGhDFR*) and *Rhodobacter capsulatus* (*coRcTAL1*). As basal design, single copies of these 11 genes, interspaced by an ARS sequence, were integrated in the 16AB *E. coli* DNA chunk on the circular and linear NeoChrs (Figure 3 and 4A), leading to NeoChr30 (IMF41) and NeoChr31 (IMF42), respectively. The two strains grew with a similar specific growth rate, but as compared to their parental strain devoid of anthocyanin pathway, they displayed a 15-17% slower specific growth rate (IMF34 and IMF40, *p*-values of 0.01 and 0.00 respectively, Figure 4B), in agreement with previous observations (7). The viability of IMF41 was similar to that of IMF42, but IMF41 was slightly less viable than its ancestor IMF40 (7% lower, *p*-values of 0.03, Figure 4C and D). Flow cytometric DNA quantification, short-and long-read sequencing showed that both linear and circular chromosomes were present in a single copy in IMF41 and IMF42 (Supplementary Figure S5 and S6).

In both IMF41 and IMF42, insertion of the plant genes on the NeoChr caused the duplication and inversion of the last four genes in the anthocyanin pathway, *coAtF3H*, *coGhDFR*, *coAtANS* and *coAt3GT*. Sequence analysis revealed that this unexpected event resulted from homologous recombination between the *ScFBA1* promoter (upstream *ScFBA1*) and the *SeFBA1* promoter (upstream *coAtC4H*), although slightly different regions of the promoters recombined in IMF41 and IMF42 (Supplementary Figure S9). The two promoters share 58% homology with identical stretches of 24 nucleotides. The absence of homology in the region where the second recombination event occurred suggested the non-homologous recombination of the two *pSePYK1*-SHR CJ ends (Supplementary Figure S9). The presence of the same chromosomal rearrangement in both the linear and circular chromosome, involving rare micro-homology (7-24 bp of 100% homology) and non-homologous recombination events, and the absence of *ARS106*, suggested that the *ARS106* fragment was supplied in suboptimal concentration during transformation. As the same duplication occurred in the strains with linear and circular NeoChr, the impact of NeoChr shape can still be investigated. Plant proteins are typically difficult to express in a yeast environment, a positive rather than negative effect of this duplication on anthocyanin production was therefore anticipated. If desired the duplicated genes can be removed using CRISPR/Cas9 editing.

Short- and long-read sequencing revealed the occurrence of a 46 bp insertion at the beginning of the *coAtANS* gene. The presence of this mutation in the *coAtANS* expression cassette showed that it occurred prior to assembly in the NeoChr, and was not caused by *in vivo* assembly. This insertion led to an early stop codon, however the presence of a start codon further along the protein still enabled the synthesis of a 277 amino acids long peptide (Supplementary Figure S10). This 80 amino acids shortening from the N-terminus most likely strongly affected the anthocyanidin synthase functionality in IMF41 and IMF42. With the exception of a single non-synonymous mutation in the gene encoding the 4-coumarate:CoA ligase (*At4CL3*, Thr-15-Ala) in IMF42, plant genes were exempt of mutations.

Grown in aerobic shake flask cultures in chemically defined medium, IMF41 and IMF42 produced detectable levels of anthocyanins. The total amount of aromatics produced by the two strains was similar ( $1.69 \pm 0.08$  and  $1.61 \pm 0.06$  mM) and anthocyanins down to dihydrokaempferol (DHK) could be reliably detected and quantified (Figure 1, Supplementary Table S4). The production of DHK revealed the functional expression of at least eight out of the eleven NeoChr-borne plant genes in *S. cerevisiae*. Considering that metabolites downstream *At4CL3* such as phloretic acid and dihydrokaempferol were detected in both IMF41 and IMF42, the mutation of *At4CL3* had no detectable effect (Figure 1).

Over 98% of the produced aromatic products were represented by intermediates (cinnamic and coumaric acid) or byproducts upstream chalcone synthase (phloretic acid) (Figure 1), a result in line with previous reports of chalcone synthase being a bottleneck for anthocyanin production, and a typical target for gene-dosage engineering (7,21). While the synthesis of the end products P3G and K3G is challenging, detectable levels of other metabolites downstream dihydrokaempferol were expected (kaempferol, pelargonidin and leucopelargonidin) (7). The absent or barely detectable levels of these metabolites, both in culture supernatants and cell extracts (Figure 1 and Supplementary Table S4) confirmed the poor functionality of the truncated anthocyanidin synthase *AtAns*. Overall, these results showed that the linear or circular structure of the NeoChrs did not affect the physiology of the engineered strains or anthocyanin production, and that the two-step assembly of a NeoChr expressing the *de novo* anthocyanin synthesis pathway enabled anthocyanin production.

### **Neochromosome engineering for improved *de novo* anthocyanin biosynthesis**

The two engineering targets identified to improve anthocyanin synthesis were implemented by CRISPR/Cas9 editing of the NeoChrs. Targeting the linear NeoChr31 (strain IMF42), four additional *coAtCHS3* copies were integrated in four different loci on the NeoChr spaced by essential genes (in Chunk 7BC, chunk 15CD, SHR N and chunk 9CD, Figure 3). In an attempt to further boost anthocyanin synthesis, four additional

*coAtCHS3* copies were integrated in native chromosomes (at *CAN1*, *X2*, *YPRCtau3* and *SPR3* loci (33-36)), resulting in a total of nine *coAtCHS3* copies (IMF47). Finally, the *coAtANS* gene was repaired leading to strain IMF48. The structure of the NeoChr was not affected by these additional genetic interventions.

The performance of IMF48, IMF42 and IMF41 was compared in pH-controlled, aerated bioreactors, a culture tool previously shown to enable higher anthocyanin titers than shake flasks (7). The strains were grown in minimal, chemically defined medium without additional growth supplements (*e.g.* yeast extract or peptone), to purely evaluate *de novo* production of anthocyanins from glucose. The physiology of the three strains was similar, despite the eight additional *coAtCHS3* copies carried by IMF48 (Supplementary Table S5). Irrespective of the strain, flavonoids were produced both during exponential growth on glucose and after the diauxic shift (Supplementary Figure S11). Confirming the results from shake flask cultures, the structure of the NeoChr in IMF41 and IMF42 affected neither yeast physiology nor flavonoid formation (Table 2, Supplementary Figure S11 and Supplementary Table S5). Overall, strain IMF48 outperformed IMF41 and IMF42 carrying a single *coAtCHS3* copy and a truncated *coAtANS* (Table 2, Supplementary Figure S11 and Supplementary Table S5). IMF48 produced 14-fold more flavonoids downstream chalcone synthase than IMF41 and IMF42, and the production of pelargonidin, kaempferol and K3G, metabolites downstream *AtAns*, was increased by ca. 30-, 350- and 500-fold, respectively in IMF48 as compared to IMF41 and IMF42 (Table 2). P3G, not detected intracellularly nor extracellularly in IMF41 and IMF42 cultures, was produced with a titer of  $0.049 \pm 0.007 \mu\text{M}$  by IMF48 at the end of the glucose consumption phase (Supplementary Figure S11). Notably, this is the first report of P3G extracellular production by yeast cultures. The concentration of P3G decreased by ca. three-fold at the end of the diauxic phase, presumably due to degradation by periplasmic  $\beta$ -glucosidases (*e.g.* *Exg1*), upon glucose exhaustion (Supplementary Figure S11). These enzymes are interesting targets for further improvement of P3G production. IMF48 accumulated more coumaric acid and *p*-hydroxyphenylethanol, aromatic intermediates upstream *AtChs3*, an unexpected response to the increased chalcone synthase expression that suggested some yet unidentified regulation of the plant enzymes in the yeast environment (Table 2). No extracellular naringenin was detected for any of the tested strains. The absence of naringenin but increased intra- and extracellular dihydroxykaempferol concentration with increasing *coAtCHS3* copy number and flux through the anthocyanin pathway, showed that flavonoid hydroxylase (F3H) was most probably not limiting P3G production and was therefore not a target for genetic engineering at that stage. This apparent overcapacity of flavonoid hydroxylase might be related to its presence in two copies on the NeoChrs. Conversely, the strong increase in flavonoid concentration upstream the last enzyme, 3GT (kaempferol and pelargonidin, Table 2),

**Table 2 - Characterization of anthocyanin production in bioreactors.** Determination of the intermediates of the anthocyanin pathway in *S. cerevisiae* strains IMF41, IMF42 and IMF48. Extracellular and intracellular metabolite concentration at the end of aerobic and pH controlled (pH 5.0) bioreactor batch cultures on glucose. The concentrations (mM) of metabolites upstream from, and including, naringenin were measured by HPLC while metabolites downstream were quantified by LC-MS method. The data represents the average  $\pm$  standard deviation of independent biological duplicates. Intermediates of the anthocyanin pathway coumaroyl-CoA, naringenin chalcone, and leucopelargonidin were not measured. An asterisk indicates statistical significance when comparing extracellular concentrations of IMF48 to concentrations of IMF41 and IMF42 (Student *t*-test, *p*-value threshold 0.05, two-tailed, homoscedastic).

BD Below Detection, NM Not Measured

	IMF41		IMF42		IMF48	
	Extracellular (mM)	Intracellular ( $\mu$ mol/g dry weights)	Extracellular (mM)	Intracellular ( $\mu$ mol/g dry weights)	Extracellular (mM)	Intracellular ( $\mu$ mol/g dry weights)
Phenylpyruvate	6.68E-02 $\pm$ 0.00E+00	NM	4.47E-02 $\pm$ 4.95E-03	NM	BD	NM
2-phenylethanol	BD	NM	BD	NM	BD	NM
<i>p</i> -hydroxyphenylethanol	8.94E-02 $\pm$ 2.55E-03	NM	7.87E-02 $\pm$ 9.90E-04	NM	1.40E-01 $\pm$ 2.83E-03*	NM
Cinnamic acid	BD	NM	BD	NM	BD	NM
Coumaric acid	8.46E-01 $\pm$ 2.38E-02	NM	7.89E-01 $\pm$ 6.22E-03	NM	2.05E+00 $\pm$ 1.98E-02*	NM
Phloretic acid	5.46E-01 $\pm$ 3.24E-02	NM	5.92E-01 $\pm$ 1.15E-02	NM	5.98E-01 $\pm$ 2.42E-02	NM
Naringenin	BD	NM	BD	NM	BD	NM
Dihydrokaempferol	3.29E-02 $\pm$ 1.64E-04	7.26E-02 $\pm$ 4.44E-02	3.32E-02 $\pm$ 2.01E-04	1.41E-01 $\pm$ 1.93E-02	2.90E-01 $\pm$ 3.98E-03*	1.15E+00 $\pm$ 2.00E-01
Kaempferol	1.04E-04 $\pm$ 8.51E-06	1.20E-03 $\pm$ 8.84E-04	8.61E-05 $\pm$ 1.64E-05	1.58E-03 $\pm$ 6.99E-04	3.55E-02 $\pm$ 5.63E-03*	3.64E+00 $\pm$ 1.18E+00
Pelargonidin	1.46E-03 $\pm$ 1.99E-05	6.25E-03 $\pm$ 3.96E-03	1.64E-03 $\pm$ 2.73E-05	1.00E-02 $\pm$ 3.26E-03	4.14E-02 $\pm$ 5.63E-04*	4.96E-01 $\pm$ 7.76E-02
Kaempferol 3-O-glucoside	2.74E-05 $\pm$ 5.38E-07	3.42E-05 $\pm$ 2.88E-05	2.17E-05 $\pm$ 6.21E-07	1.18E-04 $\pm$ 8.03E-05	9.77E-03 $\pm$ 1.09E-04*	1.44E-02 $\pm$ 1.70E-03
Pelargonidin 3-O-glucoside	BD	BD	BD	BD	1.45E-05 $\pm$ 4.78E-06*	2.51E-04 $\pm$ 1.95E-05
Total aromatics (a)	1.47 $\pm$ 0.04	-	1.14 $\pm$ 0.03	-	3.3 $\pm$ 0.23*	-
Total flavonoids (b)	0.03 $\pm$ 0.002	-	0.03 $\pm$ 0.002	-	0.51 $\pm$ 0.19*	-
Fraction of flavonoids (=b/a)	2%	-	2%	-	15%*	-

suggested that despite the duplication of its gene, this enzyme might be limiting for K3G and P3G production and a target for further improvement. The titers of produced flavonoids were modest. However, the fraction of flavonoids over all aromatic compounds produced by IMF48 was 15%, representing a substantial improvement compared to the 2% produced by IMF41 and IMF42. Additionally, the total production of aromatic compounds was increased by 2.3-fold in IMF48.

The combined *coAtCHS3* copy number increase and *coAtANS* repair markedly affected the flux distribution in the aromatics pathway. To estimate the respective contribution of chalcone and anthocyanidin synthase to these changes, the parental strain of IMF48, a strain carrying nine *coAtCHS3* copies, but with impaired *coAtANS* was tested (IMF47, Figure 4A). In line with increased chalcone synthase activity, IMF47 produced three times more flavonoids than IMF42, while the production of metabolites upstream chalcone synthase was unchanged. However, P3G was still not detected in IMF47 cultures. The fraction of flavonoids over total aromatics was 6% in IMF47 (Supplementary Table S4). Repair of *coAtANS* in IMF48 further increased flavonoids production by *ca.* four-fold and slightly, but significantly, increased total aromatics production as compared to IMF47 (1.4-fold, Supplementary Table S4). Altogether these data revealed that both increased *coAtCHS3* copy number and *coAtANS* repair contributed to the improvement of anthocyanin production in IMF48.

## Discussion

The present study illustrates the amazing potential of *S. cerevisiae* for fast and extensive genome remodeling via synthetic chromosome engineering. Supernumerary, specialized chromosomes could be easily and rapidly assembled, in a single transformation round, from 30 transcription units and 13 accessory DNA parts. Beyond this technical 'tour de force', the NeoChrs resembled native chromosomes in terms of replication and segregation. The stable maintenance of the NeoChrs at one copy number and their harmlessness to the host are important features for their implementation as metabolic engineering platforms. Plant-derived chemicals have a broad range of biotechnological applications, but their *de novo* microbial production requires remodeling of the host metabolism, as well as functional transplantation of the plant pathway. Equipped with pathway swapping, NeoChrs enabled the facile implementation of three complex interventions: i) the remodeling of native metabolic networks (glycolysis and PPP), ii) the provision of an optimized metabolic route from prokaryotic origin (shikimate pathway) as surrogate for a native route, and iii) the implementation of a new, heterologous pathway from plant origin (anthocyanin synthesis). These modifications readily enabled the synthesis of anthocyanins from glucose, and the first report of extracellular P3G production in yeast or any other single

microbial host. The implementation of NeoChrs not only accelerated strain construction, and thereby genome remodeling, but also prevented interferences with native chromosomes. While chromosome construction was carefully designed in two steps to probe the limits of *in vivo* NeoChr assembly, the present data suggest that single step chromosome assembly would have readily enabled anthocyanin production in a shorter time frame. Once assembled, the NeoChrs could be edited at will using CRISPR-Cas9. CRISPR-Cas9 editing efficiency is locus-dependent, and a small set of robust genomic integration sites have been validated to date (33,34,37). Carefully designed, mostly non-coding NeoChrs, harboring strategically located, optimized CRISPR-Cas9 programmed sites, could become ideal landing pads for large sets of (heterologous) genes. Finally, the successful assembly of chromosomes from 43 parts suggests that the limit of *in vivo* assembly has not been reached and even larger chromosomes can be assembled if required.

Synthetic Genomics is a young research field (38,39), and the design principles for optimal, tailor-made NeoChrs are ill-defined. The synthetic chromosomes rebuilt in the Sc2.0 initiative uses native chromosomes as scaffolds and yet reproduces the native organization of the chromosomes, albeit omitting non-essential elements and adding some short sequences (40,41). With the possibility to construct and test any genetic design, *de novo*-assembled NeoChrs are fantastic testbeds to explore the genetic and physiological impact of chromosome sequence and structure. In this study, some design guidelines were formulated and tested, leading to stable chromosomes, easy to screen and with functionally expressed transcription units. The 100-130 kb linear and circular NeoChrs in this study showed equal growth rate and mitotic stability. This is in good agreement with earlier studies showing approximate equal stability of linear and circular chromosomes of 100 kb, while beneath this size circular chromosomes appear more stable and above this size linear chromosomes are more stable (11,42-45). A remarkable observation was the low assembly efficiency of linear chromosomes as compared to their circular counterparts. As opposed to transcription units, telomeres and centromeres have specific, cell-cycle dependent localizations within the nucleus (46-49). It is conceivable that, for linear chromosomes equipped with one centromere and two telomeres, accessibility and spatial organization of these DNA parts conflict with homologous recombination, a difficulty potentially alleviated for circular chromosomes equipped with a single centromere and no telomeres. Additionally, the frequent occurrence of non-homologous repair in the linear NeoChrs, an otherwise rare mechanism active throughout the entire cell cycle (13), might indicate temporal incompatibility between telomere and centromere availability and homologous recombination, mostly active during the S/G<sub>2</sub> phase (13). For biotechnological applications, circular chromosomes, easier to isolate from yeast, are

therefore recommended, and can be equipped with a telomerase to enable ulterior linearization if required (11,12).

In their current design, the NeoChrs are extremely information-dense, with short intergenic regions and a concatenation of highly transcribed genes. While this genetic design leads to functional expression of the NeoChr-born genes and is not harmful for the host cells, many fundamental questions regarding optimal genetic design remain to be systematically explored, such as the cell's tolerance to 'transcriptional hotspots' (4,5,50), the impact of transcription units localization, orientation and distancing on gene expression, or the requirement for multiple selection markers for chromosome stability. Furthermore, while NeoChrs harboring spaced homologous sequences were genetically stable, the presence of homology between DNA parts might lead to unwanted recombination events during *in vivo* assembly of the NeoChrs. In this work the DNA parts were designed as 'homology-free' as possible, a design principle difficult to apply considering the limited availability of strong, constitutive promoters. Homology between DNA parts could be kept to a minimum thanks to the implementation of promoters from *S. cerevisiae* relatives (27). Nevertheless, the highest homology between promoters was still 75% (with stretches of identical sequences up to 42 nucleotides (*pSkTDH3* and *pTDH3*)). Along this line the *S. cerevisiae* toolbox could be further enriched by mining non-*Saccharomyces* yeasts genetic diversity for functional but sequence divergent promoters (27,51). In a more distant future, progress in the design of synthetic promoters should enable the construction of libraries of *in silico*-designed, artificial promoters with minimal homology (52,53).

From the present study, we can envisage future microbial cell factories with modular genomes in which core metabolic network and processes, localized on satellite, specialized NeoChrs can be swapped for alternative configurations. Following the 32% reduction of the 111 genes of yeast central carbon metabolism (16,54) the present strategy for *in vivo* assembly of NeoChrs can be applied to construct yeast strains carrying specialized NeoChrs as exclusive expression platforms for central carbon metabolism. Combined with pathway swapping (4), such strains would enable fast and easy remodeling of large sets of core cellular functions. As yeast is tolerant to chromosome ploidy variation (55), other strategically designed NeoChr could carry other industrially-relevant pathways (*e.g.* nitrogen or fatty acids metabolism) and processes (*e.g.* protein secretion). Additional NeoChrs, similar to the test NeoChrs from the present study, but tailored for CRISPR/Cas9 targeting, could also serve as landing pads dedicated to the addition of functionalities.

## Materials and Methods

### Strains, growth medium and maintenance.

All *S. cerevisiae* strains used and constructed in this study are derived from the CEN.PK family (Supplementary Table S6-S7) (56).

For liquid cultures yeast was cultivated in 50-/100-/500 mL shake flasks containing, respectively 10-/20-/100 mL media in an Innova 44 Incubator shaker (New Brunswick Scientific, Edison, NJ, USA) at 30 °C and 200 rpm. Cultures on solid media were incubated at 30°C until single colonies were visible. For non-selective growth, yeast strains were cultivated on Yeast extract Peptone Dextrose (YPD) medium containing: 10 g L<sup>-1</sup> Bacto yeast extract, 20 g L<sup>-1</sup> Bacto peptone and 20 g L<sup>-1</sup> glucose. For selective growth to maintain plasmids or NeoChrs, Synthetic Medium (SM) was used, consisting of: 3 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 5 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 1 mL L<sup>-1</sup> of a trace element solution (57). Alternatively, synthetic medium with urea as sole nitrogen source was used, consisting of 3 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 5 g L<sup>-1</sup> K<sub>2</sub>SO<sub>4</sub>, 2.3 g L<sup>-1</sup> urea and 1 mL L<sup>-1</sup> of a trace element solution (58). Media were set to pH 6 by 1M KOH addition and for solid media, 20 g L<sup>-1</sup> Bacto agar was added. Autoclaving was performed for 20 min at 110°C and 120°C for YPD and SM medium, respectively. Thereafter SM medium was supplemented with 1 mL L<sup>-1</sup> of a filter sterilized vitamin solution and 20 g L<sup>-1</sup> of glucose separately autoclaved for 20 min at 110°C. For auxotrophic strains, SM was supplemented with 125 mg L<sup>-1</sup> histidine and/or 150 mg L<sup>-1</sup> uracil. Disruption of the *URA3* marker was verified by growth on SMD with 150 mg L<sup>-1</sup> uracil and 1 g L<sup>-1</sup> 5-FluoroOrotic Acid (SMD-5-FOA). For the selection based on the markers *hphNT1*, *KanMX* and *amdS*, SM medium without nitrogen source was prepared by replacing (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> with 6.6 g L<sup>-1</sup> K<sub>2</sub>SO<sub>4</sub>. For *hphNT1* and *KanMX*, 2.3 g L<sup>-1</sup> urea was used as nitrogen source and 200 mg L<sup>-1</sup> hygromycin (Hyg) and 200 mg L<sup>-1</sup> G418 were added to the medium, respectively. For *amdS*, 1.8 g L<sup>-1</sup> filter sterilized acetamide was employed as nitrogen source.

All *E. coli* strains were cultivated in Lysogeny Broth (LB) medium containing: 10 g L<sup>-1</sup> tryptone, 5.0 g L<sup>-1</sup> yeast extract and 5 g L<sup>-1</sup> NaCl. For plasmid selection 100 mg mL<sup>-1</sup> ampicillin (ampR), 50 mg mL<sup>-1</sup> kanamycin (kanR), or 25 mg mL<sup>-1</sup> chloramphenicol (camR), was supplemented to the medium. Liquid cultivation was performed in 5 mL medium in a 15 ml Greiner Tubes at 37°C and 200 rpm in an Innova 4000 shaker (New Brunswick Scientific). Cultures on solid media were incubated at 37°C until single colonies were visible.

*S. cerevisiae* and *E. coli* strains were stored at -80°C in 1 mL vials containing cultures mixed with glycerol (30% v/v).

## **Molecular biology techniques**

Genomic DNA from *E. coli* (migula) Castellani and Chalmers (ATCC 47076) or *S. cerevisiae* used for strain construction purposes was isolated using the QIAGEN Blood & Cell Culture Kit with 100/G Genomic-tips (Qiagen, Hilden, Germany) or alternatively for yeast with the YeaStar genomic DNA kit (Zymo Research, Irvine, CA). *E. coli* DNA from a mixed population of *E. coli* XL1-Blue and *E. coli* BL21 used for construction of the test NeoChrs was isolated as described by Postma *et al.* (5). Plasmids were isolated from *E. coli* using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MO) or the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, MA), according to the manufacturer's instructions.

All PCRs for strain construction purposes were performed with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) using either desalted or PAGE purified (in case of ORFs) primers (Sigma-Aldrich). PCR products were verified by separation on 1% (w/v) or 2% (w/v) agarose (TopVision Agarose, Thermo Fisher Scientific) gels in 1x Tris-acetate-EDTA (TAE) buffer (Thermo Fisher Scientific) or 1x Tris-Borate-EDTA (TBE) (Thermo Fisher Scientific) buffer. For size determination GeneRuler DNA Ladder mix (Sigma-Aldrich) or GeneRuler DNA Ladder 50 bp (Sigma-Aldrich) were used. For DNA staining 10  $\mu\text{L L}^{-1}$  SERVA (SERVA Electrophoresis GmbH, Heidelberg, Germany) was added to the agarose gel solution. DNA was purified using either the Zymoclean Gel DNA Recovery kit (Zymo Research), the GenElute PCR Clean-Up kit (Sigma-Aldrich), the GeneJET PCR Purification Kit (Thermo Fisher Scientific) or using AMPure XP beads (Beckman Coulter, Brea, CA) according to the suppliers' protocols. Purity of DNA was checked using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and the concentration was measured either by the NanoDrop 2000 (Thermo Fisher Scientific) or by the Qubit dsDNA BR Assay kit (Thermo Fisher Scientific) using the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). Gibson assembly used to construct gRNA plasmids and some expression plasmids was performed with the NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, Ipswich, MA) in a final volume of 5  $\mu\text{L}$  according to the supplier's instruction.

Chemical *E. coli* XL1-Blue transformation was performed as described by Inoue *et al.* (59) and correct assembly of plasmids was verified by diagnostic PCR or restriction analysis. *S. cerevisiae* was transformed using the lithium acetate/polyethylene glycol method (60). For diagnostic PCR, DNA was isolated by resuspending some culture in 0.2 M NaOH or by using the method described by Looke *et al.* (61). All diagnostic PCRs were performed using DreamTaq PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's instruction. For yeast, single colony isolates were obtained by three consecutive re-streaks on solid selective medium.

## Plasmid construction

All plasmids used in this study are listed in Supplementary Table S8.

gRNA plasmids. GuideRNA (gRNA) plasmids (Supplementary Table S8A) for targeting Cas9 to specific loci were constructed as described by Mans *et al.* (37). Primers to construct and verify the gRNA plasmid are listed in Supplementary Table S9-S10.

Golden Gate part plasmids. Part plasmids compatible with Golden Gate Assembly, harboring promoter, gene or terminator were constructed using the Yeast Toolkit principle (24). A range of part plasmids were present in-house and previously described (Supplementary Table S8B). Some of the promoters, genes and terminators flanked by *BsaI* and *BsmBI* restriction sites were ordered from GeneArt (Thermo Fisher Scientific). The promoters and terminators listed in Supplementary Table S8C were subcloned by GeneArt (Thermo Fisher Scientific) in the entry vector pUD565 and could be directly used for construction of expression plasmids. The pentose phosphate pathway genes were ordered from GeneArt but subcloned in-house into entry vector pUD565 (Supplementary Table S8D). Finally, some part plasmids were made by amplifying the target region with primers containing part type specific overhangs (Supplementary Table S11) and assembling in entry vector pUD565 by *BsmBI* cloning. These plasmids are listed in Supplementary Table S8E with the respective primers and template for target region amplification. Internal *BsmBI* or *BsaI* sites were removed as described by Hassing *et al.* (62) and verified by Sanger sequencing (Baseclear, Leiden, The Netherlands). For two parts (*coAtANS* #1 and *pSePYK1*) no correct *E. coli* part plasmid transformant was found. Therefore, the PCR fragments containing yeast toolkit flanks were directly assembled into expression cassettes (Supplementary Table S8G) Plasmids were verified by PCR and/or restriction analysis (Supplementary Table S12).

Golden Gate expression plasmids. All expression plasmids were made using *BsaI* mediated Golden Gate assembly (24,62) in the pGGKd012 GFP dropout plasmid. All expression cassettes and the part plasmids (or PCR fragments) used for construction are outlined in Supplementary Table S8F-G. Plasmids were verified by PCR and/or restriction analysis (Supplementary Table S13). Correct assembly of some of the expression plasmids was verified by Sanger sequencing (Baseclear).

Gibson assembly expression plasmids. For two expression cassettes, namely: *pRPS3-coAtCPR1-tIDH2* and *pSeTPI1-At4CL3-tSDH2*, the individual parts contained too many *BsaI/BsmBI* sites, making Golden Gate assembly impossible. Therefore, the individual parts were amplified with PCR primers containing homology flanks and the expression cassettes were constructed via Gibson Assembly (Supplementary Table S8H and S11).

## Strain construction

### Construction of strains harboring test NeoChrs

Test NeoChrs, NeoChr10 and NeoChr11 consisted of 43 fragments namely: 36 non-coding 2.5 kb *E. coli* DNA fragments, *CEN6/ARS4*, *ARS1*, *ARS417*, *mRuby2*, *mTurquoise2*, *HIS3* and two TeSS fragments (Supplementary Table S14). The TeSS fragments were amplified from the telomerator plasmid pLM092 (12). To prevent possible circularization of the linear NeoChr, in NeoChr11 the sequence of the right TeSS was changed by modifying a few base pairs (PCR amplification with a mutated primer). Fragments were amplified by PCR using primers with 60 bp SHR flanks (63) (Supplementary Table S15). NeoChrs were assembled by transforming 200 fmol of each non-coding 2.5 kb *E. coli* DNA fragment, 200 fmol of each fluorescent marker, 200 fmol of the TeSS fragments, and 100 fmol of *CEN6/ARS4*, *HIS3* and the two *ARS* fragments. Transformants were checked by FACS, CHEF and long-read Nanopore sequencing. One transformant (NeoChr10.13) missing only 2113 bp of fragment 7A was stocked as strain IMF22.

### Construction of strains harboring NeoChrs designed for anthocyanin production

Construction of the host strain IMX2770. Before assembly of coding NeoChrs a suitable starting strain was engineered by several rounds of CRISPR/Cas9 gene deletions as described by Mans *et al.* (37). As parental strain, the SwYG strain IMX589 from Kuijpers *et al.* (4) was used. In this strain the minor paralogs of glycolysis are deleted and the major paralogs are centralized at the *sga1* locus on chromosome IX. From this strain, the *amdSYM* marker located between the major paralogs of glycolysis was deleted using *in vivo* assembly of a pMEL10 gRNA plasmid backbone (amplified with primer 6005, Supplementary Table S16) and a gRNA insert made from annealing primers (11588 & 11589, Suppl. Table S15). The DSB was repaired with a 120 bp repair fragment homologous to the flanking SHRs K and L, made by annealing of complementary primers (11590 & 11591, Supplementary Table S16), resulting in strain IMX1433 and IMX1769 before and after plasmid recycling, respectively. Subsequently the minor paralogs of the pentose phosphate pathway, *GND2*, *NQM1*, *SOL4* and *TKL2*, were deleted by transformation of two gRNA plasmids (pUDR286 & pUDR590, Supplementary Table S8A) and 120 bp repair fragments (Supplementary Table S17) homologous to the 60 bp upstream and downstream of the ORF. The strains were stocked before and after discarding the gRNA plasmids, resulting in respectively IMX2154 and IMX2204. Next as much as possible of the promoter, gene and terminator of the *ura3* and *his3* as well as of the functional *SpHIS5* gene were removed using gRNA plasmids pUDR426 and pUDR546 and repair fragments (Supplementary Tables S8A and S18) obtaining strain IMX2234 after plasmid recycling. Finally, in the last round of deletion, the *ARO10* gene was removed with gRNA plasmid pUDR406 and a 120 bp repair fragment (Supplementary Tables S8A and S19). Again, the plasmid was removed

and the strain was stocked as IMX2270. Assembly of NeoChr25 and NeoChr26. In strain IMX2270 the circular NeoChr26 and the linear NeoChr25 were assembled. The fragments for assembly were identical between the two NeoChrs namely: 7 pentose phosphate pathway genes (*ZWF1*, *TKL1*, *GND1*, *RKI1*, *TAL1*, *RPE1*, *SOL3*), 13 glycolysis genes (*HXK2*, *PGK1*, *FBA1*, *TPI1*, *TDH3*, *GPM1*, *ENO2*, *PGI1*, *PYK1*, *PDC1*, *ADH1*, *PFK1*, *PFK2*), 10 codon optimized *E. coli* shikimate pathway genes (*coEcAroG<sup>ibr</sup>*, *coEcAroB*, *coEcAroE*, *coEcAroL*, *coEcAroA*, *coEcAroD*, *coEcAroC*, *coEcTyrA<sup>ibr</sup>*, *coEcPheA<sup>ibr</sup>*, *coEcTyrB*), *HIS3* selection marker, two fluorescent markers (*mRuby2*, *mTurquoise2*), *CEN6/ARS4* and two ARS sequences. For NeoChr26 (circular), the telomerator fragment was transformed, while for NeoChr25 (linear) two TeSS were used instead. The fragments were amplified by PCR from the corresponding expression cassettes for all transcriptional units or from genomic DNA and other plasmids for auxiliary parts using primers with 60 bp SHR flanks (Supplementary Table S20-S21). For each fragment 200 fmol was used, with the exception of *HIS3* and *CEN6/ARS4* fragments, where 100 fmol were added instead. The pooled transformation mix were concentrated with Vivacon 500 (Sartorius AG, Gottingen, Germany) up to a final volume of 50  $\mu$ l and transformed in IMX2270. Both the NeoChr25 (linear) and NeoChr26 (circular) transformants were screened by FACS and long-read Nanopore sequencing, and NeoChr25 was additionally also verified by CHEF. After sequence confirmation using short-read whole genome sequencing, the strains were stocked as IMF27 (NeoChr25) and IMF29 (NeoChr26). Removal of single locus glycolysis from native chromosome. In IMF29 (circular NeoChr) the glycolytic cassette was removed from the *sga1* locus by induction of two double strand breaks (DSB) in the flanks of the cassette with the gRNA plasmid pUDR413, and providing a 120 bp repair fragment homologous to the upstream and downstream region of *sga1* (Supplementary Table S8 and S22). For IMF27 (linear NeoChr) the glycolytic cassette was replaced by a *KIURA3* transcriptional-unit by providing a *KIURA3* expression unit as repair fragment, obtained by PCR amplification from pMEL10 with primers containing flanks homologous to *sga1* locus (Supplementary Table S22). After discarding the gRNA plasmids the strains were stocked as IMF31 (linear NeoChr) and IMF32 (circular NeoChr). The circular NeoChr in strain IMF32 contained a mutation in the *RKI1* ORF. The mutation was repaired by inducing a DSB in the vicinity of the mutation with the gRNA plasmid pUDR756, targeting only the watermarked gene but not the native copy, and repairing it with a 120 bp repair fragment (Supplementary Table S8 and S23). The repair fragment contained the non-mutated sequence and a silent mutation of the PAM sequence. A correct transformant was confirmed by Sanger sequencing (Baseclear) and named IMF35. Deletion of major PPP paralogs from their native chromosomal loci. In a first transformation round, *ZWF1*, *SOL3*, *GND1* and *RKI1* were removed from IMF31 and IMF35 using the gRNA plasmids pUDR703 and pUDR700 and 120 bp repair fragments (Supplementary Table S6 and S24). Plasmid removal resulted in strains

IMF33 and IMF36 with linear and circular NeoChr, respectively. In a second transformation *RPE1*, *TKL1* and *TAL1* were removed using the gRNA plasmids pUDR701 and pUDR702 and 120 bp repair fragments (Supplementary Table S7 and S24). After plasmid recycling, the strains were stocked as IMF34 and IMF40, for the linear and circular NeoChr strains respectively. Integration of plant anthocyanin pathway. The next step in strain construction was the integration of single copies of the plant genes of the anthocyanin pathway on the NeoChrs, *AtPAL1*, *coRcTAL1*, *coAtC4H*, *coAtCPR1*, *At4CL3*, *AtCHI1*, *coAtCHS3*, *coAtF3H*, *coGhDFR*, *coAtANS*, *coAt3GT*, under expression of strong, constitutive yeast promoters. The fragments were amplified from their corresponding Golden Gate expression plasmids with primers containing 60 bp SHRs for *in vivo* assembly (Supplementary Table S25-S26). They were integrated in a single locus of the linear (strain IMF34) and circular (strain IMF40) NeoChrs by induction of a DSB in the *E. coli* fragment 16AB using gRNA plasmid pUDR765 (Figure 3, Supplementary Table S7). The two outer fragments contained 60 bp homology to a stretch of 60 bp in the *E. coli* fragment 16AB and in the flanking SHR DL. After verification of correct integration by PCR (Supplementary Table S26), the gRNA plasmid was removed by unselective growth on SMD and the strains were stocked as IMF41 (circular NeoChr) and IMF42 (linear NeoChr). Integration of multiple copies of coAtCHS3. To integrate multiple copies of *coAtCHS3*, the *coAtCHS3* expression unit was amplified from pUDC352 using primers adding 60 bp homology flanks to the integration sites Chunk 7BC, chunk 15CD, SHR N and chunk 9CD in the linear NeoChr31 and *CAN1*, *X2*, *YPRCtau3* and *SPR3* in native chromosomes (Supplementary Table S6, S7 and S27). The *coAtCHS3* expression units carrying homology with native chromosomes were transformed to IMF42 using the gRNA plasmids pUDR771 and pUDR772 for targeted editing. After recycling of the gRNA plasmids, strain IMF44 (Lin, 5x *CoAtCHS3* (4 copies in native genome)) was obtained. IMF44 was then transformed with suitable *coAtCHS3* cassettes and gRNA plasmids pUDR780 and pUDR781 targeting four loci in NeoChr31 resulting in strain IMF47, containing nine copies of the *coAtCHS3* (five on NeoChrs and four on native chromosomes). coAtANS repair. *coAtANS* was repaired by integration of a correct copy of *coAtANS* in the *mTurquoise2* gene of NeoChr33 in IMF47. The sequence of pUDC398 (*pSeENO2-coAtANS-tFUM1*) was verified and the transcriptional unit was amplified using primers 18740/18741 (Supplementary Table S28) adding flanks with homology to *mTurquoise2* on NeoChr33. IMF47 was transformed with the template DNA and pUDR400 (gRNA-*mTurquoise2*) for targeting *mTurquoise2*. After PCR confirmation (Supplementary Table S28), pUDR400 was recycled and the final strain was stocked as IMF48 (linear NeoChr34, 9x *coAtCHS3*, correct *coAtANS*).

#### Strain verification by PCR and whole genome sequencing

All strains constructed by CRISPR-Cas9 were verified by diagnostic PCR.

In addition, whole genome short read sequencing was performed in-house for IMF27 and IMF29 on an Illumina MiSeq Sequencer (Illumina, San Diego, CA) as described previously (5,32). Strains IMF41, IMF42, and IMF47 were sequenced on a NovaSeq 6000 at GenomeScan Leiden (GenomeScan, Leiden, NL).

Long-read sequencing of IMF22, IMF27, IMF29, IMF41, IMF42, IMF47 and IMF48 was performed in-house on MinION flow cells using the SQK-LSK109 sequencing kit with the EXP-NBD104 expansion kit (Oxford Nanopore Technologies, Oxford, United Kingdom). Average DNA size and integrity were verified with the TapeStation 2200 (Agilent Technologies, Santa Clara, CA). Before sequencing, flow cell quality was assessed by running the MinKNOW platform QC. Samples NeoChr10.10, NeoChr10.13 (IMF22), Neochr10.47, NeoChr10.54, NeoChr10.16, NeoChr10.62, NeoChr10.67, NeoChr10.69, NeoChr11.19, NeoChr11.22, NeoChr25.25, NeoChr25.47, NeoChr25.53, NeoChr25.56 (IMF27), NeoChr26.2, NeoChr26.4 (IMF29), NeoChr26.6, Neochr26.9 and NeoChr26.1 were sequenced on a FLO-MIN106 flow cell with sequencing kit SQK-LSK108. Samples NeoChr12 (IMF23), NeoChr30 (IMF41), NeoChr31 (IMF42), NeoChr33 (IMF47) and NeoChr34 (IMF48) were sequenced on a FLO-MIN111 with sequencing kit SQK-LSK109. Basecalling was performed for samples with NeoChr10 and Neochr11 by using Albacore (version 2.3.1, Oxford Nanopore). Demultiplexing of the fastq files of the NeoChr10 and NeoChr11 samples was performed with Porechop (<https://github.com/rrwick/Porechop>). Basecalling and demultiplexing was performed with Guppy (Oxford Nanopore) for samples with NeoChr25 and NeoChr26 with version 3.1.5, samples IMF41, IMF42, IMF47 with version 4.4.2 and IMF48 with version 4.5.4. All resulting fastq files were filtered on length (> 1kb) followed by *de novo* assembly by Canu version 2.0 (64).

The sequences were deposited to NCBI as a BioProject under the accession number PRJNA738851.

### **Characterization of NeoChrs fidelity, stability and toxicity.**

Fluorescence-based sorting by flow cytometry. Flow cytometry employing an BDFACSAria™ II SORP Cell Sorter (BD Biosciences, Franklin Lakes, NJ), was used to sort single cells in the chromosome stability experiment, as well as for screening of fluorescent transformants bearing the constructed NeoChrs as both described earlier by Postma *et al.* (5). Karyotyping with CHEF analysis. Chromosomes separation and size determination was performed using contour-clamped homogeneous electric field electrophoresis as described earlier by Postma *et al.* (5). Quantification of strain viability and NeoChr stability. NeoChr stability was assessed based on plating cells from shake flask cultures on selective and non-selective growth medium with respect to the NeoChr as described by Postma *et al.* (5). On the second- and fifth-day strain viability and stability was measured, on the third and fourth day, only the cell density

(OD<sub>660</sub>) of overnight cultures was measured. For the test-chromosome bearing strains (IMF22 and IMF23) viability and stability was based on sorting of 96 cells, while for CEN.PK113-7D, IMF27, IMF29, IMF34, IMF40, IMF41, IMF42 and IMF47, 384 cells were sorted on a microtiter plate. Determination of specific growth rates. For determination in shake flasks, strains were inoculated from a -80°C freezer stock in a 500 mL shake flask with 100 mL medium and grown until late-exponential phase. Cells were transferred to fresh 100 mL medium and grown until mid-exponential phase. Finally, from these cultures, shake flasks containing 100 mL medium were inoculated at an OD<sub>660</sub> of 0.3. Cell density was measured with a 7200 Visible spectrophotometer (Cole-Parmer, Staffordshire, United Kingdom). A maximum specific growth rate ( $\mu_{\max}$ ) was calculated from at least biological duplicates, using five data points and at least two doublings in the exponential phase. For growth rate determination in multi titer plates, a Growth Profiler (EnzyScreen, Heemstede, NL) was used as previously described (5) using plate 06 and constants: a: 0.084327, b:  $5.35 \times 10^{-8}$ , c: 4.398348, d: -0.41959. Growth rates were calculated from six biological replicates, using ODs between 1 and 8.

### **Characterization of aromatics production by the engineered strains**

Shake flasks. The strains were inoculated in biological triplicates in 500 mL shake flasks containing 100 mL synthetic medium with 20 g L<sup>-1</sup> glucose. Urea was used as sole nitrogen source to prevent acidification of the medium, thereby allowing respiration of the produced ethanol. After 72 hours of cultivation, the optical density was measured using a Jenway 7200 spectrometer (Jenway, Staffordshire, United Kingdom) at 660 nm and aromatics were quantified as described below. Aerobic batch bioreactors. Strains IMF41, IMF42 and IMF48 were characterized in bioreactors. 2L bioreactors (Applikon, Delft, The Netherlands) were filled with 1.0 L synthetic medium, containing 20 g L<sup>-1</sup> glucose and ammonium sulfate as nitrogen source. Exponentially growing cells were used to inoculate the bioreactors at an initial biomass concentration of around 0.12 g L<sup>-1</sup>. The cultivations were performed at 30 °C, 800 RPM using 0.5 L min<sup>-1</sup> pressurized air to sparge the bioreactor with oxygen. Automated addition of 2 M KOH and 3 M HCL ensured maintenance of the culture pH at 5.0. Samples were taken at regular time intervals for optical density, metabolite concentrations and cell dry weights determinations. Cell dry weight, organic acids, sugars and ethanol concentrations were measured as previously described (62,65). Analysis of aromatics. Extracellular naringenin and upstream aromatics were quantified by HPLC as previously described (62). HPLC analysis of aromatic compounds up until naringenin. For extracellular aromatic compounds, a sample containing broth was mixed 1:1 with 96% ethanol, vortexed thoroughly, spun down for 5 minutes at 14800 RPM and the supernatant was used for further analysis. The aromatic compounds up until naringenin (2-phenylethanol (2PE), *p*-hydroxyphenylethanol

(*p*OH2PE), phenylacetic acid (PAA), *p*-hydroxyphenylacetic acid (*p*OHPPAA), phenylpyruvic acid (PPY), coumaric acid (COUM), cinnamic acid (CIN), phloretic acid (PHLOR) and naringenin (NAR) were measured using an Agilent Zorbax Eclipse plus C18 column (4.6 x 100mm, 3.5  $\mu$ m) (Agilent). As mobile phase, 0.020 M  $\text{KH}_2\text{PO}_4$  set at pH 2.0 containing 1% acetonitrile was used at a flow rate of 0.8 mL  $\text{min}^{-1}$  at an operating temperature of 40 °C. The amount of acetonitrile was gradually increased to 10% within 6 minutes, then to 40% after 23 minutes, followed by a decrease in amount to 1% after 30 minutes. The compounds were detected using a diode array and a multiple wavelength detector (Agilent G1315C) at different wavelengths: 200 nm for PAA, 210 nm for PPY, 214 nm for 2PE, *p*OH2PE, *p*OHPPAA and PHLOR, 270 nm for CIN and finally 280 nm for NAR and COUM. The extracellular concentrations in the supernatant of the aromatic compounds kaempferol (KEA), dihydrokaempferol (DHK), kaempferol 3-O-glucoside (K3G), pelargonidin (PEL) and pelargonidin 3-O-glucoside (P3G) were detected using LS-MS, as described in the next section. Additionally, since P3G has never been measured extracellular before, the intracellular concentrations of P3G, and its precursors kaempferol, dihydrokaempferol, K3G and pelargonidin were also measured. A certain amount of cell culture was spun down for 5 minutes at 5000 RPM, washed once with  $\text{dH}_2\text{O}$ , resuspended in 0.5-1 ml methanol (0.75% HCL) and the samples were stored overnight at -80 °C. Next, the samples were lyophilized for 24 h using a Mini Lyotrap freeze-dryer (LTE Scientific TLD, UK) operated at -80 °C, connected to a Pirani 501 manometer (Edwards Vacuum, UK) using a RV8 pump (Edwards Vacuum, UK). Finally, the pellet was resuspended in 1 mL methanol (2.0% HCL) and stored overnight at -80 °C. Mass spectrometric analysis of anthocyanin pathway compounds. Identification and quantification of compounds from the anthocyanin pathway downstream of naringenin was performed using an ACQUITY UPLC chromatography system (Waters, UK) coupled online to a high-resolution Orbitrap mass spectrometer (Q-Exactive Focus, Thermo Fisher Scientific, Germany). For chromatographic separation, a reverse phase separation column (ACQUITY UPLC BEH C18, 1.0 mm  $\times$  100 mm, 3  $\mu$ m particle size, part No 186002346, Waters UK) was used at room temperature using  $\text{H}_2\text{O}$  plus 0.1% formic as mobile phase A, and acetonitrile plus 0.1% formic acid as mobile phase B. A gradient was maintained at 50  $\mu\text{L}/\text{min}$  at 7.5% B over 5 minutes. Solvent B was then increased to 80% over 4 minutes, and kept constant for additional 3 minutes before equilibrating back to the starting conditions. The metabolite extracts were taken from -80°C immediately before injection, brought to room temperature, vortexed and 15  $\mu\text{L}$  crude extract were mixed with 85  $\mu\text{L}$  1mM HCL. The mixture was carefully vortexed and centrifuged using a bench top centrifuge for 1 minute to remove insoluble materials. 5  $\mu\text{L}$  were subsequently injected onto the UPLC reverse phase separation system. The mass spectrometer was operated alternating in full scan and PRM mode. Full scan was acquired from 250-700  $m/z$  in ESI positive mode (+3.25 kV), at a resolution of 70K. Parallel reaction

monitoring was performed for the precursor masses for dihydrokaempferol (DHK, Cas No. 104486-98-8) 289.07 m/z [M+H]<sup>+</sup> using a NCE of 26, kaempferol (KEA, Cas No. 520-18-3) 287.05 m/z [M+H]<sup>+</sup> using a NCE of 30, kaempferol 3-O-glucoside (K3G, Cas No. 480-10-4) 449.10 m/z [M+H]<sup>+</sup> using a NCE of 24, pelargonidin (PEL, Cas No. 134-04-3) 271.06 m/z [M]<sup>+</sup> using a NCE of 30 and pelargonidin 3-O-glucoside (Cas No. 18466-51-8) m/z 433.10 [M+H]<sup>+</sup> using a NCE of 24. Fragment ions were measured at fixed first mass of 75 m/z, a resolution of 35K, a max IT of 100ms and an AGC target of 1e5, by acquiring 2 microscans. Raw data were analyzed using XCalibur 4.1 (Thermo) where retention and unique fragments for each individual compound were compared to commercial standards. For quantification, peak intensities of identified compounds from the samples were summed using Matlab 2020b, and compared against an external calibration curve established using commercial standards. The standards were purchased from Sigma Aldrich (dihydrokaempferol Cat No. 91216, kaempferol Cat No. 60010, kaempferol 3-O-glucoside Cat No. PHL89237, pelargonidin chloride Cat No. PHL80084, pelargonidin 3-O-glucoside chloride Cat No. PHL89753). The mass spectrometer was calibrated using the Pierce™ LTQ ESI positive ion calibration solution (Thermo Fisher Scientific, Germany). Detection of P3G is illustrated in Supplementary Figure S12.

## Data availability

Genome sequence data are available at NCBI as Bioproject under the accession number PRJNA738851.

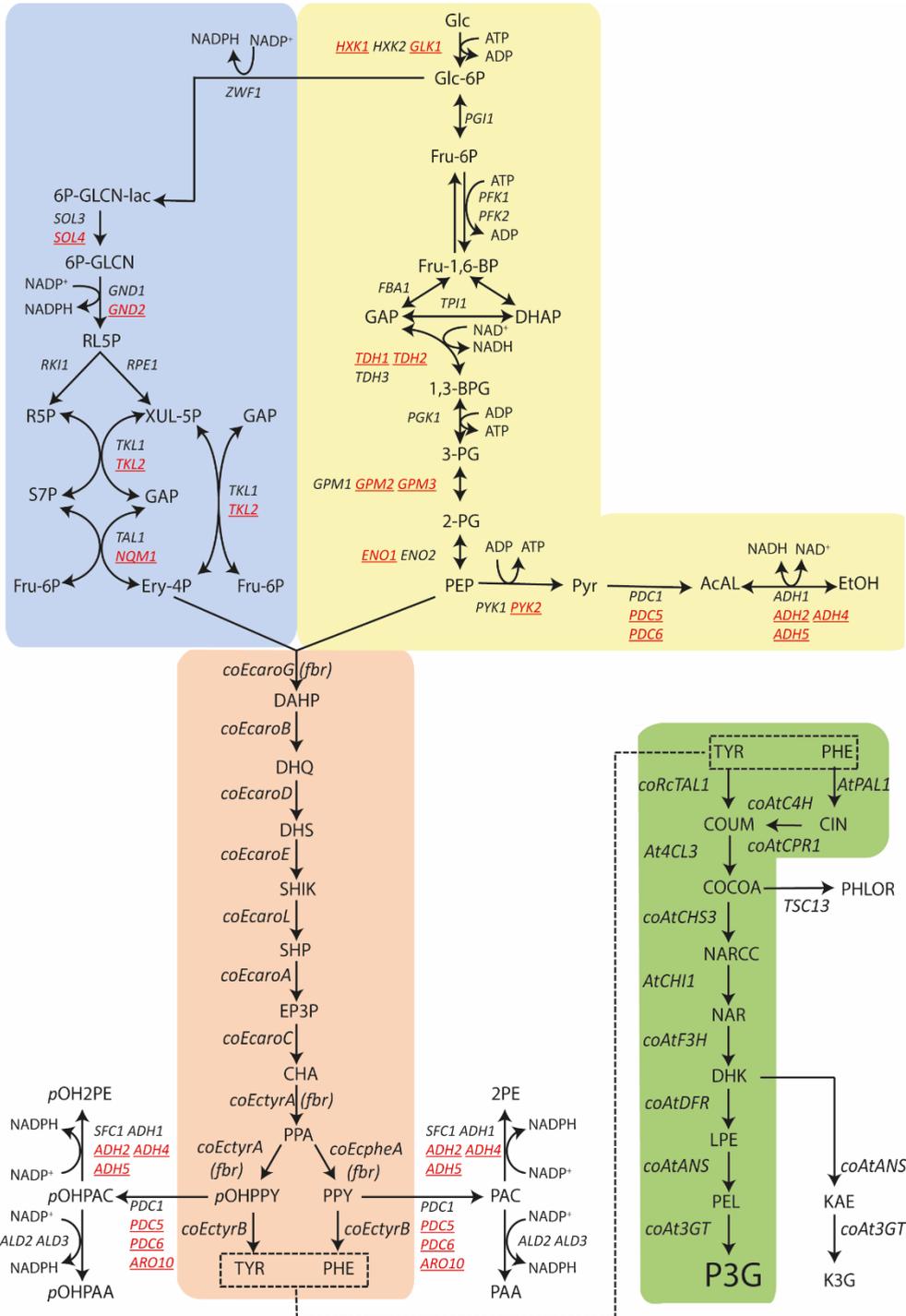
## Acknowledgements

We thank Carol de Ram for technical support for the analysis of metabolites by LC-MS.

## Supplementary data

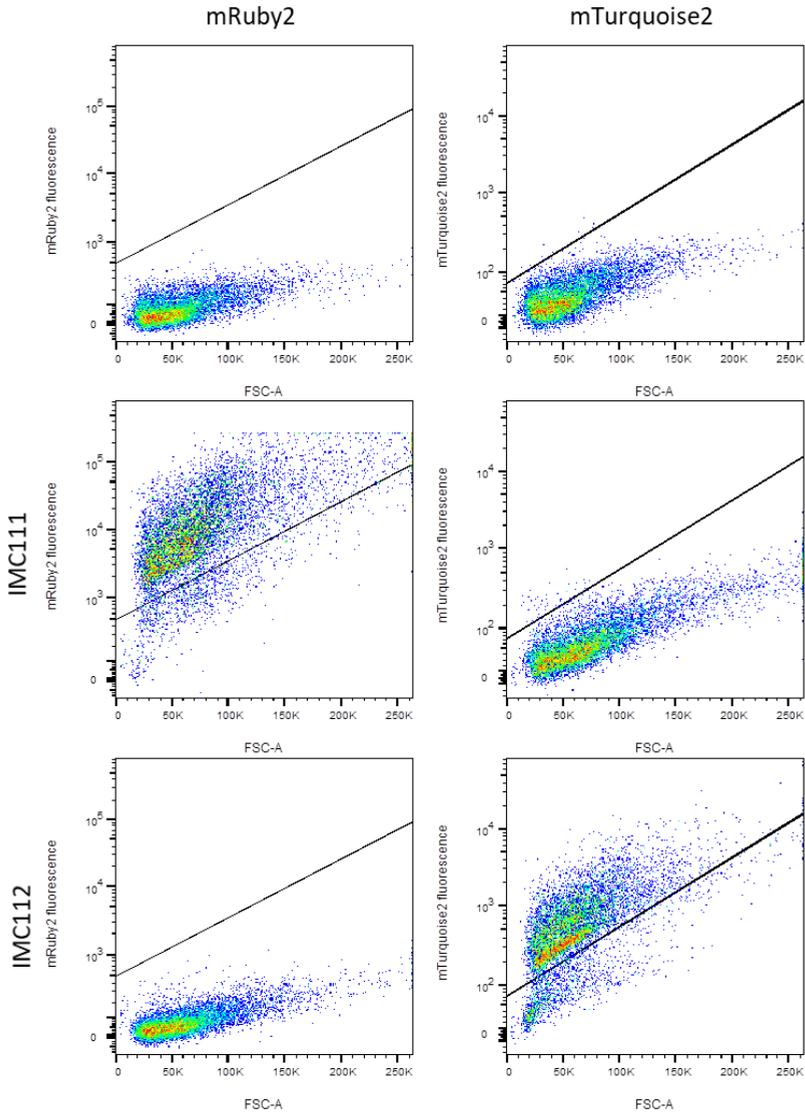
### Figure S1 Reactions from glucose to pelargonidin 3-O-glucoside

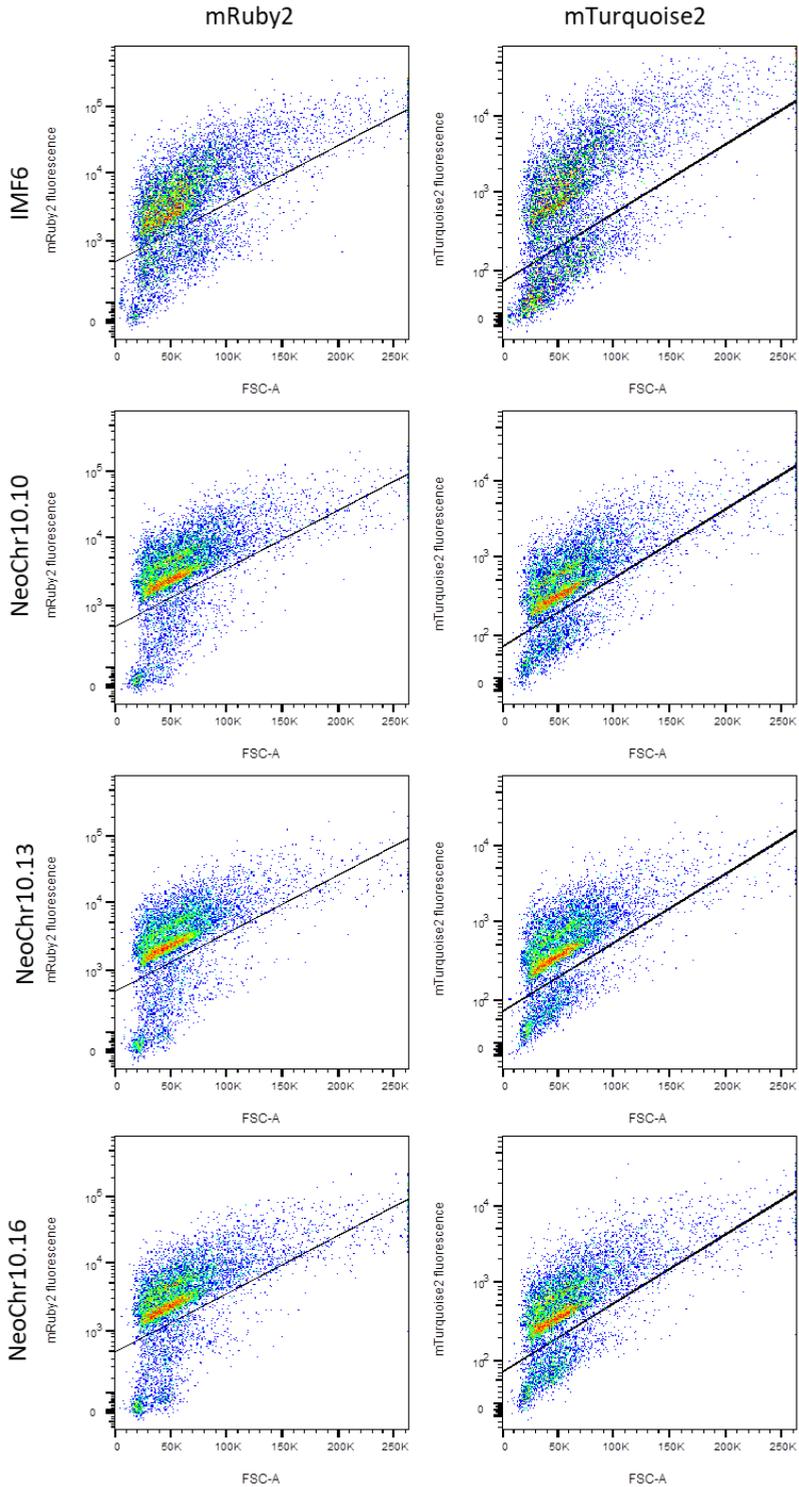
Yellow: glycolysis and ethanolic fermentation. Blue: pentose phosphate pathway. Brown: *E. coli* shikimate pathway. Green: plant anthocyanin pathway. The gene names encoding the enzymes involved in the indicated reactions are indicated in italics. Genes deleted in the present study are indicated in red and underlined. *Ec* is *E. coli*, *At* is *Arabidopsis thaliana*, *Rc* is *Rhodobacter capsulatus*, *co* is codon optimized. *Fbr* is feedback resistant, Glc is glucose, Glc-6P is glucose 6-phosphate, Fru-6p is fructose-6-phosphate, Fru-1,6-BP is fructose 1,6-bisphosphate, GAP is glyceraldehyde 3-phosphate, DHAP is dihydroxyacetone, 1,3-BPG is 1,3-bisphosphoglycerate, 3-PG is 3-phosphoglycerate, 2-PG is 2-phosphoglycerate, PEP is phosphoenolpyruvate, Pyr is pyruvate, AcAL is acetaldehyde, EtOH is ethanol, 6p-GLCN-lac is 6-phosphogluconolactone, 6p-GLCN is 6-phosphogluconate, RL5P is ribulose 5-phosphate, R5P is ribose 5-phosphate, S7P is sedoheptulose 7-phosphate, XUL-5P is xylulose 5-phosphate, GAP is glyceraldehyde 3-phosphate, Ery-4P is erythrose 4-phosphate, DAHP is 3-deoxy-D-arabino-heptulosonate-7-P, DHQ is 3-dehydroquinate, DHS is 3-dehydroshikimate, SHIK is shikimate, SHP is shikimate 3-phosphate, EP3P is 5-enolpyruvyl-shikimate 3-phosphate, CHA is chorismate, PPA is prephenate, PPY is phenylpyruvate, PAC is phenylacetaldehyde, 2PE is 2-phenylethanol, PAA is phenylacetic acid, PHE is L-phenylalanine, *p*OHPPY is *p*-hydroxyphenylpyruvate, *p*OH-PAC is *p*-hydroxyphenylacetaldehyde, *p*OH-2PE is *p*-hydroxyphenylethanol, *p*OH-PAA is *p*-hydroxyphenylacetic acid, TYR is tyrosine, COUM is coumaric acid, CIN is cinnamic acid, COCOA is coumaroyl-CoA, NARCC is naringenin chalcone, PHLOR is phloretic acid, NAR is naringenin, DHK is dihydrokaempferol, KAE is kaempferol, K3G is kaempferol 3-O-glucoside, LPE is leucopelargonidin, PEL is pelargonidin and P3G is pelargonidin 3-O-glucoside.

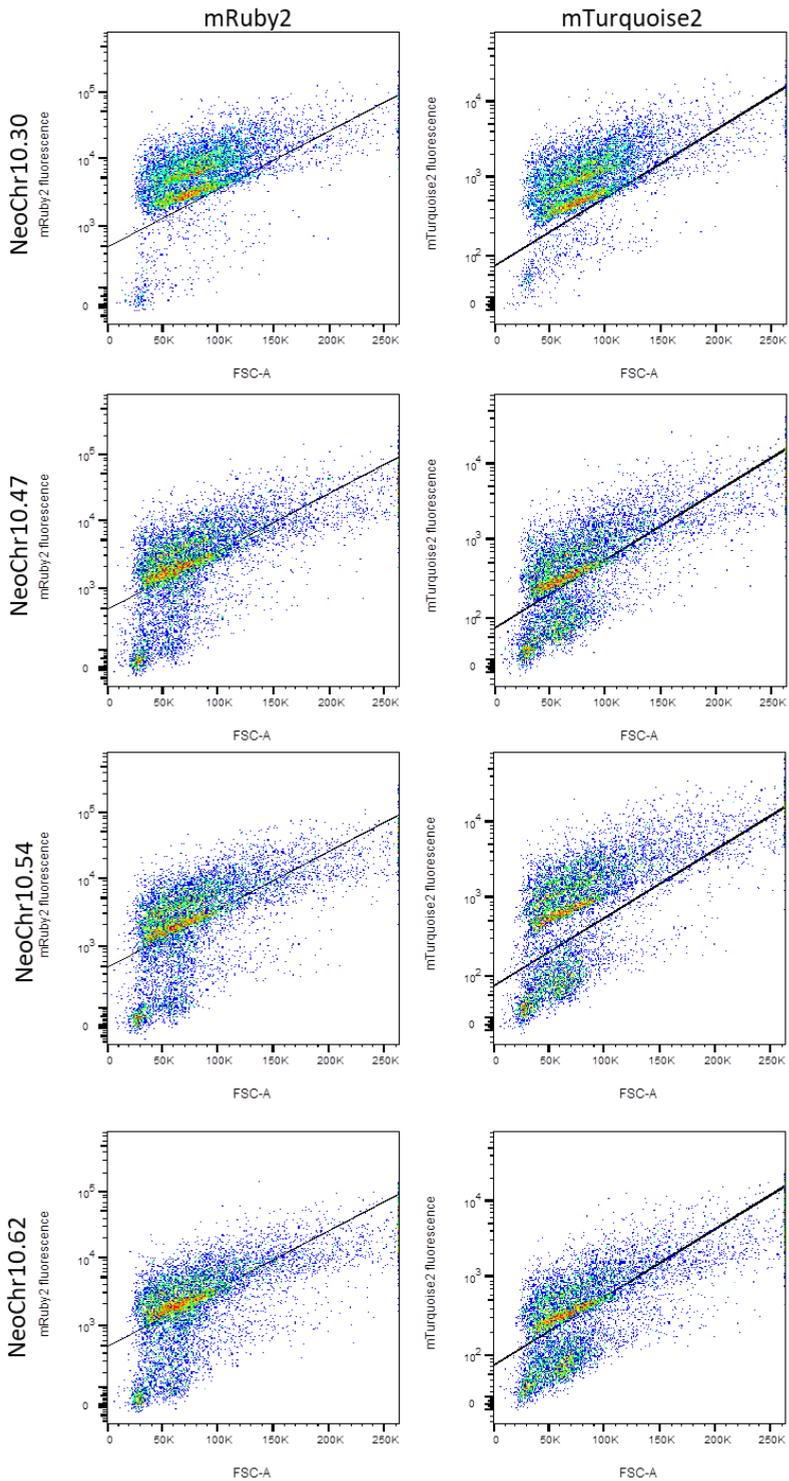


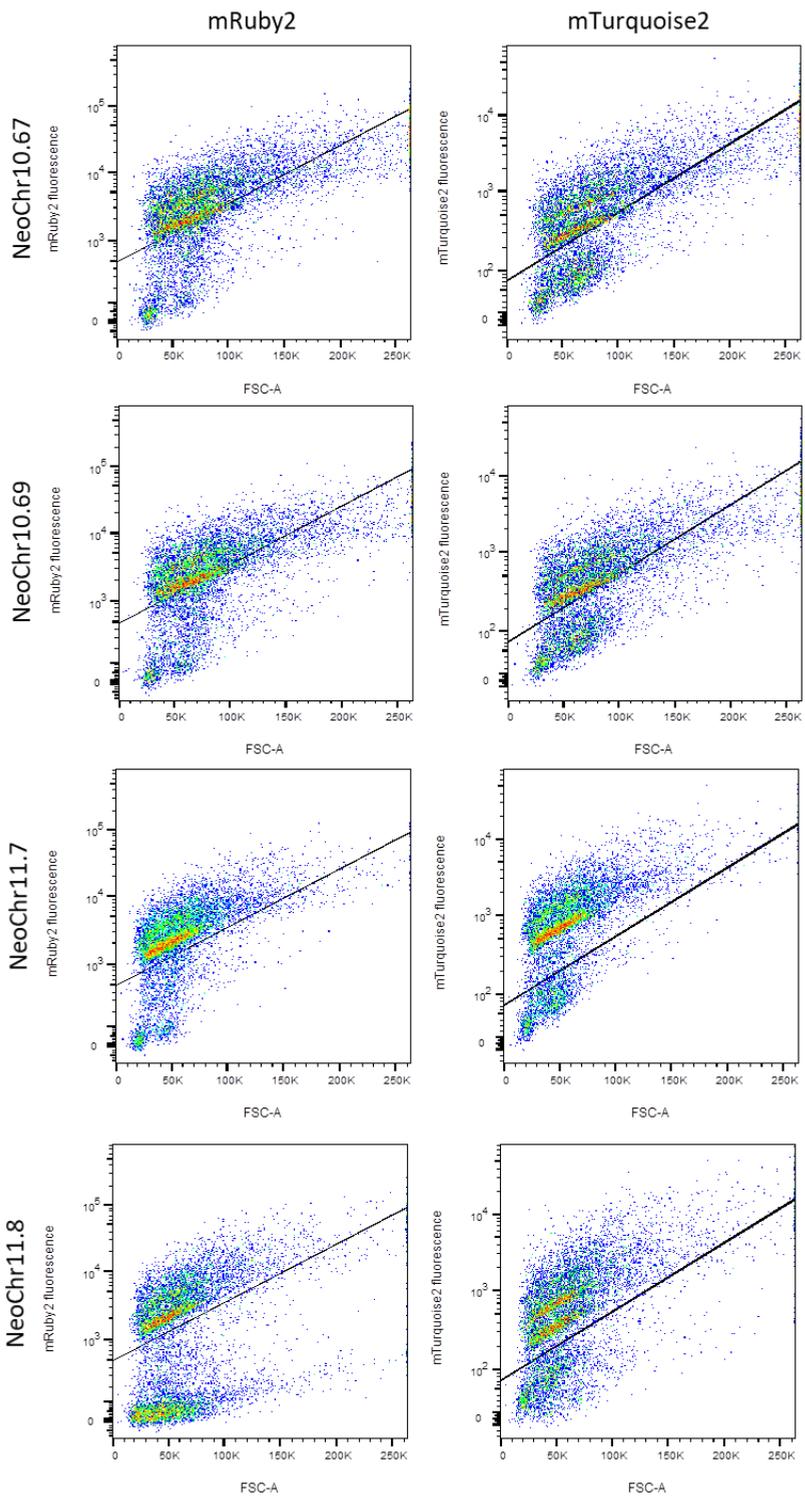
**Figure S2 Flow cytometric analysis of the test linear neochromosomes**

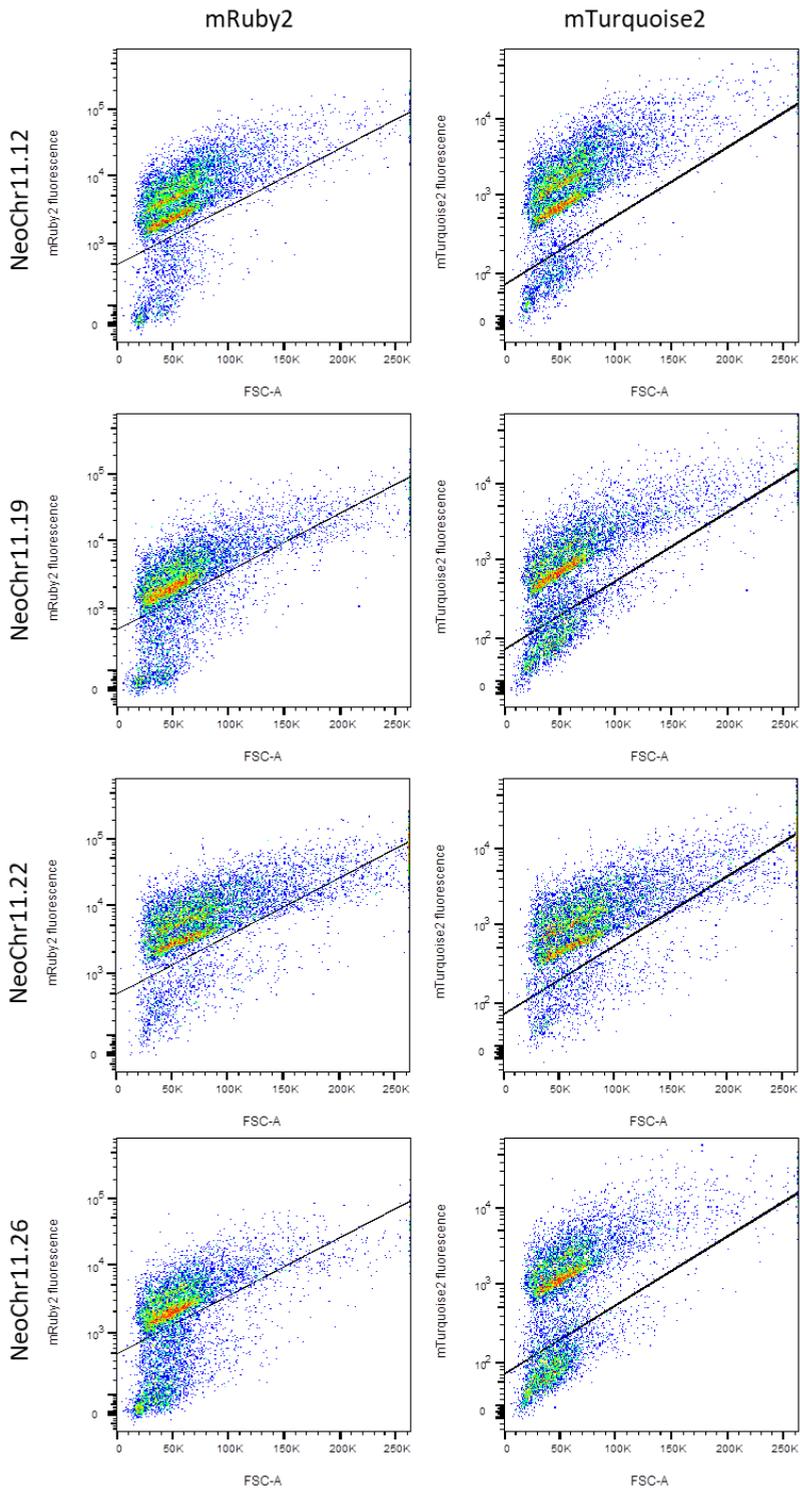
Cells from shake flask cultures were analysed by FACS. The fluorescence is plotted on the y-axis and the forward scatter (FSC-A) on the x-axis. Negative control: CEN.PK113-7D. Positive controls: IMC111 (mRuby2), IMC112 (mTurquoise2) and IMF6 (mRuby2 and mTurquoise2). Gates for fluorescence of the two different fluorescent proteins were drawn based on the IMC111 and IMC112 controls. Approximately 10000 events are shown for each plot.

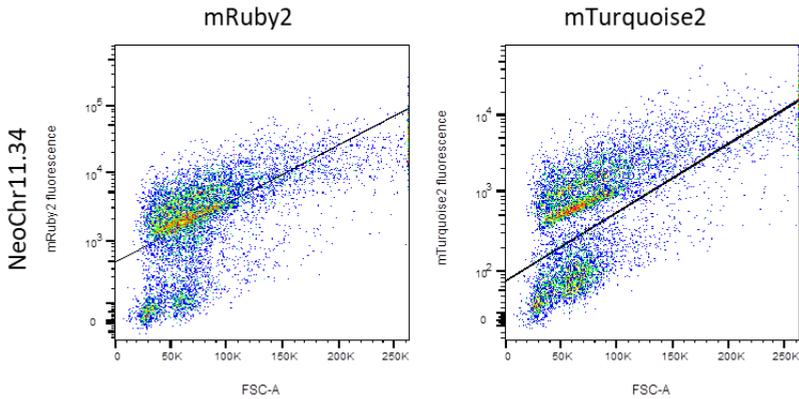






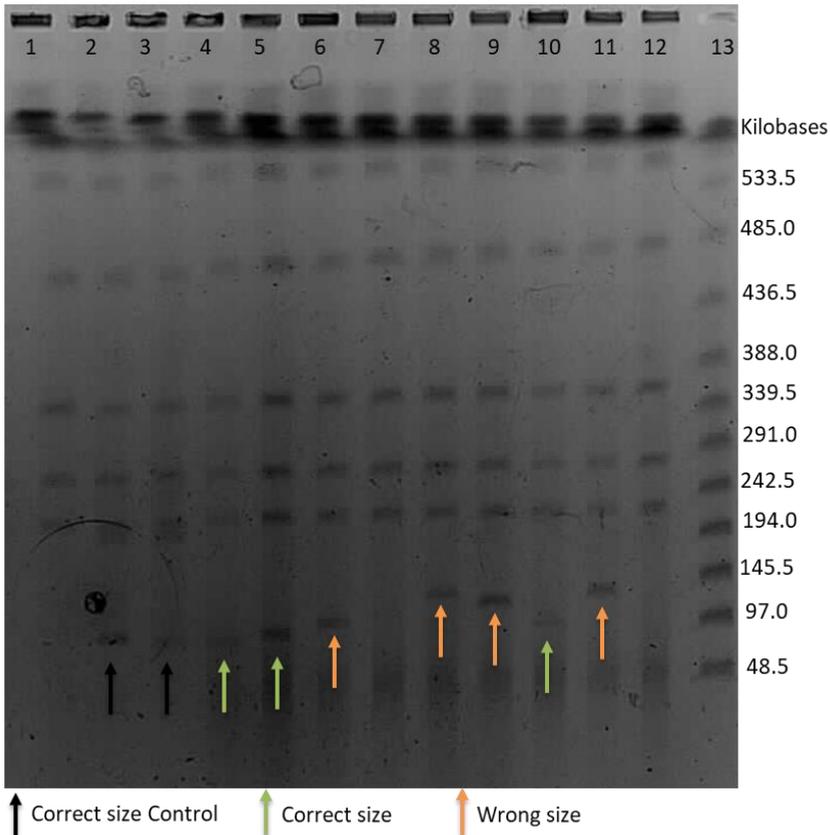




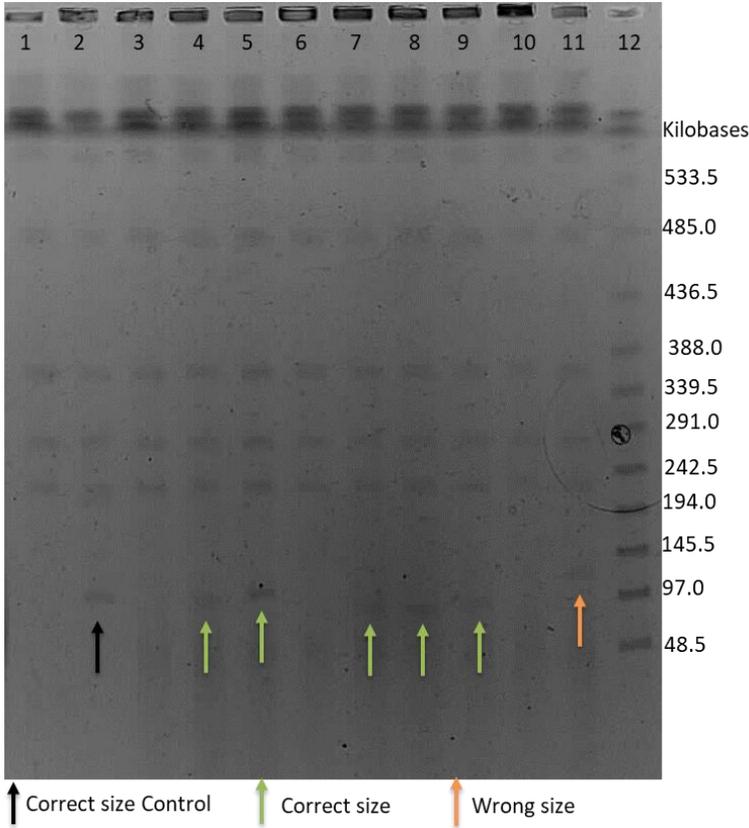


**Figure S3 Separation of the test linear neochromosomes on pulsed-field electrophoresis**

**A.** 1) IMX1338: control strain without neochromosome. 2) IMF6: control strain with 100 kb in plug linearized neochromosome NeoChr1. 3) IMF23: strain with 100 kb in plug linearized NeoChr12. 4) NeoChr10.10: correct size. 5) NeoChr10.13: correct size. 6) NeoChr10.16: wrong size. 7) NeoChr11.7: no visible neochromosome. 8) NeoChr11.8: wrong size. 9) NeoChr11.12: wrong size. 10) NeoChr11.19: correct size. 11) NeoChr11.22: wrong size. 12) NeoChr11.26: no visible neochromosome. 13) Size ladder



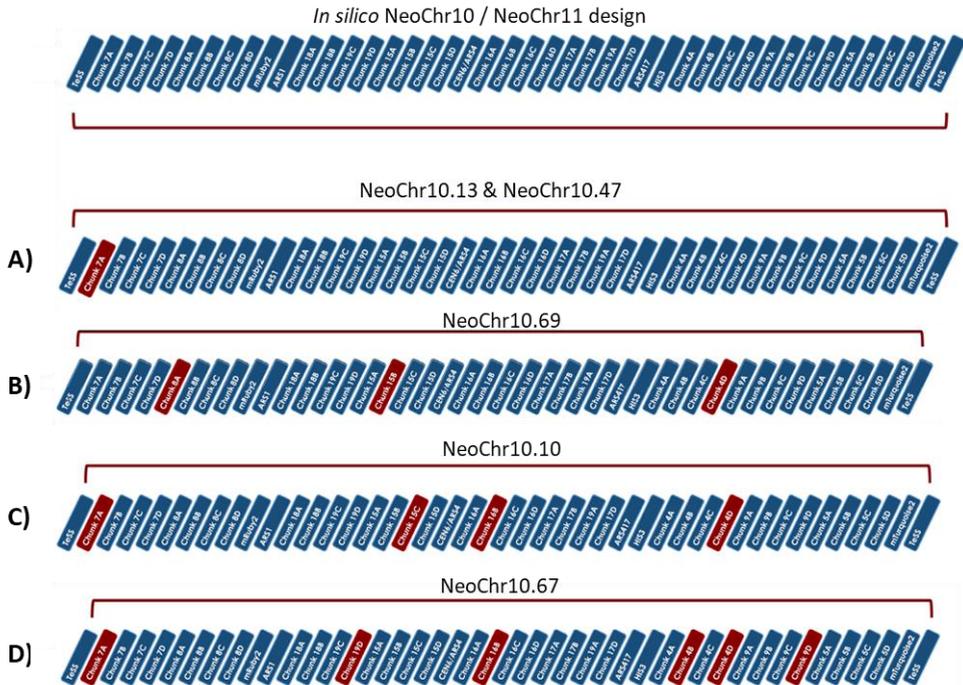
**B.** 1) IMX1338: control strain without neochromosome. 2) IMF6: control strain with 100 kb in plug linearized neochromosome (NeoChr1). 3) NeoChr10.30: no visible neochromosome. 4) NeoChr10.47: correct size. 5) NeoChr10.54: correct size. 6) NeoChr10.60: no visible neochromosome. 7) NeoChr10.62: correct size. 8) NeoChr10.67: correct size. 9) NeoChr10.69: correct size. 10) NeoChr11.29: no visible neochromosome. 11) NeoChr11.34: wrong size. 12) Size ladder



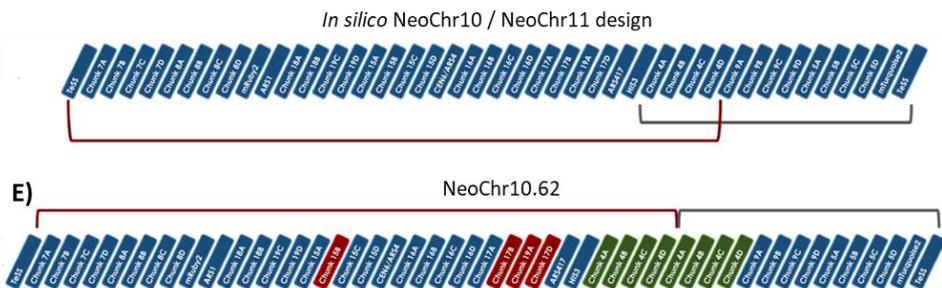
**Figure S4 Sequencing results of test linear neochromosomes**

*In silico* fragment configuration in NeoChr10 / NeoChr11 as well as *in vivo* fragment configuration of the neochromosome transformants as measured by long-read nanopore sequencing. The fragments of the *in silico* design which are present in the neochromosome transformants are connected by the same colored line. A dotted line indicates an area which is inverted. Fragments are color coded as follows: blue represents a correctly assembled fragment; red represents a missing fragment; green represents a duplicated fragment; and yellow indicates an inverted fragment.

A) NeoChr10.13 and NeoChr10.47 are missing an internal part of chunk 7A. B) NeoChr10.69 is missing 3 chunks: 8A, 15B and 4D. C) NeoChr10.10 is missing 4 chunks: 7A, 15C, 16B and 4D. D) NeoChr10.67 is missing 6 chunks: 7A, 19D, 16B, 4B, 4D and 9D



E) NeoChr10.62 is missing 4 chunks: 15B, 17B, 19A and 17D. In addition, a region containing the chunks 4A, 4B, 4C and 4D is duplicated

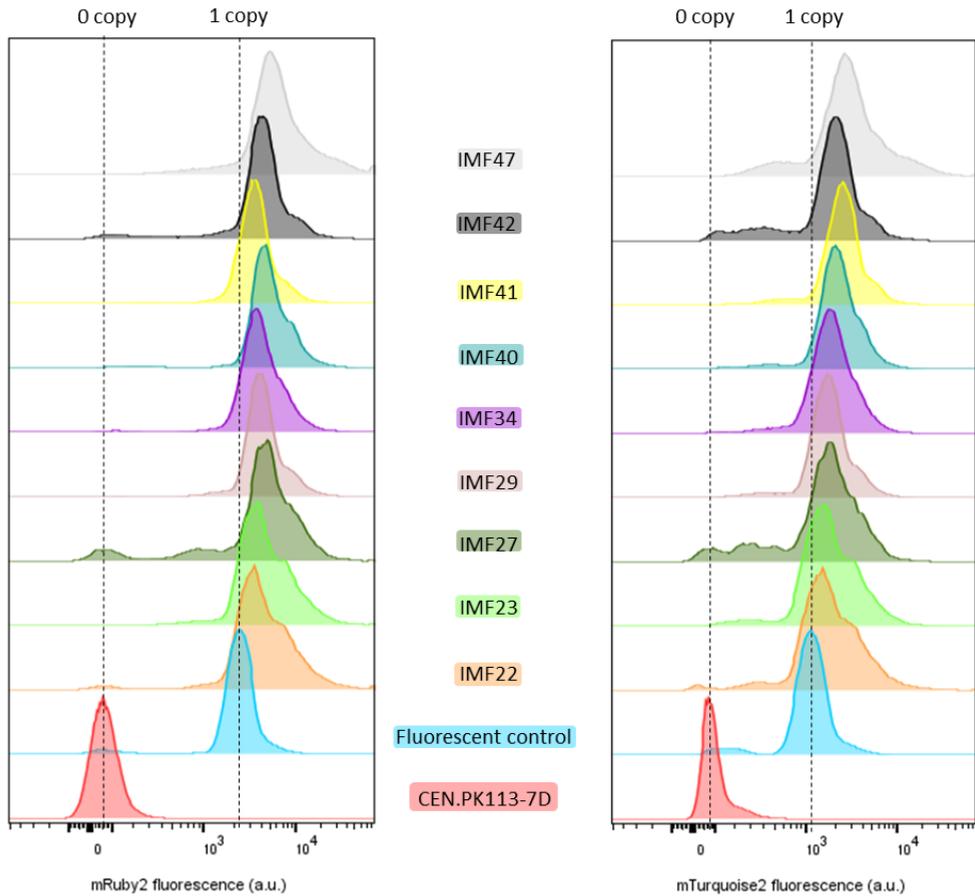


F) Neochr10.54 has a large inversion from 8D until 19A, from this region 2 chunks are missing: 15C and 18B. This region is linked to a region spanning from 19A (which is thus duplicated) until the right telomere. From this region chunk 17D is missing



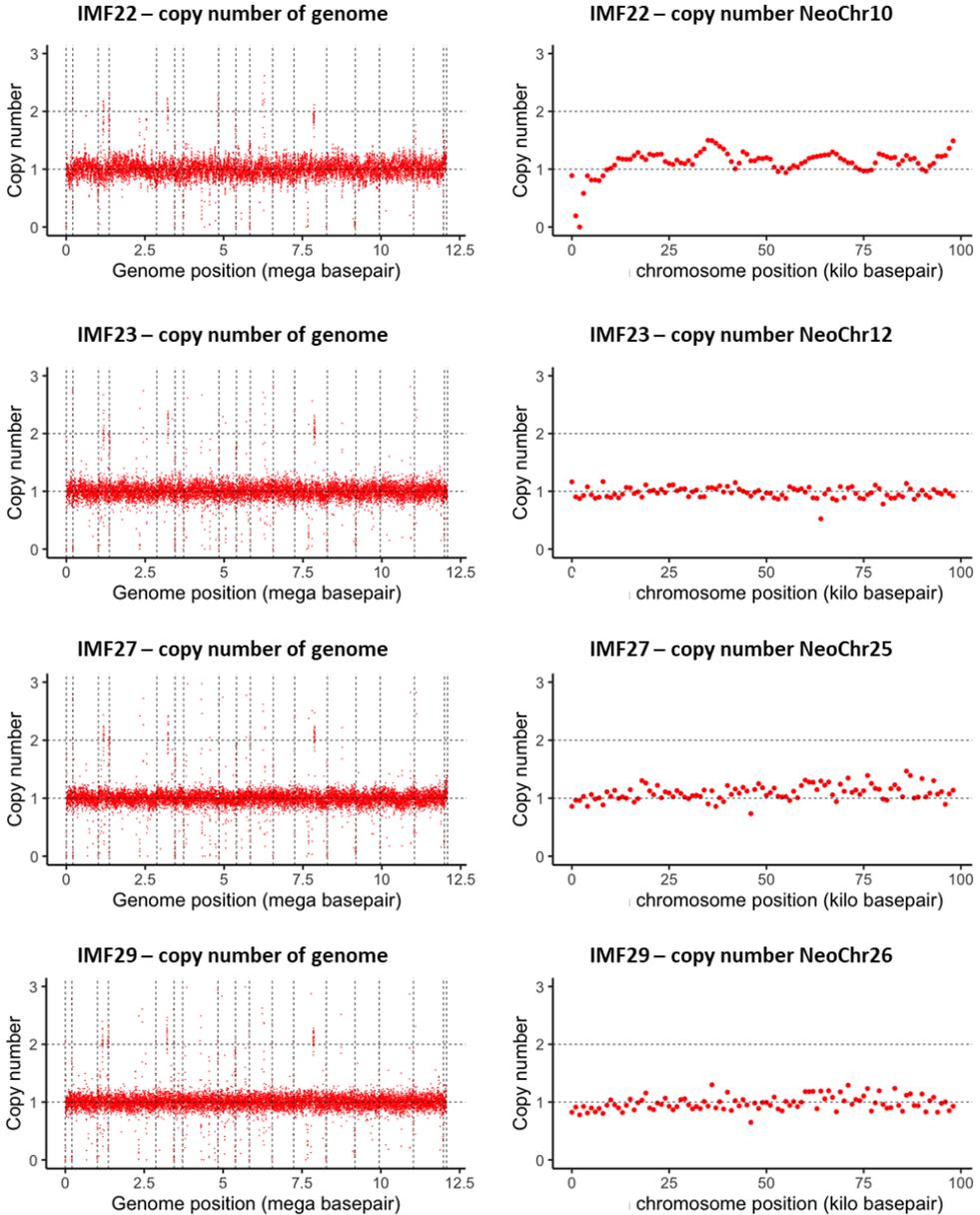
**Figure S5 NeoChr copy number estimation based on fluorescence**

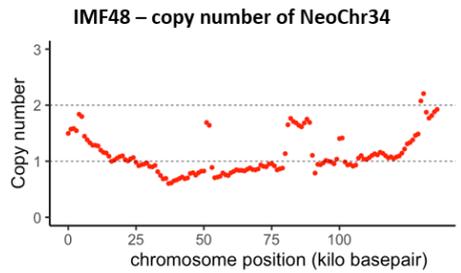
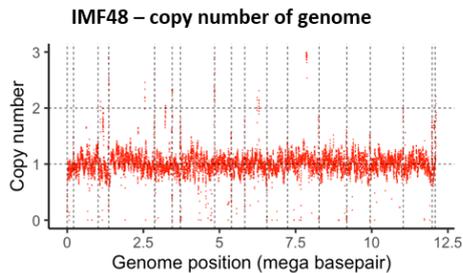
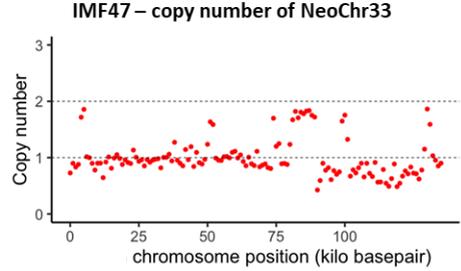
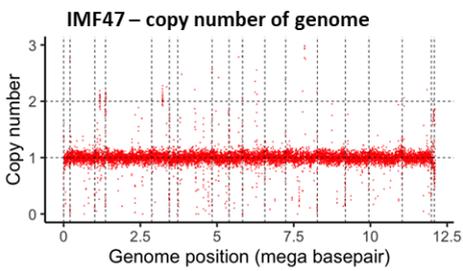
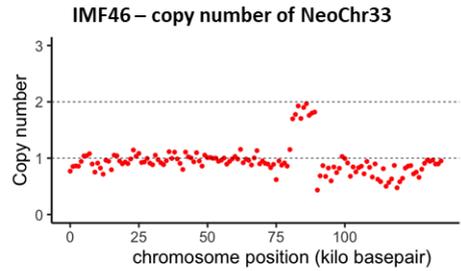
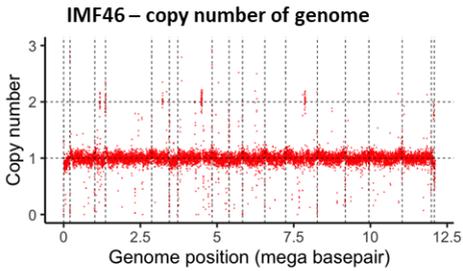
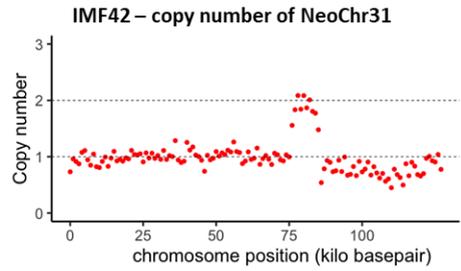
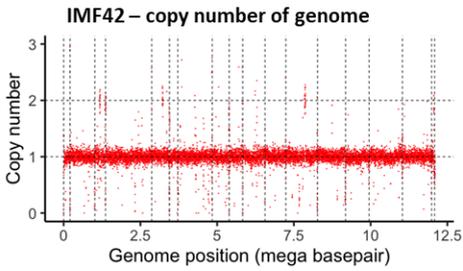
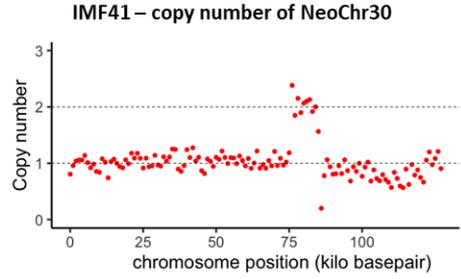
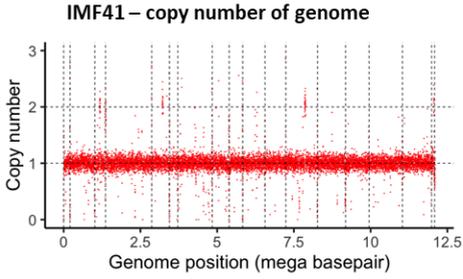
mRuby2 and mTurquoise2 fluorescence was measured by flow cytometry. CEN.PK113-7D with no fluorescent markers was used as negative control. IMX2224 and IMX2226 with a single copy of *mRuby2* and *mTurquoise2* integrated in the genome, respectively, were used as positive controls. All strains showed a fluorescence corresponding to the expected NeoChr copy number.



**Figure S6 NeoChr copy number estimation based on sequencing**

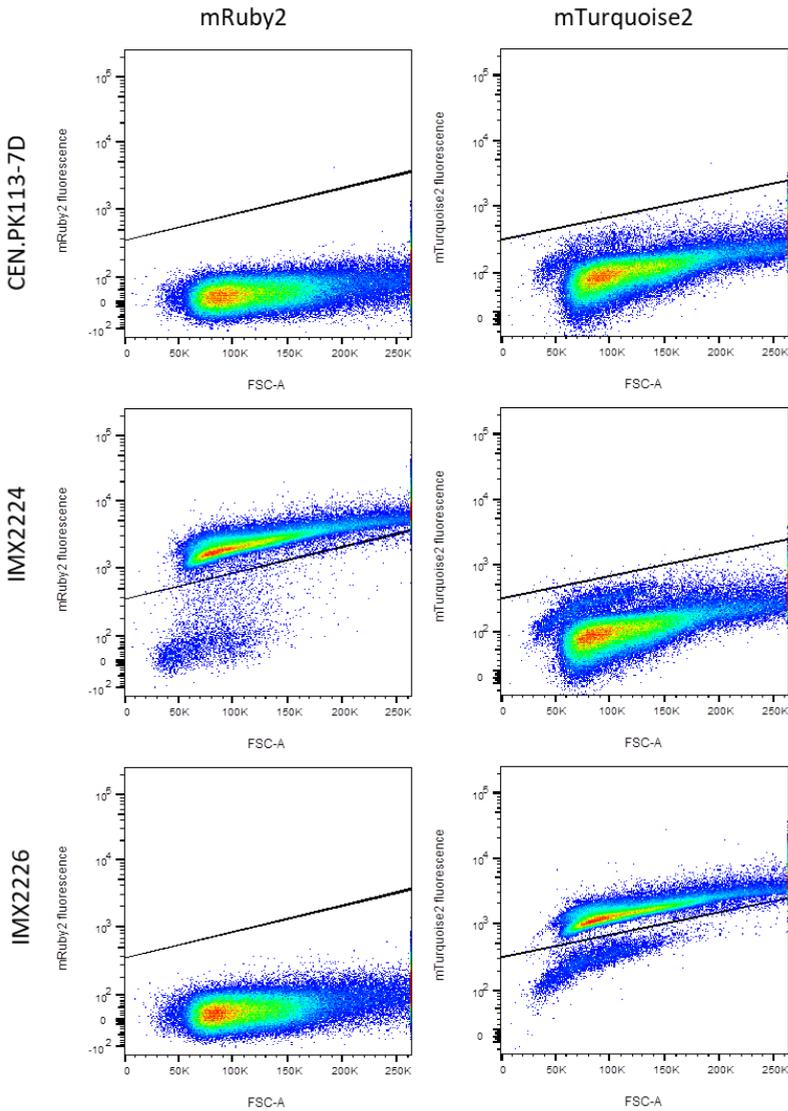
IMF22 and IMF48 were analyzed by long-read Nanopore sequencing and IMF23, IMF41, IMF42 and IFM47 by short-read Miseq sequencing. Plots on the left represent the copy number of native chromosomes, while plots on the right show the NeoChrs copy number

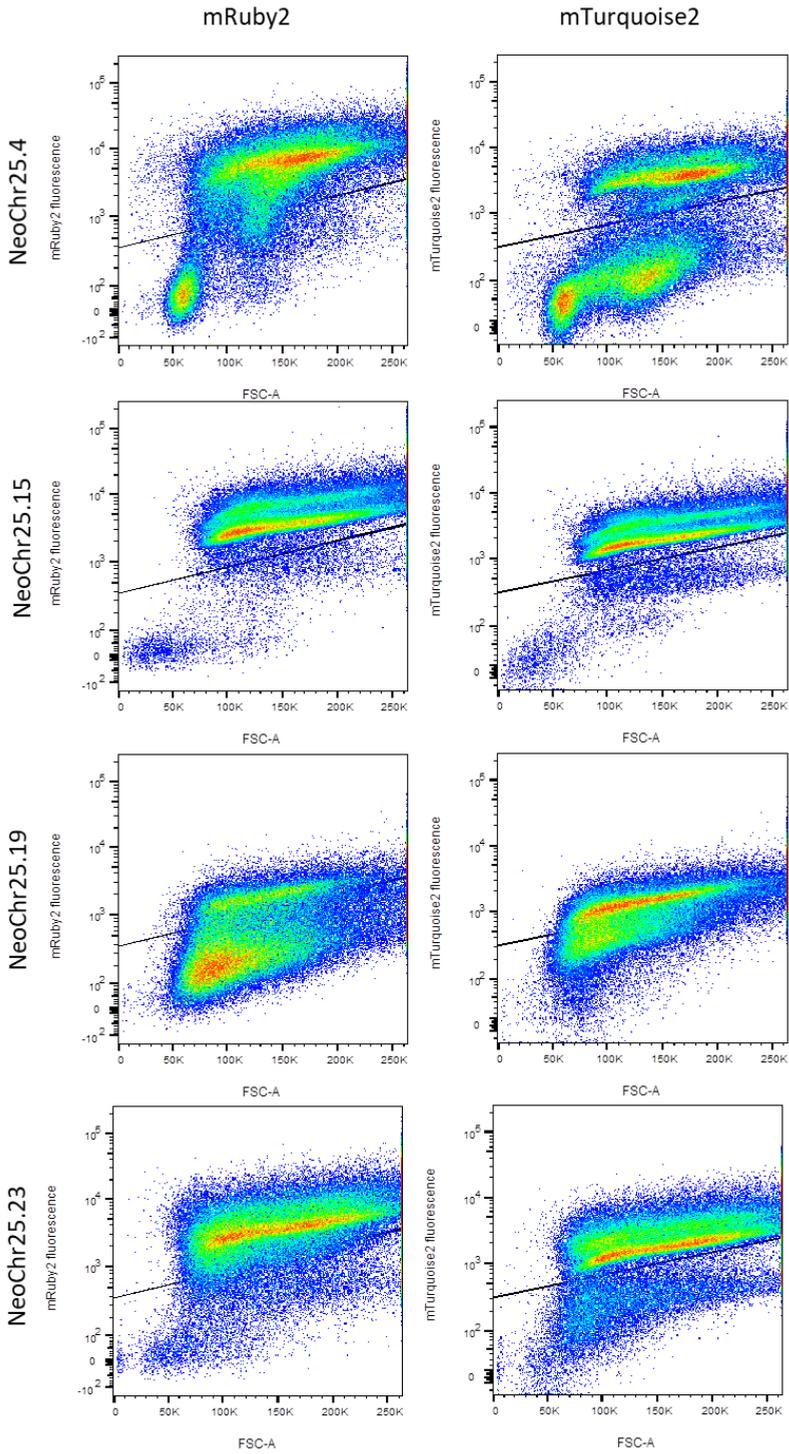


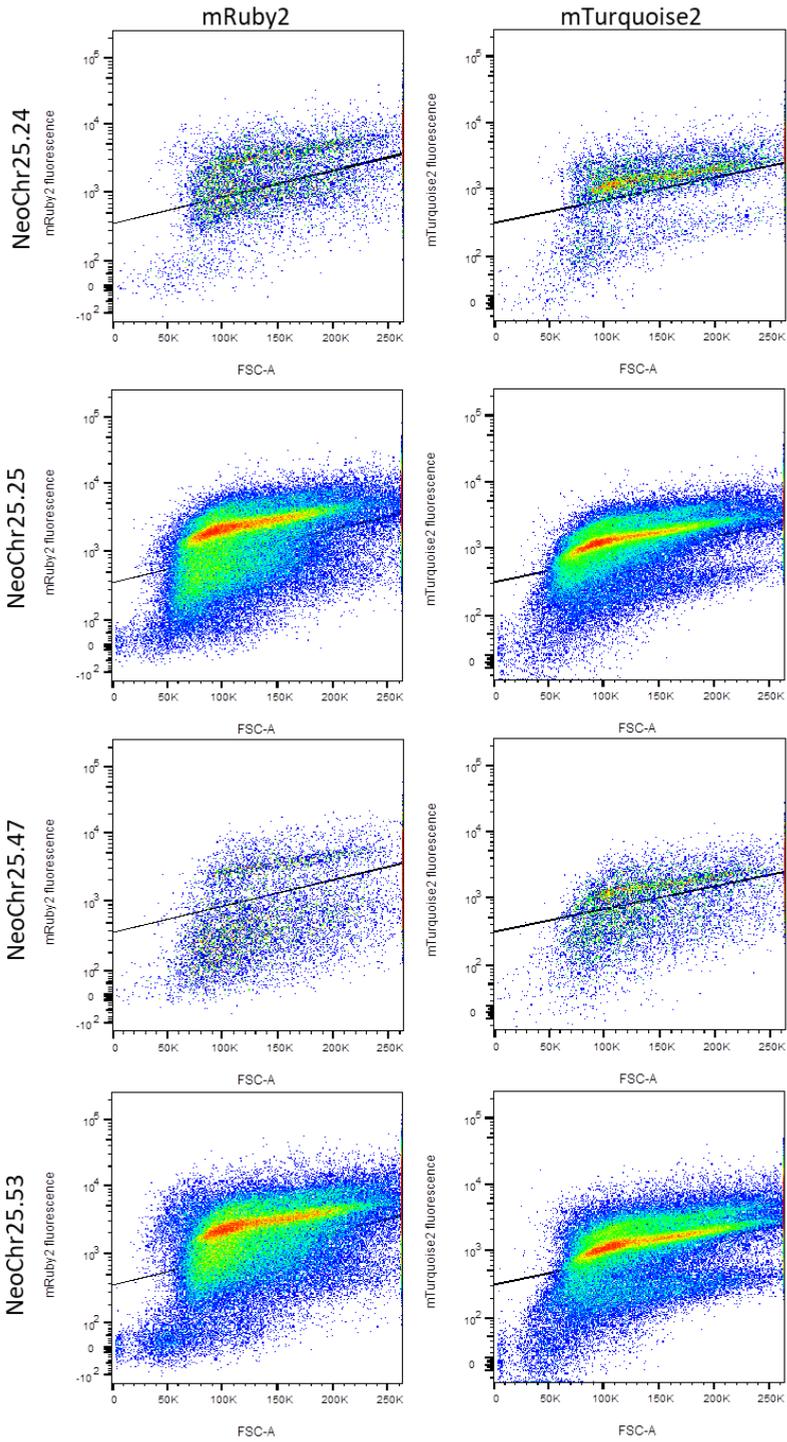


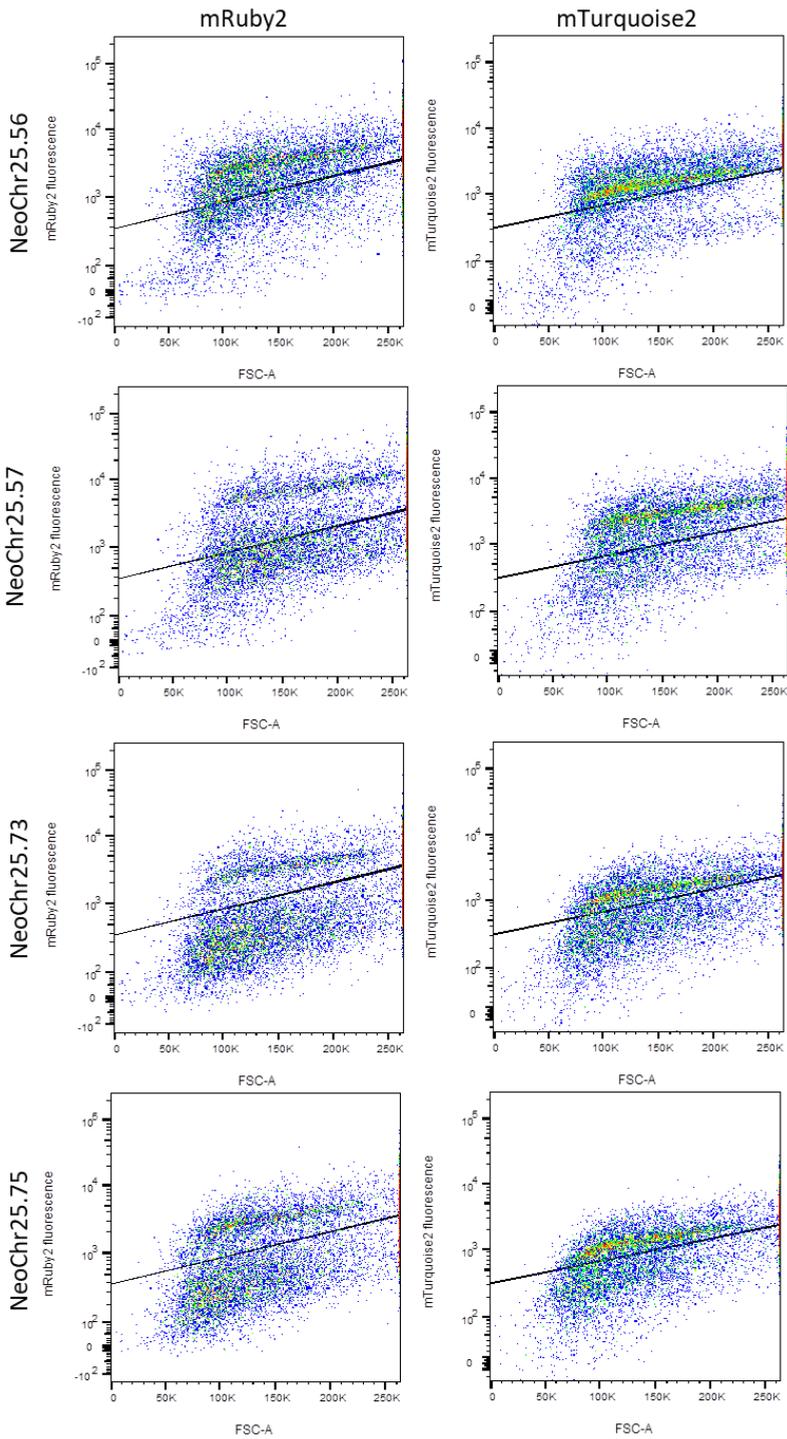
**Figure S7 Flow cytometric analysis of (linear) NeoChr25 and (circular) NeoChr26 designed for anthocyanin production**

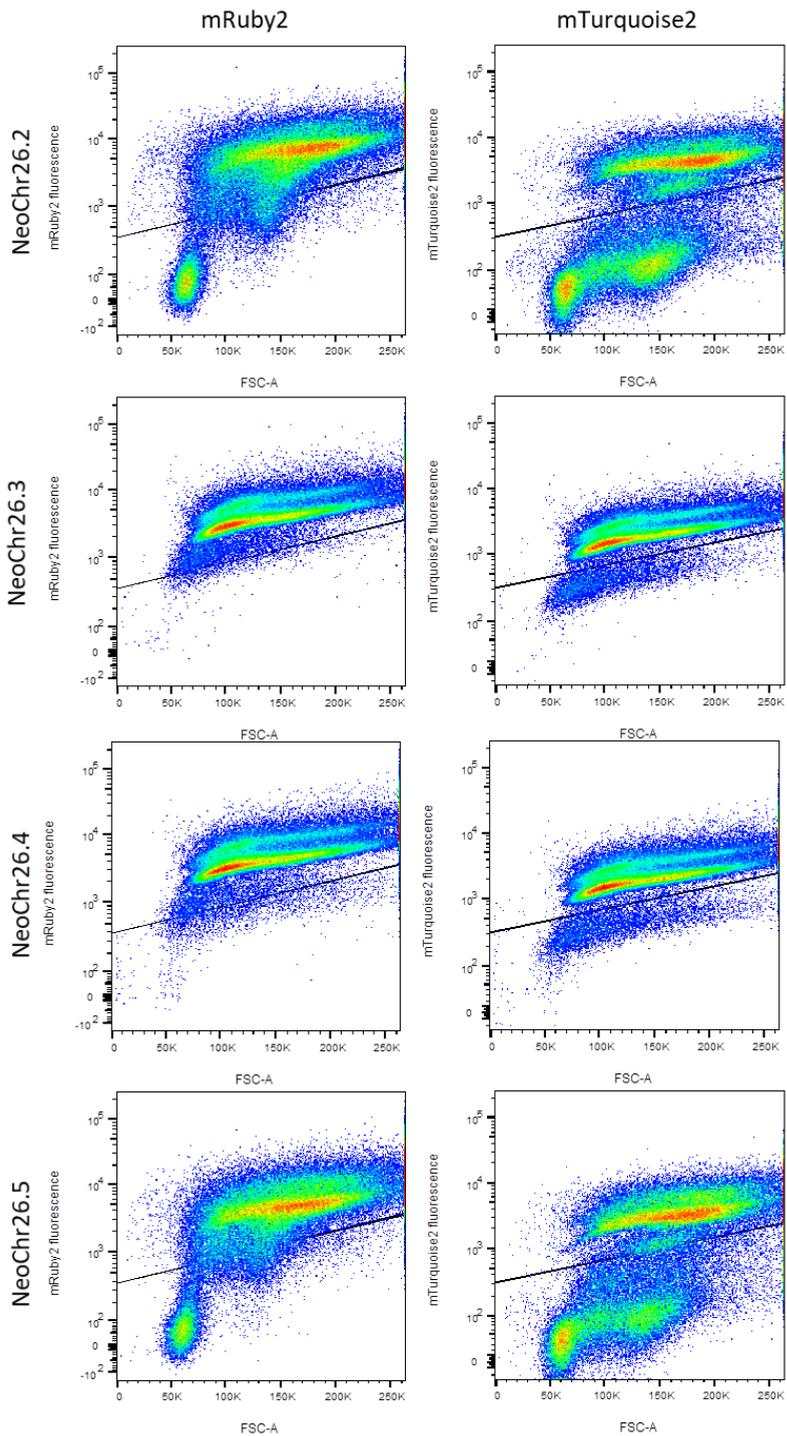
Cells from shake flask cultures were analyzed by FACS. The fluorescence is plotted on the y-axis and the FSC-A on the x-axis. Negative control: CEN.PK113-7D. Positive controls: IMX2224 (mRuby2), IMX2226 (mTurquoise2). Gates for fluorescence of the two different fluorescent proteins were drawn based on the IMX2224 and IMX2226 controls. Approximately 10000 or 100000 events are shown for each plot

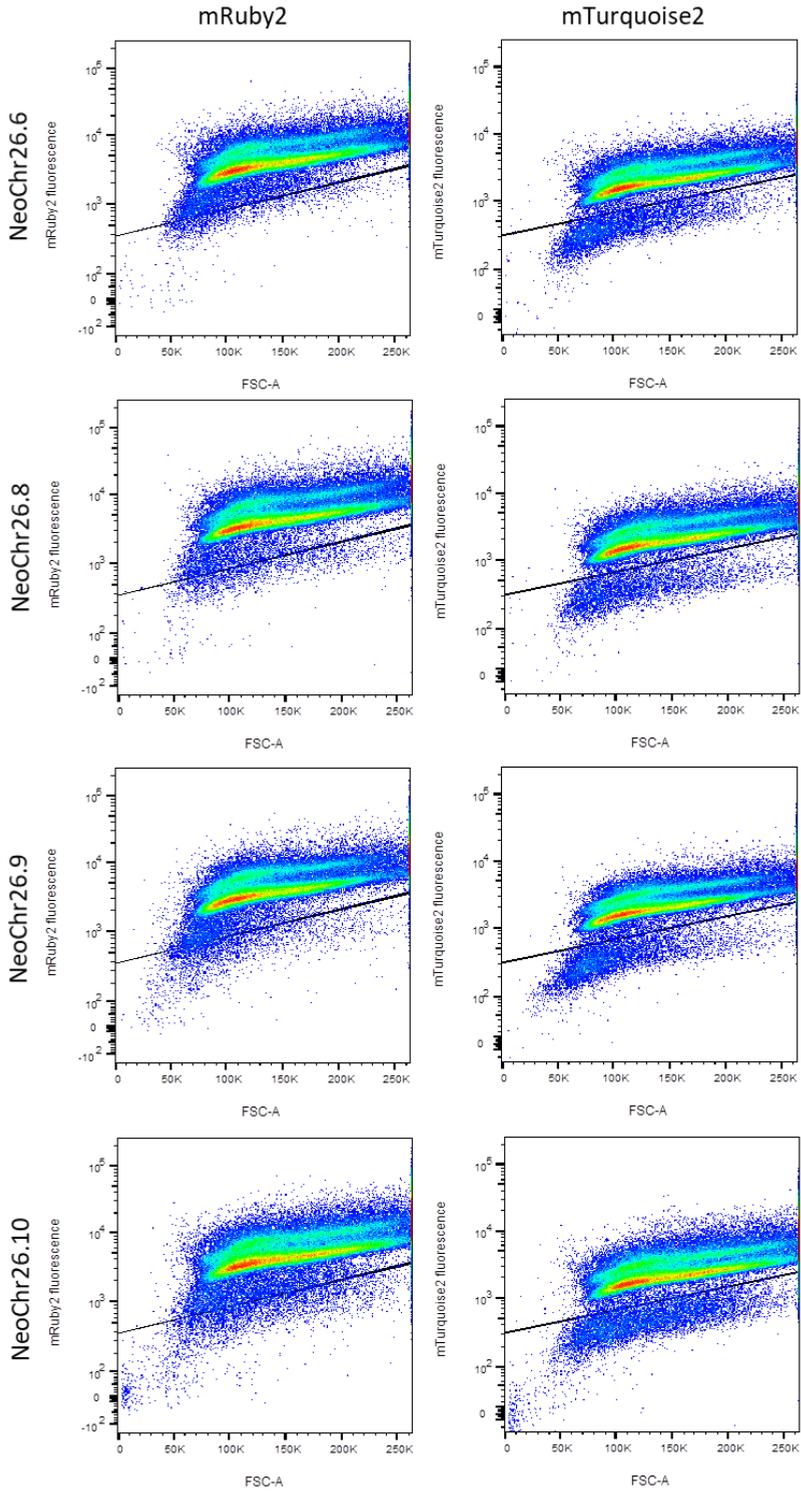


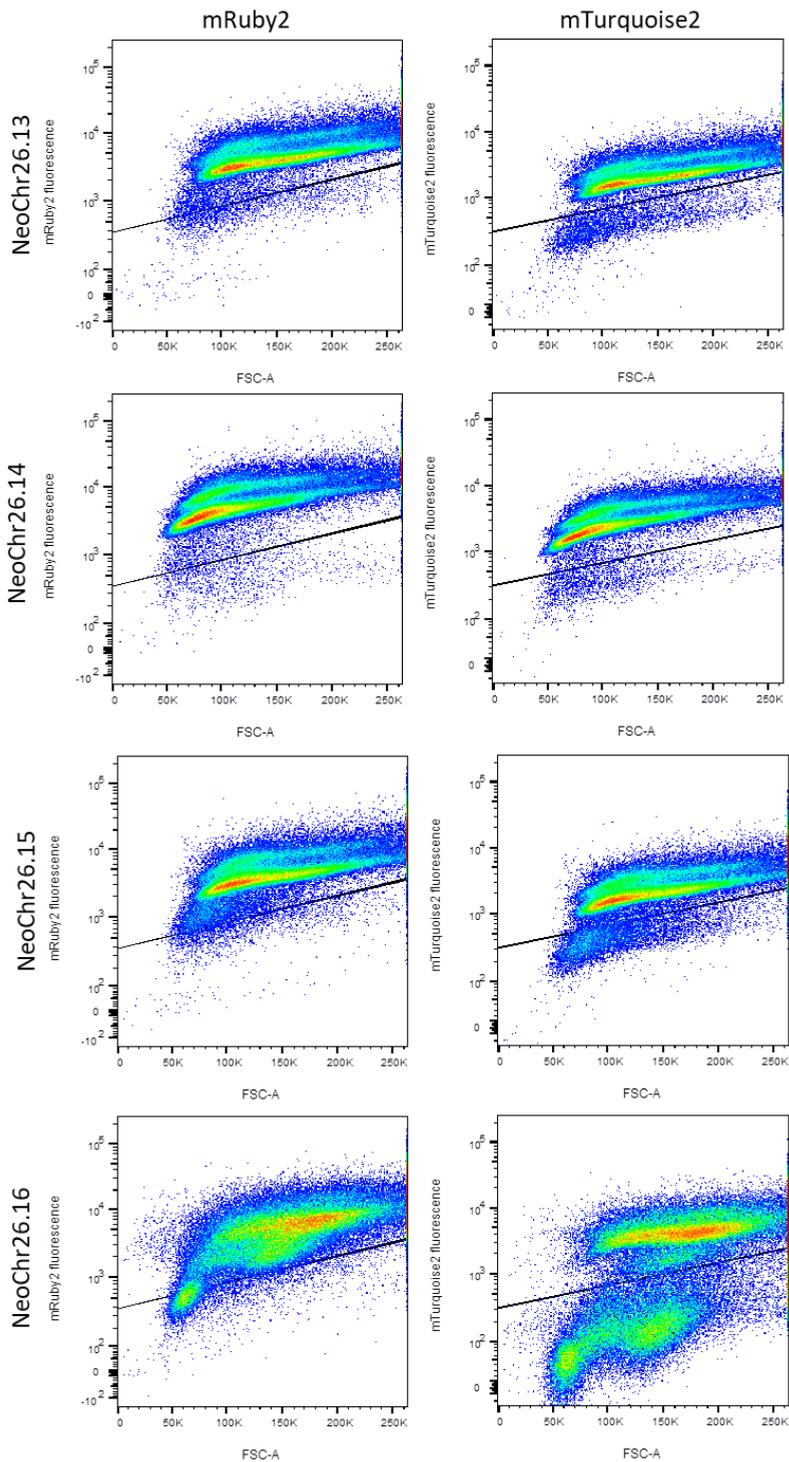






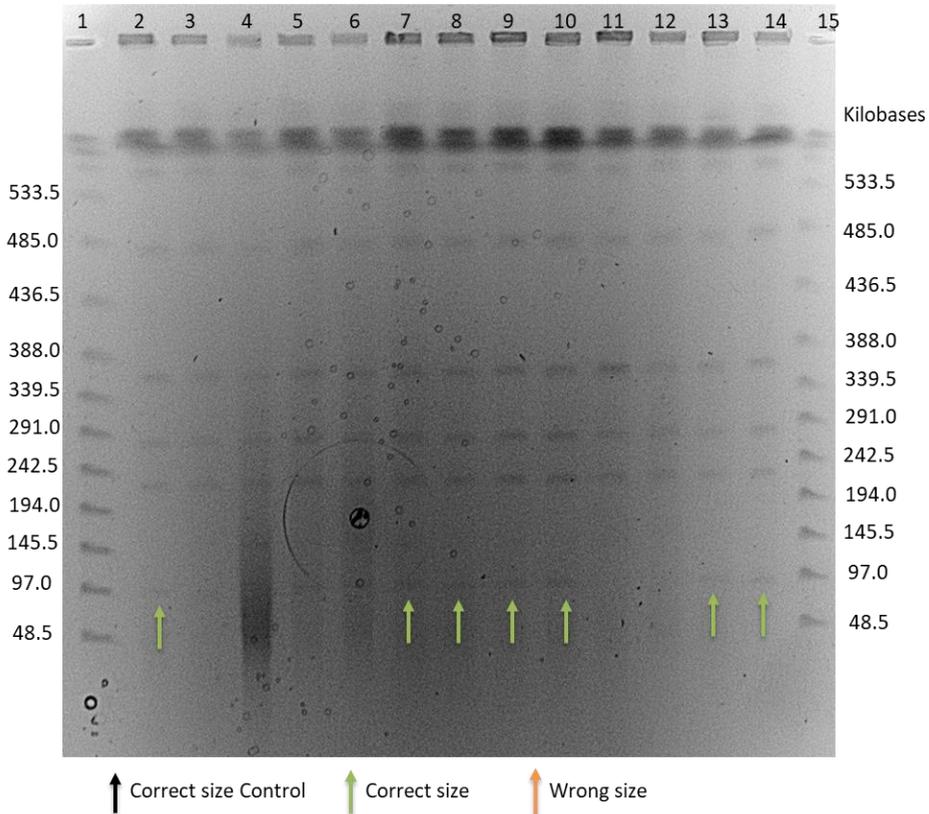






**Figure S8 Separation of (linear) NeoChr25 transformants on pulsed-field electrophoresis.**

Pulsed-field electrophoresis was used to estimate the size of NeoChr25 in several yeast transformants. 1) Size ladder. 2) NeoChr25.4: correct size. 3) NeoChr25.15: no visible neochromosome. 4) NeoChr25.19: no visible neochromosome. 5) NeoChr25.23: no visible neochromosome. 6) NeoChr25.24: no visible neochromosome. 7) NeoChr25.25: correct size. 8) NeoChr25.47: correct size. 9) NeoChr25.53: correct size. 10) NeoChr25.56: correct size. 11) NeoChr25.57: no visible neochromosome. 12) NeoChr25.73: no visible neochromosome. 13) NeoChr25.75: correct size. 14) IMF22: positive control. 15) Size ladder

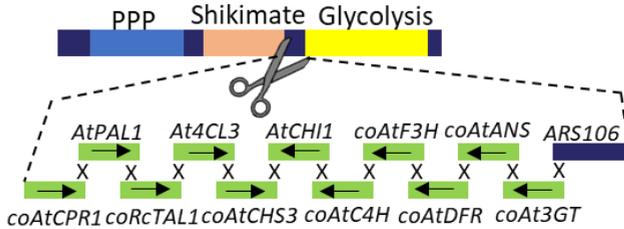


**Figure S9 Duplication and inversion of four plant genes in linear NeoChr25 and circular Neochr26.**

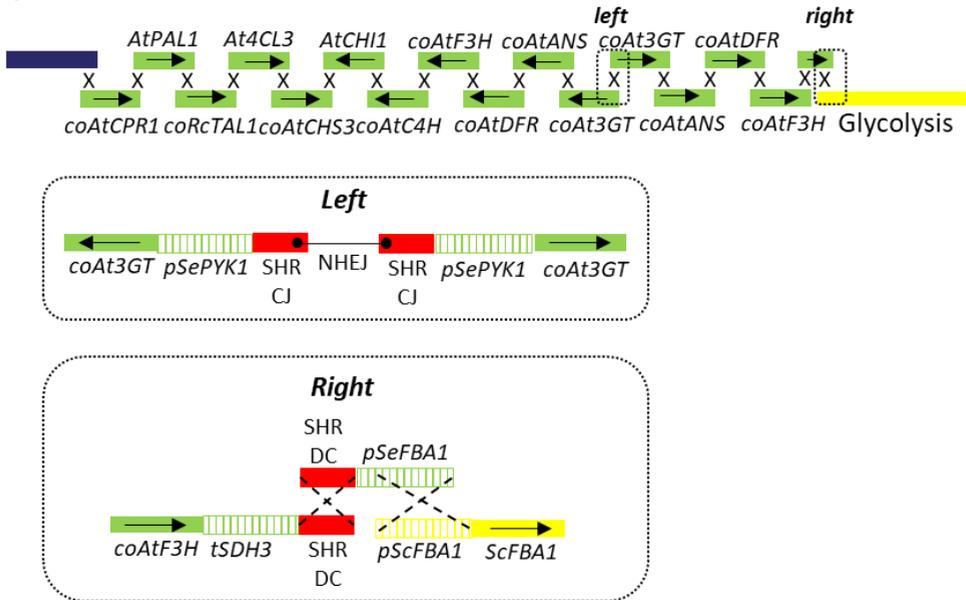
An unexpected recombination was observed upon integration of the genes encoding the anthocyanin production pathway in the linear and circular NeoChrs. **A)** Schematic representation of the *in silico* design for the integration of the anthocyanin pathway in the circular NeoChr26 of IMF40 resulting in IMF41 and the linear NeoChr25 of IMF34 resulting in IMF42. **B)** Schematic representation of the genetic organization observed in IMF41 and IMF42. The last four genes in the anthocyanin pathway (*coAtF3H*, *coGhDFR*, *coAtANS* and *coAt3GT*) were duplicated and inversed, and *ARS106* was absent. The dashed boxes illustrate the recombination events that occurred on the left and right flank of this duplicated region. For the left flank, there was probably an exonuclease and subsequent Non-Homologous End Joining (NHEJ) event between the two SHR CJ, since there was no homology between the inverted and non-inverted

sequences. In the sequenced IMF41 strain (circular) 57 bp of SHR CJ was retained and in the sequenced IMF42 (linear) 51 bp of SHR CJ was retained. For the right flank, in the IMF41 strain (circular) the first 649 bp showed exact homology to *pSeFBA1*, while the last 414 bp showed exact homology to *pScFBA1* (100% homology overlap of 7 bp). In the sequenced IMF42 strain (linear) the first 29 bp showed exact homology to *pSeFBA1* and the last 710 bp showed exact homology to *pScFBA1* (overlap of 100% homology is 24 bp).

**A) *In silico* anthocyanin pathway integration design**



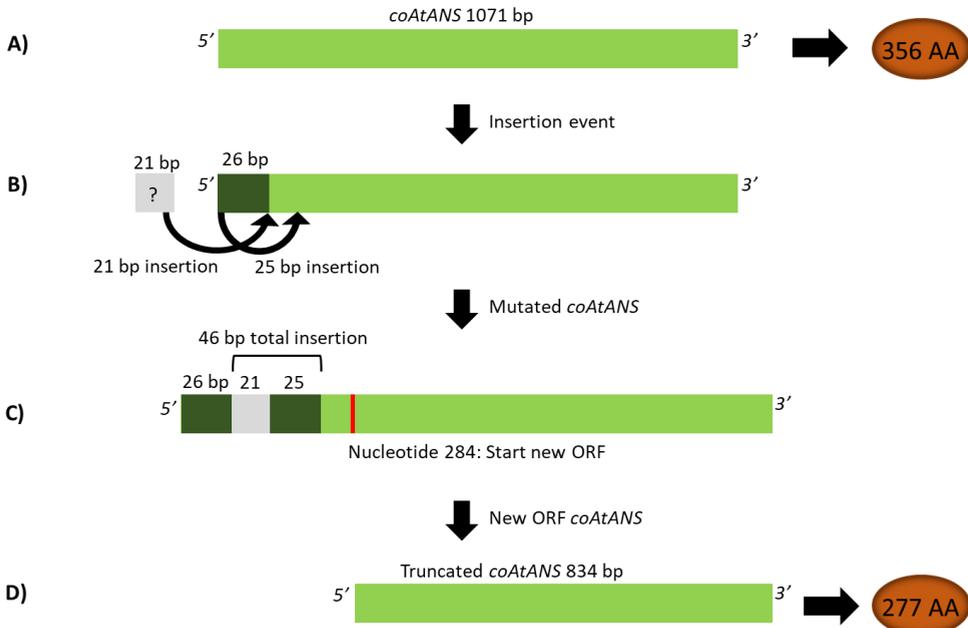
**B) *In vivo* anthocyanin pathway integration**



**Figure S10 schematic representation of *coAtANS* mutation in strains IMF41, IMF42, IMF44 and IMF47**

After sequencing, the *coAtANS* gene in strains IMF41, IMF42, IMF44 and IMF47 turned out to be mutated. **A)** The original *coAtANS* has a length of 1071 bp and encodes for an enzyme consisting of 356 amino acids. **B)** In strains IMF41, IMF42, IMF44 and IMF47, 21 nucleotides of non-homologous DNA (indicated in grey) together with 25 of the first 26 nucleotides of the

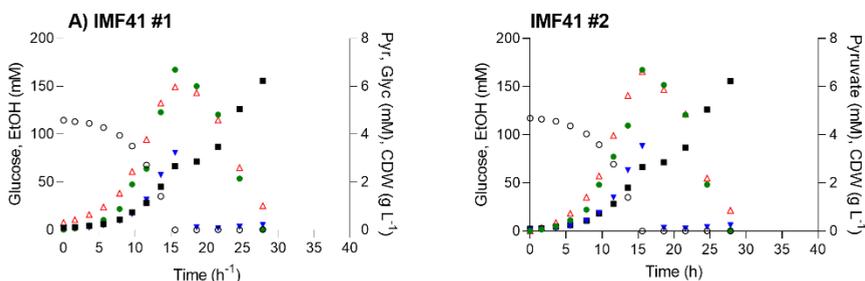
*coAtANS* gene (indicated in dark green) were inserted right after the 26<sup>th</sup> nucleotide. **C)** This insertion resulted in a total insertion of 46 nucleotides disrupting the original ORF. However, this also resulted in a new ORF starting from the 284<sup>th</sup> nucleotide. **D)** The new ORF of the truncated *coAtANS* has a length of 834 bp and encodes for an enzyme consisting of 277 amino acids.

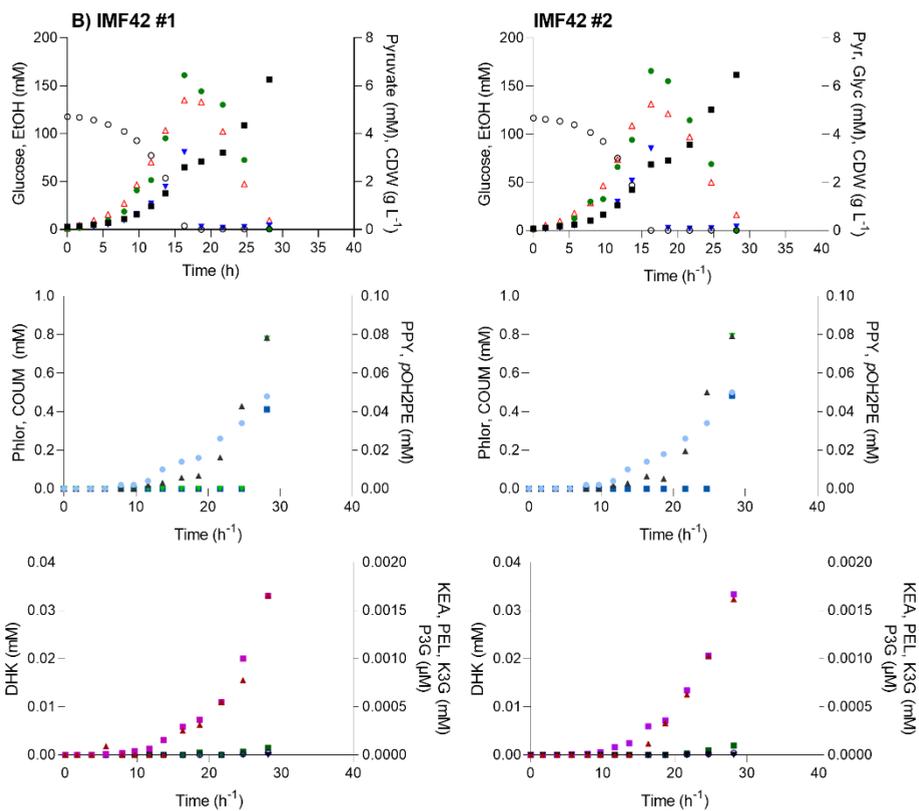
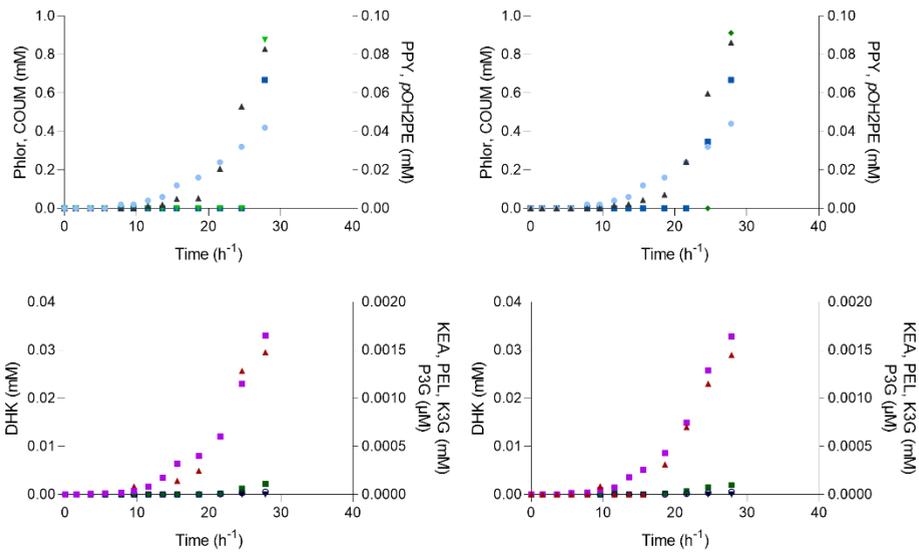


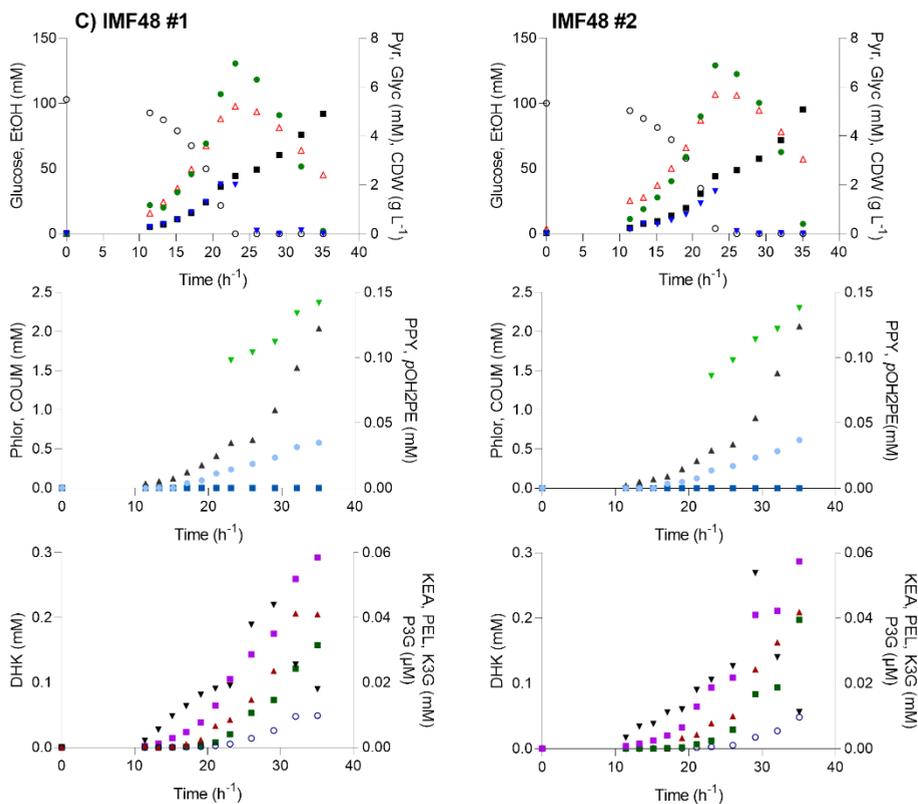
**Figure S11 Substrates and products profiles during aerobic batch cultivation in bioreactors of IMF41, IMF42 and IMF48.**

**A)** IMF41 (Cir, 1x *coAtCHS3*), **B)** IMF42 (Lin, 1x *coAtCHS3*), and **C)** IMF48 (Lin, 9x *coAtCHS3*, *coAtANS*), were grown at 30°C in aerobic batch cultures in bioreactors, in chemically defined medium with 20 g L<sup>-1</sup> glucose as sole carbon source (SMD). Biological duplicates were performed and are shown in two columns as #1 and #2.

Row 1) ■ CDW (g L<sup>-1</sup>), ○ Glucose (mM), ● EtOH (mM), ▼ PYR (mM), ▲ Glyc (mM)  
 Row 2) ■ PPY (mM), ▲ COUM (mM), ● Phlor (mM), ▼ pOH2PE (mM)  
 Row 3) ■ DHK (mM), ▲ PEL (mM), ■ KEA (mM) ○ K3G (mM) ▼ P3G (μM)







**Figure S12 Detection of pelargonidin and pelargonidin 3-O-glucoside by LC-MS**

**A)** Extracted ion chromatogram for the pelargonidin 3-O-glucoside (P3G) mass peak with the composition  $C_{21}H_{21}O_{10}^+$  and the  $m/z$  of 433.1. Data shown for the cell pellet extract of IMF48 duplicate #1 (Table 2), grown in aerobic bioreactor (sample, upper trace), for a blank injection (trace in the middle) analyzed just before the sample and for a synthetic P3G standard shown in the lower trace (Pelargonidin 3-O-glucoside chloride, Sigma Aldrich, Cat No. PHL89753).

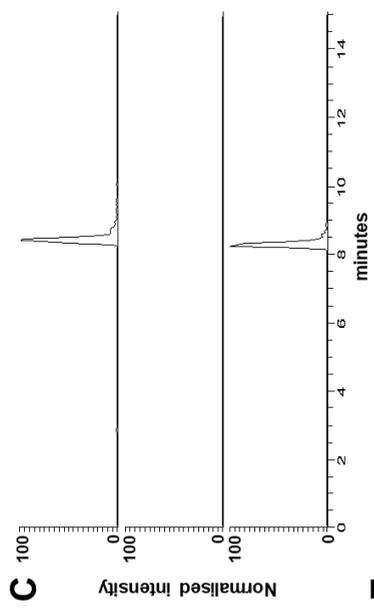
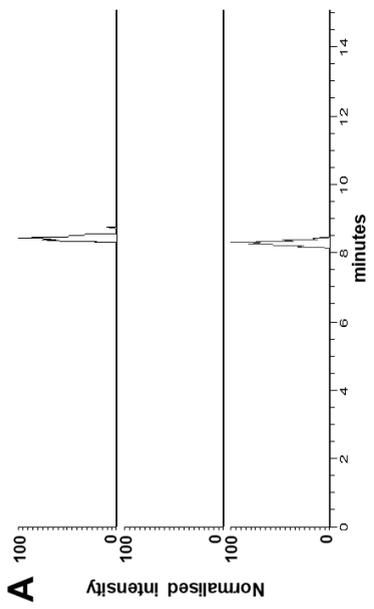
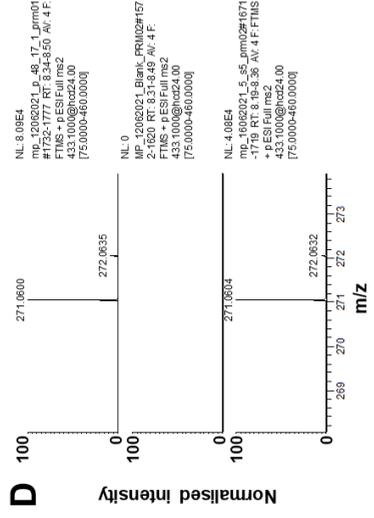
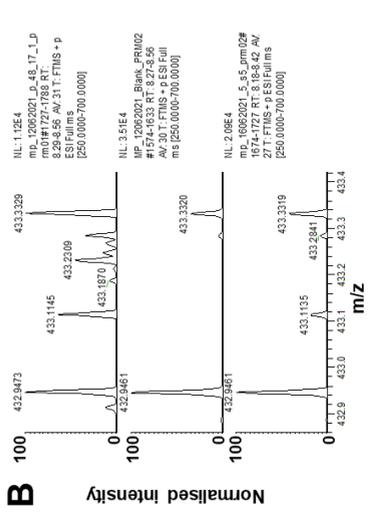
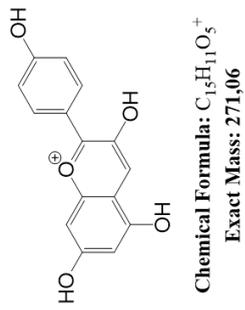
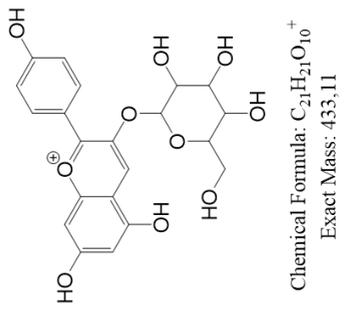
**B)** The mass spectra show the accurate mass of P3G observed in the sample (upper mass spectrum) and the standard (lower mass spectrum). No corresponding P3G peak was observed for the blank injection (spectrum in the middle) analyzed before the sample.

**C)** Extracted ion chromatogram of the pelargonidin (PEL) fragment with the composition  $C_{15}H_{11}O_5^+$ , and a  $m/z$  of 271.06 Da. The corresponding fragment was observed in the sample (upper mass spectrum) and the standard (lower mass spectrum). No corresponding PEL fragment peak was observed for the blank injection analyzed just before the sample. (Pelargonidin chloride, Sigma Aldrich, Cat No PHL80084).

**D)** The spectra show the accurate mass of the PEL major fragment with the composition  $C_{15}H_{11}O_5^+$  and a  $m/z$  of 271.06 Da, as observed for the sample (upper spectrum) and the standard

(lower spectrum). No corresponding fragment mass peak was observed for the blank injection (spectrum in the middle), which was performed just before the sample.

**E)** The table summarized the chemical compositions of P3G and the major fragment of pelargonidin (PEL) (loss of the sugar unit), the resulting theoretical  $m/z$  values, the observed  $m/z$  values and the mass deviations (ppm). The observed mass deviations for standard and sample peaks were  $< 5$  ppm compared to their theoretical  $m/z$  values.



**E**

COMPOSITION	THEORETICAL M/Z	SAMPLE M/Z	Appm	STANDARD M/Z	Δppm
$C_{21}H_{21}O_{10}^+$	433.1129	433.1145	3.69	433.1135	1.39
$C_{15}H_{11}O_5^+$	271.0601	271.0600	-0.37	271.0604	1.11

**Table S1 Promoter-gene-terminator combinations in the NeoChrs**

Promoters, genes or terminators originate from *S. cerevisiae* unless indicated by: *Ec* is *Escherichia coli*, *At* is *Arabidopsis thaliana*, *Rc* is *Rhodobacter capsulatus*, *Gh* is *Gerbera hybrida* *Se* is *Saccharomyces eubayanus*, *Sk* is *Saccharomyces kudriavzevii*, *co* is codon optimized. Watermarked *S. cerevisiae* genes (32) are indicated with an \*.

Promoter	ORF	Terminator
<b>Genes from glycolysis and ethanolic fermentation are expressed from their native promoters and terminators</b>		
<i>pFBA1</i>	<i>FBA1*</i>	<i>tFBA1</i>
<i>pPGM1</i>	<i>PGM1*</i>	<i>tPGM1</i>
<i>pHXK2</i>	<i>HXK2*</i>	<i>tHXK2</i>
<i>pPDC1</i>	<i>PDC1*</i>	<i>tPDC1</i>
<i>pPFK1</i>	<i>PFK1*</i>	<i>tPFK1</i>
<i>pPFK2</i>	<i>PFK2*</i>	<i>tPFK2</i>
<i>pPGK1</i>	<i>PGK1*</i>	<i>tPGK1</i>
<i>pPYK1</i>	<i>PYK1*</i>	<i>tPYK1</i>
<i>pTPI1</i>	<i>TPI1*</i>	<i>tTPI1</i>
<i>pADH1</i>	<i>ADH1*</i>	<i>tADH1</i>
<i>pTDH3</i>	<i>TDH3*</i>	<i>tTDH3</i>
<i>pENO2</i>	<i>ENO2*</i>	<i>tENO2</i>
<i>pPGI1</i>	<i>PGI1*</i>	<i>tPGI1</i>
<b>Genes from the pentose phosphate pathway are expressed from their native promoters and terminators</b>		
<i>pZWF1</i>	<i>ZWF1*</i>	<i>tZWF1</i>
<i>pTKL1</i>	<i>TKL1*</i>	<i>tTKL1</i>
<i>pGND1</i>	<i>GND1*</i>	<i>tGND1</i>
<i>pRK11</i>	<i>RK11*</i>	<i>tRK11</i>
<i>pTAL1</i>	<i>TAL1*</i>	<i>tTAL1</i>
<i>pRPE1</i>	<i>RPE1*</i>	<i>tRPE1</i>
<i>pSOL3</i>	<i>SOL3*</i>	<i>tSOL3</i>
<b>Auxotrophic markers are expressed from their native promoters and terminators</b>		
<i>pHIS3</i>	<i>HIS3</i>	<i>tHIS3</i>
<i>pURA3</i>	<i>URA</i>	<i>tURA3</i>
<b>Fluorescent markers are expressed from <i>S. cerevisiae</i> promoters and terminators. Promoters identified from (24)</b>		
<i>pCCW12</i>	<i>mRuby2</i>	<i>ENO1</i>
<i>pTEF2</i>	<i>mTurquoise2</i>	<i>tSSA1</i>
<b>Genes from the <i>E. coli</i> shikimate pathway are expressed from <i>S. cerevisiae</i> promoters and terminators. Promoters identified from (24-26)</b>		
<i>pRPL3</i>	<i>coEcaroA</i>	<i>tSOL4</i>

<i>pRPL25</i>	<i>coEcaroD</i>	<i>tGPH1</i>
<i>pRPP0</i>	<i>coEcaroE</i>	<i>tCYC1</i>
<i>pHHF1</i>	<i>coEcaroGp<sup>150L</sup></i>	<i>tTEF1</i>
<i>pHTB2</i>	<i>coEcaroL</i>	<i>tPGM2</i>
<i>pRPL10</i>	<i>coEctyrA<sup>M53I A354V</sup></i>	<i>tGDB1</i>
<i>pCWP2</i>	<i>coEctyrB</i>	<i>tGLC3</i>
<i>pHHF2</i>	<i>coEcaroB</i>	<i>tTEF2</i>
<i>pRPL8A</i>	<i>coEcaroC</i>	<i>tGPD2</i>
<i>pRPL18B</i>	<i>coEcpheA<sup>T326P</sup></i>	<i>tGSY2</i>
<b>One gene from the anthocyanin pathway is expressed from a <i>S. cerevisiae</i> promoter and terminator. Promoter identified from (24-26)</b>		
<i>pTEF1</i>	<i>coAtCHS3</i>	<i>tMDH1</i>
<b>Genes from the anthocyanin pathway are expressed from a <i>S. eubayanus</i> and <i>S. kudriavzevii</i> promoters (27) and <i>S. cerevisiae</i> terminators.</b>		
<i>pSePDC1</i>	<i>AtPAL1</i>	<i>tLAT1</i>
<i>pSeGPM1</i>	<i>coRcTAL1</i>	<i>tCIT1</i>
<i>pSkADH1</i>	<i>AtCHI1</i>	<i>tSDH4</i>
<i>pSeFBA1</i>	<i>coAtC4H</i>	<i>tADH3</i>
<i>pSkTDH3</i>	<i>coAtF3H</i>	<i>tSDH3</i>
<i>pSePGK1</i>	<i>coGhDFR</i>	<i>tACO1</i>
<i>pSeENO2</i>	<i>coAtANS</i>	<i>tFUM1</i>
<i>pSePYK1</i>	<i>coAt3GT</i>	<i>tDIC</i>

**Table S2 Sequence fidelity of NeoChrs**

Mutation identified in the neochromosomes as compared to the *in silico* design and with the most relevant parental strain. The \* indicates mutations which are the same in two separate transformations and therefore probably resulting from the template DNA and not during the *in vivo* assembly. Non-synonymous mutations are indicated in bold.

Position	Region	Mutation type
<b>NeoChr25 (IMF27)</b>		
8648	<i>pTKL1</i>	C to CT
14466	SHR BQ	C to CT
20137*	<i>pTEF2 (mTurquoise2)*</i>	CAT to C*
26993	<i>pHHF2 (EcAroB)</i>	AT to A
46676*	<i>pCWP2 (EcTyrB)*</i>	A to G*
52732	SHR AE	G to GT
66608	SHR N	CA to C
73854*	<i>tENO2*</i>	C to A*
86809	<i>pPFK2</i>	GA to G
90753	SHR M	GC to G
90762	SHR M	AT to A

<b>NeoChr26 (IMF29)</b>		
14306	<i>tGND1</i>	CT to C
<b>15181</b>	<b><i>RKI1</i></b>	<b>C to A (Glu-129-Gln)</b>
20137*	<i>pTEF2 (mTurquoise2)*</i>	CAT to C*
22220	SHR DF	TC to T
22223	SHR DF	TC to T
46676*	<i>pCWP2 (EcTyrB)*</i>	A to G*
57795	SHR DL	A to AG
64374	SHR Q	T to TG
66632	<i>pPYK1</i>	CT to C
73854*	<i>tENO2*</i>	C to A*
73864	<i>tENO2</i>	GT to G
73925	SHR B	T to C
73926	SHR B	A to T
73928	SHR B	G to A
78556	<i>pPGI1</i>	C to A
88608	<i>pHIS3</i>	GA to G
90398	<i>pGPM1</i>	C to CTA
<b>NeoChr30 (IMF41) as compared to NeoChr26 (IMF29)</b>		
8648	<i>pTKL1</i>	C to CT
35602	<i>tPGM2 (coEcAroL)</i>	CT to C (In T stretch)
67393	<i>pSeTPI1 (At4CL3)</i>	G to GT (In T stretch)
71689	<i>tMDH1 (coAtCHS3)</i>	GA to G (In A stretch)
73063*	<i>pSkADH1 (AtCHI1)*</i>	AT to A*
73209*	<i>pSkADH1 (AtCHI1)*</i>	CT to C (In T stretch)*
81107	<i>tFUM1 (coAtANS)</i>	CG to C
<b>82459*</b>	<b><i>coAtANS*</i></b>	<b>Insertion of 46 bp*</b>
115179	<i>pPFK2</i>	G to GAA (In A stretch)
<b>NeoChr31 (IMF42) as compared to NeoChr25 (IMF27)</b>		
57432	Chunk 16AB	A to AC
64168	<i>tLAT1 (AtPAL1)</i>	TAA to T (In A stretch)
67393	<i>pSeTPI1 (At4CL3)</i>	G to GT (In T stretch)
<b>67753</b>	<b><i>At4CL3</i></b>	<b>A to G (Thr-15-Ala)</b>
70927	<i>coAtCHS3</i>	G to A (Leu-155-Leu)
70930	<i>coAtCHS3</i>	A to G (Arg-156-Arg)
73063*	<i>pSkADH1 (AtCHI1)*</i>	AT to A*
73209*	<i>pSkADH1 (AtCHI1)*</i>	CT to C (In T stretch)*
<b>82459*</b>	<b><i>coAtANS*</i></b>	<b>Insertion of 46 bp*</b>
<b>NeoChr33 (IMF47) as compared to NeoChr31 (IMF42)</b>		
48856	<i>pCWP2 (coEcTyrB)</i>	A to G
50730	<i>tMDH1 (coAtCHS3)</i>	A to AT (In T stretch)
85431	SHR EB	AT to A

85439	SHR EB	TG to T
85458	SHR EB	GA to G
96995	<i>tPGK1</i>	T to A

**Table S3 Amino acid substitution in native genome of NeoChr strains**

Amino acid substitutions identified in the genome of the constructed strains as compared to most relevant parental strain.

Systematic name	Name	Type	Amino acid change
<b>IMF27 compared to IMX589</b>			
YPL283W-A	-	Intron	-
YNL327W	<i>EGT2</i>	synonymous	Tyr-583-Tyr
YNL327W	<i>EGT2</i>	Non-synonymous	Thr-586-Ser
YNL161W	<i>CBK1</i>	Non-synonymous	Ser-711-Ala
<b>IMF29 compared to IMX589</b>			
YPL283W-A	-	intron	-
YPL283W-A	-	intron	-
YPL283W-A	-	intron	-
YPL283W-A	-	Non-synonymous	Gly-132-Ser
YMR160W	-	Non-synonymous	Gln-11-Arg
YNL327W	<i>EGT2</i>	Synonymous	Tyr-583-Tyr
YNL327W	<i>EGT2</i>	Non-synonymous	Thr-586-Ser
YNL161W	<i>CBK1</i>	Non-synonymous	Ser-711-Ala
<b>IMF41 compared to IMF29</b>			
YCR089W	<i>FIG2</i>	Non-synonymous	Thr-1017-Arg
YCR089W	<i>FIG2</i>	Non-synonymous	Ala-1020-Ser
YDR224C	<i>HTB1</i>	Synonymous	Ala-121-Ala
YIL137C	<i>TMA108</i>	Non-synonymous	Ser-742-Leu
<b>IMF42 compared to IMF27</b>			
YCR089W	<i>FIG2</i>	Non-synonymous	Thr-1017-Arg
YCR089W	<i>FIG2</i>	Non-synonymous	Ala-1020-Ser
YBL113C	-	Non-synonymous	His-252-Asn
<b>IMF47 compared to IMF42</b>			
YEL075W-A	-	intron	-
YHR016C	<i>YSC84</i>	intron	-
YJR143C	<i>PMT4</i>	Non-synonymous	Met-1-Ile

**Table S4 Extracellular concentration of aromatic compounds produced by engineered *S. cerevisiae* strains in shake flask cultures.**

Determination of the intermediates of the anthocyanin pathway in *S. cerevisiae* strains IMF41 (Cir NeoChr, 1X *coAtCHS3*), IMF42 (Lin NeoChr, 1X *coAtCHS3*), IMF47 (Lin NeoChr, 9X *coAtCHS3*) and IMF48 (Lin NeoChr, 9X *coAtCHS3* repaired *coAtANS*), grown in aerobic shake flask batch

cultures on glucose (20 g L<sup>-1</sup>) and urea. The data represents the average ± standard deviation of independent biological triplicates. Intermediates of the anthocyanin pathway coumaroyl-CoA, naringenin-chalcone, and leucopelargonidin were not measured. \* Indicates statistical significance when comparing IMF47 or IMF48 to IMF42, and # when comparing IMF48 to IMF47 (Student *t*-test, two-tailed, homoscedastic, *p*-value threshold 0.05).

(mM)	IMF41	IMF42	IMF47	IMF48
Phenylpyruvate	2.00E-02 ± 0.00E+00	4.33E-02 ± 4.04E-02	BD <sup>a</sup>	BD <sup>a</sup>
2-Phenylethanol	8.67E-02 ± 1.15E-02	8.67E-02 ± 2.52E-02	3.23E-01 ± 3.51E-02*	1.97E-01 ± 2.62E-03**
<i>p</i> - Hydroxyphenylethanol	2.33E-02 ± 2.08E-02	3.33E-02 ± 5.77E-03	1.10E-01 ± 1.73E-02*	BD <sup>a</sup>
Cinnamic acid	3.80E-01 ± 2.00E-02	1.60E-01 ± 2.00E-02	0.00E+00 ± 0.00E+00*	1.46E-01 ± 3.06E-03#
Coumaric acid	4.27E-01 ± 1.15E-02	5.40E-01 ± 2.00E-02	7.13E-01 ± 1.15E-02*	7.64E-01 ± 6.24E-03**
Phloretic acid	7.25E-01 ± 1.21E-02	7.18E-01 ± 8.72E-03	5.09E-01 ± 1.15E-03*	1.07E+00 ± 7.66E-03**
Naringenin	BD <sup>a</sup>	BD <sup>a</sup>	BD <sup>a</sup>	BD <sup>a</sup>
Dihydrokaempferol	2.25E-02 ± 2.57E-03	2.84E-02 ± 1.27E-03	9.49E-02 ± 4.29E-03*	3.83E-01 ± 8.34E-02**
Kaempferol	6.33E-04 ± 3.44E-05	6.96E-04 ± 3.46E-05	2.14E-03 ± 3.06E-04*	1.39E-02 ± 5.72E-03**
Pelargonidin	2.45E-03 ± 1.60E-04	1.81E-03 ± 1.98E-04	5.76E-03 ± 4.11E-04*	3.65E-02 ± 7.64E-03**
Kaempferol 3-O- glucoside	2.24E-04 ± 1.83E-05	2.24E-04 ± 5.98E-06	4.39E-04 ± 2.41E-05*	5.53E-03 ± 1.13E-03**
Pelargonidin 3-O- glucoside	BD <sup>a</sup>	BD <sup>a</sup>	BD <sup>a</sup>	2.14E-05 ± 3.93E-06**
Total aromatics before CHS	1.66 ± 0.06	1.58 ± 0.06	1.66 ± 0.05	1.98 ± 0.01**
Total anthocyanins (after CHS)	0.03 ± 0.00	0.03 ± 0.00	0.1 ± 0.0*	0.45 ± 0.10**
Total aromatics	1.69 ± 0.06	1.61 ± 0.08	1.76 ± 0.05	2.43 ± 0.11**

<sup>a</sup>BD: below detection

### Table S5 Physiological characterization of anthocyanin-producing strains grown in bioreactors

A) The specific growth rate ( $\mu$ ) and the yield (Y) of biomass (X) and ethanol (ETOH) on glucose (S)

**B)** The overall yield (Y) of glycerol (GLYC), pyruvate (PYR), coumaric acid (COUM), phloretic acid (PHLOR) and dihydrokaempferol (DHK) on glucose and ethanol (S) during aerobic bioreactor batch cultivation of IMF41 (Cir NeoChr, 1x *coAtCHS3*), IMF42 (Lin NeoChr, 1x *coAtCHS3*), and IMF48 (Lin NeoChr, 9x *coAtCHS3*, repaired *coAtANS*).

<b>A)</b>	$\mu_{MAX}$ h <sup>-1</sup>	$^aY_{X/S}$ (g g <sup>-1</sup> )	$^aY_{ETOH/S}$ (mol mol <sup>-1</sup> )
<b>IMF41</b> (Cir)	0.23 ± 0.00	0.12 ± 0.00	1.44 ± 0.05
<b>IMF42</b> (Lin)	0.22 ± 0.01	0.12 ± 0.00	1.41 ± 0.02
<b>IMF48</b> (Lin, 9x <i>coAtCHS3</i> , <i>coAtANS</i> )	0.20 ± 0.01	0.13 ± 0.00	1.29 ± 0.05

<b>B)</b>	$Y_{GLYC/S}$ (mol mol <sup>-1</sup> )	$Y_{PYR/S}$ (mol mol <sup>-1</sup> )	$Y_{X/S}$ (mol mol <sup>-1</sup> )	$Y_{COUM/S}$ (μmol mol <sup>-1</sup> )	$Y_{PHLOR/S}$ (μmol mol <sup>-1</sup> )	$Y_{DHK/S}$ (μmol mol <sup>-1</sup> )
<b>IMF41</b> (Cir)	0.056 ± 0.005	0.029 ± 0.001	0.29 ± 0.01	7.30 ± 0.08	4.71 ± 0.20	0.28 ± 0.01
<b>IMF42</b> (Lin)	0.048 ± 0.002	0.029 ± 0.001	0.30 ± 0.01	6.70 ± 0.12	5.02 ± 0.15	0.28 ± 0.00
<b>IMF48</b> (Lin, 9x <i>coAtCHS3</i> , <i>coAtANS</i> )	0.052 ± 0.003	0.019 ± 0.003	0.27 ± 0.01	20.2 ± 0.59	5.89 ± 0.35	2.85 ± 0.02

<sup>a</sup> Determined for the glucose phase only

### Tables S6 until Table S28

Can be found in the Supplementary Materials file at the data.4TU.nl repository via <https://doi.org/10.4121/16539927>

Below the titles of these tables:

**Table S6** *S. cerevisiae* strains used in this study,

**Table S7** Neochromosome configurations,

**Table S8** Plasmids,

**Table S9** pROS/pMEL gRNA primers,

**Table S10** Primers to check correct construction of gRNA plasmids,

**Table S11** Primers to make golden gate part plasmids and expression plasmids with Gibson assembly,

**Table S12** Diagnostic primers to check golden gate part plasmids,

**Table S13** Diagnostic primers to check golden gate and Gibson assembly expression plasmids, for PCR and Sanger sequencing,

**Table S14** List of NeoChr10 and NeoChr11 chromosome parts,

**Table S15** List of primers for amplifying NeoChr10 and NeoChr11 chromosome parts,

**Table S16** Primers for of *amdSYM* deletion,

**Table S17 Primers for deletion of *GND2*, *NQM1*, *SOL4* and *TKL2*,**  
**Table S18 Primers for deletion of *ura3*, *his3* and *SpHIS5*,**  
**Table S19 Primers for deletion of *ARO10*,**  
**Table S20 List of NeoChr25 (linear) and NeoChr26 (circular) chromosome parts,**  
**Table S21 List of primers for amplifying NeoChr25 and NeoChr26 chromosome parts,**  
**Table S22 Primers for glycolysis deletion,**  
**Table S23 Primers to repair *RKI1* mutation in IMF32,**  
**Table S24 Primers for deletion of native *ZWF1*, *GND1*, *SOL3*, *RKI1*, *TAL1*, *TKL1* and *RPE1* ORFs,**  
**Table S25 Parts of the “basic design” of the anthocyanin pathway,**  
**Table S26 List of primers for amplifying the fragments of the “basic design” of the anthocyanin pathway and diagnosing integration,**  
**Table S27 List of primers for amplifying the fragments of the “elaborate design” of the anthocyanin pathway with several copies of the chalcone synthase and diagnostic PCR and**  
**Table S28 List of primers for amplifying the correct *CoAtANS* transcriptional unit and the diagnostic primers used to confirm correct integration**

## References

1. Keasling, J.D. (2010) Manufacturing molecules through metabolic engineering. *Science*, **330**, 1355-1358.
2. Chubukov, V., Mukhopadhyay, A., Petzold, C.J., Keasling, J.D. and Martín, H.G. (2016) Synthetic and systems biology for microbial production of commodity chemicals. *NPJ Syst. Biol. Appl.*, **2**, 16009.
3. Escalera-Fanjul, X., Quezada, H., Riego-Ruiz, L. and Gonzalez, A. (2019) Whole-genome duplication and yeast's fruitful way of life. *Trends Genet.*, **35**, 42-54.
4. Kuijpers, N.G., Solis-Escalante, D., Luttik, M.A., Bisschops, M.M., Boonekamp, F.J., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2016) Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 15060-15065.
5. Postma, E.D., Dashko, S., van Breemen, L., Taylor Parkins, S.K., van den Broek, M., Daran, J.-M. and Daran-Lapujade, P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **49**, 1769-1783.
6. Zha, J., Wu, X. and Koffas, M.A. (2020) Making brilliant colors by microorganisms. *Curr. Opin. Biotechnol.*, **61**, 135-141.
7. Levisson, M., Patinios, C., Hein, S., de Groot, P.A., Daran, J.M., Hall, R.D., Martens, S. and Beekwilder, J. (2018) Engineering *de novo* anthocyanin production in *Saccharomyces cerevisiae*. *Microb. Cell Fact.*, **17**, 103.
8. Eichenberger, M., Hansson, A., Fischer, D., Durr, L. and Naesby, M. (2018) *De novo* biosynthesis of anthocyanins in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **18**, foy046.
9. McClintock, B. (1938) The production of homozygous deficient tissues with mutant characteristics by means of the aberrant mitotic behavior of ring-shaped chromosomes. *Genetics*, **23**, 315-376.
10. Xie, Z.-X., Li, B.-Z., Mitchell, L.A., Wu, Y., Qi, X., Jin, Z., Jia, B., Wang, X., Zeng, B.-X., Liu, H.-M. *et al.* (2017) "Perfect" designer chromosome V and behavior of a ring derivative. *Science*, **355**, eaaf4704.
11. Walker, R.S.K. (2017) *De novo* biological engineering of a tRNA neochromosome in yeast, <https://era.ed.ac.uk/handle/1842/28921>.
12. Mitchell, L.A. and Boeke, J.D. (2014) Circular permutation of a synthetic eukaryotic chromosome with the telomerase. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 17003-17010.
13. Zhao, X., Wei, C., Li, J., Xing, P., Li, J., Zheng, S. and Chen, X. (2017) Cell cycle-dependent control of homologous recombination. *Acta Biochim. Biophys. Sin. (Shanghai)*, **49**, 655-668.
14. Pâques, F. and Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.*, **63**, 349-404.
15. Aylon, Y. and Kupiec, M. (2004) New insights into the mechanism of homologous recombination in yeast. *Mutat. Res.*, **566**, 231-248.
16. Postma, E.D., Couwenberg, L.G.F., van Roosmalen, R.N., Geelhoed, J., de Groot, P.A. and Daran-Lapujade, P. (2021) Top-down, knowledge-based genetic reduction of yeast central carbon metabolism. *bioRxiv*, 2021.2008.2024.457526.
17. Luttik, M.A., Vuralhan, Z., Suir, E., Braus, G.H., Pronk, J.T. and Daran, J.M. (2008) Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: quantification of metabolic impact. *Metab. Eng.*, **10**, 141-153.
18. Hu, C., Jiang, P., Xu, J., Wu, Y. and Huang, W. (2003) Mutation analysis of the feedback inhibition site of phenylalanine-sensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase of *Escherichia coli*. *J. Basic Microbiol.*, **43**, 399-406.

19. Lütke-Eversloh, T. and Stephanopoulos, G. (2005) Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of *Escherichia coli*: generation and characterization of tyrosine-insensitive mutants. *Appl. Environ. Microbiol.*, **71**, 7224-7228.
20. Zhou, H., Liao, X., Wang, T., Du, G. and Chen, J. (2010) Enhanced l-phenylalanine biosynthesis by co-expression of *pheA(fbr)* and *aroF(wt)*. *Bioresour. Technol.*, **101**, 4151-4156.
21. Koopman, F., Beekwilder, J., Crimi, B., van Houwelingen, A., Hall, R.D., Bosch, D., van Maris, A.J., Pronk, J.T. and Daran, J.M. (2012) *De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. *Microb. Cell Fact.*, **11**, 155.
22. Romagnoli, G., Luttik, M.A., Kötter, P., Pronk, J.T. and Daran, J.M. (2012) Substrate specificity of thiamine pyrophosphate-dependent 2-oxo-acid decarboxylases in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **78**, 7538-7548.
23. Vuralhan, Z., Morais, M.A., Tai, S.L., Piper, M.D. and Pronk, J.T. (2003) Identification and characterization of phenylpyruvate decarboxylase genes in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **69**, 4534-4541.
24. Lee, M.E., DeLoache, W.C., Cervantes, B. and Dueber, J.E. (2015) A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.*, **4**, 975-986.
25. Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A. *et al.* (2013) Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.*, **9**, 701.
26. Knijnenburg, T.A., Daran, J.-M.G., van den Broek, M.A., Daran-Lapujade, P.A., de Winde, J.H., Pronk, J.T., Reinders, M.J. and Wessels, L.F. (2009) Combinatorial effects of environmental parameters on transcriptional regulation in *Saccharomyces cerevisiae*: a quantitative analysis of a compendium of chemostat-based transcriptome data. *BMC genomics*, **10**, 53.
27. Boonekamp, F.J., Dashko, S., van den Broek, M., Gehrman, T., Daran, J.-M. and Daran-Lapujade, P. (2018) The genetic makeup and expression of the glycolytic and fermentative pathways are highly conserved within the *Saccharomyces* genus. *Front. Genet.*, **9**, 504.
28. Wei, L., Wang, Z., Zhang, G. and Ye, B. (2017) Characterization of terminators in *Saccharomyces cerevisiae* and an exploration of factors affecting their strength. *Chembiochem*, **18**, 2422-2427.
29. Curran, K.A., Karim, A.S., Gupta, A. and Alper, H.S. (2013) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab. Eng.*, **19**, 88-97.
30. Yamanishi, M., Ito, Y., Kintaka, R., Imamura, C., Katahira, S., Ikeuchi, A., Moriya, H. and Matsuyama, T. (2013) A genome-wide activity assessment of terminator regions in *Saccharomyces cerevisiae* provides a "terminatome" toolbox. *ACS Synth. Biol.*, **2**, 337-347.
31. Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr., Hieter, P., Vogelstein, B. and Kinzler, K.W. (1997) Characterization of the yeast transcriptome. *Cell*, **88**, 243-251.
32. Boonekamp, F.J., Dashko, S., Duiker, D., Gehrman, T., van den Broek, M., den Ridder, M., Pabst, M., Robert, V., Abeel, T., Postma, E.D. *et al.* (2020) Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences. *ACS Synth. Biol.*, **9**, 1361-1375.
33. Mikkelsen, M.D., Buron, L.D., Salomonsen, B., Olsen, C.E., Hansen, B.G., Mortensen, U.H. and Halkier, B.A. (2012) Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.*, **14**, 104-111.

34. Flagfeldt, D.B., Siewers, V., Huang, L. and Nielsen, J. (2009) Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast*, **26**, 545-551.
35. DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.*, **41**, 4336-4343.
36. Kuijpers, N.G., Chroumpi, S., Vos, T., Solis-Escalante, D., Bosman, L., Pronk, J.T., Daran, J.M. and Daran-Lapujade, P. (2013) One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **13**, 769-781.
37. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J. and Daran, J.M. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **15**, 1-15.
38. Zhang, W., Mitchell, L.A., Bader, J.S. and Boeke, J.D. (2020) Synthetic genomes. *Annu. Rev. Biochem.*, **89**, 77-101.
39. Coradini, A.L.V., Hull, C.B. and Ehrenreich, I.M. (2020) Building genomes to understand biology. *Nat. Commun.*, **11**, 6177.
40. Pretorius, I.S. and Boeke, J.D. (2018) Yeast 2.0-connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Res.*, **18**, foy032.
41. Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y. *et al.* (2017) Design of a synthetic yeast genome. *Science*, **355**, 1040-1044.
42. Murray, A.W. and Szostak, J.W. (1983) Construction of artificial chromosomes in yeast. *Nature*, **305**, 189-193.
43. Clarke, L. and Carbon, J. (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, **287**, 504-509.
44. Hieter, P., Mann, C., Snyder, M. and Davis, R.W. (1985) Mitotic stability of yeast chromosomes: a colony color assay that measures nondisjunction and chromosome loss. *Cell*, **40**, 381-392.
45. Sleister, H.M., Mills, K.A., Blackwell, S.E., Killary, A.M., Murray, J.C. and Malone, R.E. (1992) Construction of a human chromosome 4 YAC pool and analysis of artificial chromosome stability. *Nucleic Acids Res.*, **20**, 3419-3425.
46. Wellinger, R.J. and Zakian, V.A. (2012) Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. *Genetics*, **191**, 1073-1105.
47. Ebrahimi, H. and Donaldson, A.D. (2008) Release of yeast telomeres from the nuclear periphery is triggered by replication and maintained by suppression of Ku-mediated anchoring. *Genes Dev.*, **22**, 3363-3374.
48. Straatman, K.R. and Louis, E.J. (2007) Localization of telomeres and telomere-associated proteins in telomerase-negative *Saccharomyces cerevisiae*. *Chromosome Res.*, **15**, 1033-1050.
49. Taddei, A. and Gasser, S.M. (2012) Structure and function in the budding yeast nucleus. *Genetics*, **192**, 107-129.
50. Hamperl, S. and Cimprich, K.A. (2016) Conflict resolution in the genome: how transcription and replication make it work. *Cell*, **167**, 1455-1467.
51. Langenberg, A.K., Bink, F.J., Wolff, L., Walter, S., von Wallbrunn, C., Grossmann, M., Heinisch, J.J. and Schmitz, H.P. (2017) Glycolytic functions are conserved in the genome of the wine yeast *Hanseniaspora uvarum*, and pyruvate kinase limits its capacity for alcoholic fermentation. *Appl. Environ. Microbiol.*, **83**.
52. Machens, F., Balazadeh, S., Mueller-Roeber, B. and Messerschmidt, K. (2017) Synthetic promoters and transcription factors for heterologous protein expression in *Saccharomyces cerevisiae*. *Front. Bioeng. Biotechnol.*, **5**, 63.

53. Chen, B., Lee, H.L., Heng, Y.C., Chua, N., Teo, W.S., Choi, W.J., Leong, S.S.J., Foo, J.L. and Chang, M.W. (2018) Synthetic biology toolkits and applications in *Saccharomyces cerevisiae*. *Biotechnol. Adv.*, **36**, 1870-1881.
54. Solis-Escalante, D., Kuijpers, N.G., Barrajon-Simancas, N., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2015) A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell*, **14**, 804-816.
55. Gorter de Vries, A.R., Knibbe, E., van Roosmalen, R., van den Broek, M., de la Torre Cortés, P., O'Herne, S.F., Vijverberg, P.A., El Masoudi, A., Brouwers, N., Pronk, J.T. *et al.* (2020) Improving industrially relevant phenotypic traits by engineering chromosome copy number in *Saccharomyces pastorianus*. *Front. Genet.*, **11**, 518.
56. Entian, K.-D. and Kötter, P. (2007) 25 Yeast genetic strain and plasmid collections. In: Stansfield, I. and Stark, M. J. R. (eds.), *Methods in Microbiology*. Academic Press, Vol. **36**, pp. 629-666.
57. Verduyn, C., Postma, E., Scheffers, W.A. and Van Dijken, J.P. (1992) Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, **8**, 501-517.
58. Luttik, M.A., Kötter, P., Salomons, F.A., van der Klei, I.J., van Dijken, J.P. and Pronk, J.T. (2000) The *Saccharomyces cerevisiae* *ICL2* gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme A metabolism. *J. Bacteriol.*, **182**, 7007-7013.
59. Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23-28.
60. Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.*, **350**, 87-96.
61. Looke, M., Kristjuhan, K. and Kristjuhan, A. (2011) Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques*, **50**, 325-328.
62. Hassing, E.J., de Groot, P.A., Marquenie, V.R., Pronk, J.T. and Daran, J.G. (2019) Connecting central carbon and aromatic amino acid metabolisms to improve *de novo* 2-phenylethanol production in *Saccharomyces cerevisiae*. *Metab. Eng.*, **56**, 165-180.
63. Kuijpers, N.G., Solis-Escalante, D., Bosman, L., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb. Cell Fact.*, **12**, 47.
64. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, **27**, 722-736.
65. Postma, E., Scheffers, W.A. and van Dijken, J.P. (1989) Kinetics of growth and glucose transport in glucose-limited chemostat cultures of *Saccharomyces cerevisiae* CBS 8066. *Yeast*, **5**, 159-165.

## Outlook

### **Implementation of specialized neochromosomes as novel, efficient way for the construction of yeast cell factories**

Microbial cell factories have become increasingly important for the sustainable production of a range of bulk and fine chemicals from renewable feedstocks (1-3). *Saccharomyces cerevisiae*, also known as baker's yeast, is a popular host for these applications due to its long industrial track record and excellent accessibility to modern genome-editing techniques (4,5). The high efficiency of homologous recombination (HR) of DNA strands in *S. cerevisiae* has enabled a tremendous acceleration of genetic modifications. In combination with CRISPR/Cas-based genome-editing technologies, the native HR machinery of *S. cerevisiae* allows for simultaneous introduction of multiple 'seamless' gene deletions, insertions and modifications into the yeast genome (6-8). However, even in *S. cerevisiae*, the development of new cell factories is still hampered by the number of alterations that can be simultaneously achieved in a single transformation step, which is currently at nine (8,9). This limitation stems from insufficient repair efficiency of the double strand breaks introduced via the CRISPR system, which result in cell death. Study and modifications in this repair reaction (donor DNA design and delivery, cell synchronization, increasing HR with respect to alternative repair pathways) could possibly increase the multiplex genome editing efficiency (8). In addition, the lack of predictability of genome engineering on product titer, production rate, and product yield (TRY) typically necessitates multiple, iterative rounds of engineering (10). The research described in this thesis therefore explored a new approach for fast and efficient construction of yeast cell factories.

In the native *S. cerevisiae* genome, genes encoding the enzymes of any given metabolic pathway are scattered over multiple chromosomes, which complicates fast and combinatorial approaches for pathway optimization. Relocation of all genes encoding enzymes of a metabolic pathway (or a set of proteins involved in another cellular process) to a single chromosomal locus offers the opportunity to rapidly and easily exchange entire pathways for alternative designs. Based on relocation of pathway genes to a native yeast chromosome, this approach was successfully explored and tested in studies on the 'pathway swapping concept' of Kuijpers *et al.* (11). In this thesis a new, more radical approach was explored, namely the use a synthetic supernumerary neochromosome as orthogonal platform for expression and remodeling of multiple (metabolic) pathways.

The concept of 'synthetic' chromosomes in yeast is not new and the first functional small circular chromosome was already made in 1980 (12). Since then, Yeast Artificial

Chromosomes (YACs) have been used for the study of complex eukaryotic genomes in terms of architecture and gene functions (13,14). Lately, a range of synthetic (minimal) genomes were assembled in *S. cerevisiae*, before potential transplantation in a recipient cell (15-22). Finally, in the Sc2.0 project, all native yeast chromosomes have been replaced by synthetic versions that carry some deliberate alterations (23). The approach described in this thesis is profoundly different, both in terms of the experimental approach and in the design of the synthetic chromosomes. While previous studies relied on existing *S. cerevisiae* chromosomal sequences as backbones or as scaffolds for sequence replacement, the present study demonstrates single-step assembly of entirely synthetic chromosomes with a modular design that is specifically geared towards rational engineering of metabolic pathways and cellular processes.

By harnessing the HR capacity of yeast, we were able to synthesize complete, supernumerary synthetic neochromosomes in a single transformation step. Circular and linear neochromosomes with a size of 100 kb were efficiently assembled from up to 44 transcriptional-sized fragments. While this number of fragments is already impressive, there is no reason to assume that the limit of HR capacity, in terms of number of fragments that can be simultaneously assembled or neochromosome size, has been reached. It is therefore highly relevant to further explore the limit of the HR capacity of native *S. cerevisiae* strains and, subsequently, to investigate how this capacity can be further extended by genetic engineering (24,25). Such research may also help to provide insight in the molecular requirements for efficient HR and, thereby, contribute to its transfer to other eukaryotic cell factories and laboratory model organisms (26-28). Furthermore, this thesis demonstrates that newly assembled neochromosomes can be used as vectors for expression of multiple clustered metabolic pathways. Using a plant-derived pathway for anthocyanin production as proof of principle, it was shown that these pathways can either be derived from yeast or from other organisms. The implied vision for future research is that single-step assembly of modularly organized neochromosomes can be used to generate large numbers of different configurations of precursor-supply and product-synthesis pathways. In application-oriented research, these configurations can be used for screening of optimal TRY performance indicators. In fundamental research, our approach can be used to investigate other cellular processes (*e.g.* related to cell division, cell wall synthesis or stress tolerance). These and other potential applications of neochromosomes as well as their relation to neochromosome design will be further elaborated on below.

### **Genetic reduction can increase genome engineering efficiency**

The central carbon metabolism (CCM) of yeast plays an essential role in industrial product formation by providing metabolic precursors and energy-rich molecules, as well as for cofactor balancing. Product-specific rewiring of the CCM is therefore highly

important for optimizing TRYs (29-35). Ultimately, it would be relevant to relocate all genes encoding CCM enzymes to a synthetic neochromosome and, thereby, open up CCM for combinatorial optimization for specific industrial applications. Achieving this goal would benefit from eliminating any genetic redundancy in the native *S. cerevisiae* CCM. Results presented in **Chapter 2** demonstrated that 35 deletions in the CCM of *S. cerevisiae*, corresponding to an approximate 32% reduction of gene number, led to only a slight decrease of the specific growth rate on standard glucose chemically-defined medium. Also, in growth experiments on a wide range of growth conditions including other carbon sources and stressful environments, only minimal impacts were observed. These results show that many of the iso-enzymes of yeast are dispensable under standard laboratory conditions as well as for a wider range of environments. For many of these proteins, their fixation in the yeast genome during evolution in natural environments remains an intriguing scientific question. The 'minimal CCM' strain described in this thesis presents a perfect starting point towards a neochromosome-based reconfiguration of yeast CCM. In follow up studies, relocalization of the remaining 76 major paralogs of the minimal CCM strain to a 'CCM neochromosome' would enable large scale CCM optimization for yeast-based production of industrially relevant compounds.

The CCM minimalization effort presented in **Chapter 2** focused on a subset of pathways (glycolysis, gluconeogenesis, ethanolic fermentation, pentose phosphate pathway, TCA cycle, mitochondrial transporters, fumarate reductases, anaplerotic reactions, glyoxylate cycle, acetyl-CoA synthesis and glycerol synthesis) strategically chosen for their relevance for most industrial products. Beyond CCM, some other pathways are relevant for specific product classes, such as amino acid biosynthesis and shikimate pathway for anthocyanins, as illustrated in **Chapter 5**. Depending on the product or research question, it would therefore be useful to (re)locate other pathways (*e.g.* pathways for assimilation of alternative carbon sources, plasma membrane transporters, reserve carbohydrates *etc.*) and their regulatory genes on one or more specialized neochromosomes. On the short term, we can envision neochromosome platform strains with specific sets of clustered pathways dedicated to TRYs optimization. Looking further ahead, the yeast genome could be redesigned towards a modular genome, with sets of specialized neochromosomes tailored for industrial strains optimization.

### **Design principles and applications of specialized neochromosomes**

Results in this thesis show that assembled non-coding neochromosomes were stable in terms of mitotic segregation and genetic composition and had no impact on the physiology of the host. Also, fully coding neochromosomes exclusively containing the glycolysis, ethanolic fermentation and pentose phosphate pathway as well as the *Escherichia coli* amino acid biosynthesis and shikimate pathway were mitotically stable

and only resulted in a minor growth rate decrease of 10-16%. While these results are very promising for the use of neochromosomes as expression platforms, for initial strain construction homologous sequences within the neochromosome could lead to unwanted recombination events during *in vivo* neochromosome assembly. Recombination is less likely to occur during strain propagation as yeast's native chromosomes are very stable despite the occurrence of many homologous sequences, and subculturing of neochromosome strains with significant homology did not result in recombination (36,37). Avoiding unintended recombination events during neochromosome construction, implies that the same genetic sequence (promoters, genes, terminators, auxiliary parts etc.) cannot be used multiple times in the neochromosome design. This can become limiting considering that *S. cerevisiae* toolbox does not contain enough well-characterized strong constitutive promoter sequences, important for the design of multi-gene heterologous and/or synthetic product-pathway gene-clusters. Fortunately, research in this field is expanding, highlighted by the recent addition of strong heterologous promoters to the molecular toolbox of *S. cerevisiae* and the vast amount of research that is currently dedicated to the design of (fully) synthetic promoters (38-42). Still, there is insufficient quantitative data on the nature of homologous sequences (length of the homology sequence in combination with tolerated sequence divergence and spatial proximity of homologous sequences) and the occurrence of homologous recombination. This is exemplified by the fact that a recombination occurred between the *FBA1* promoter of *S. cerevisiae* and the *FBA1* promoter of *S. eubayanus* during anthocyanin pathway integration in the neochromosome, while these sequences did not have the highest global or local homology as compared to other promoter sequences used (38). Further systematic research on homology threshold for neochromosome *in vivo* assembly would bring valuable insights for the design of future specialized, synthetic chromosomes.

From **chapter 3**, in which the reduced growth rate of 10-30% of the single locus glycolysis strain was investigated, several important design principles could be learned. First, while natively ARS sequences are spaced every 30-40 kb, doubling this distance by integration of the 35 kb glycolytic cassette did not influence growth rate, which confirms observations by Karas *et al.*(43) that ARS sequences or only required every 100 kb or not at all if AT-rich sequences are used (44). Furthermore, the orientation of transcription with respect to replication is not important, at least in the 13 gene containing highly transcribed glycolytic cassette. Finally, clustering of genes at a single locus only marginally effects expression (and activity) of the genes in the cassette and the neighboring genes at the site of integration. It would be interesting to evaluate if the same conclusions hold true for synthetic chromosome design. In future research, the single-step chromosome-assembly strategy described in this thesis can be used to accelerate studies on the impact of chromosome architecture (*e.g.* size,

spacing of ARS sequences and clustering of highly expressed genes) on chromosome segregation and gene expression. Moreover, the current modular neochromosome design, clustering pathways to enable individual swapping, results in transcriptional hotspots. While this modular design did not result in visible negative impacts on yeast physiology in our work, further studies are required before it can be established as general principle for neochromosome design.

To efficiently evaluate design principles like sequence homology and gene placement, an experimental infrastructure is required that not only enables generation of many large neochromosomes with different (combinatorial) designs, but also the high-throughput sequence verification and quantitative analysis of the performance of the resulting yeast strains. Liquid-handling robots and automated protocols for chromosome design, assembly and strain analysis are indispensable requirements for such research. In addition, quality control of the constructed neochromosomes and strains will require access to affordable, high-throughput DNA-sequencing platforms.

While the research on these fundamental aspects, such as gene and ARS placement, requires the generation of neochromosomes true to a carefully planned design, for applied research this does not necessarily have to be the case. Initial screening of strains with different neochromosome configurations could be based only on TRY performance indicators, possibly made more efficient by use of appropriate biosensors. Sequencing of the best strains, even if they are not true to the initial design could then reveal beneficial, potentially unexpected genetic configurations. Rational chromosome design based on experimental data and modelling could even be replaced by chromosome-level combinatorial approaches. Cells, provided with multiple gene variants for each reaction (all containing the same flanks), would then be selected for the most favorable chromosomal configuration based on their performance in high-throughput screening (45). Classical genetics could also provide an interesting way of increasing genetic diversity and thus allow for higher TRY performances. This would be accomplished by changing the mating type of neochromosome harboring strains and subsequently mating cells with different neochromosome configurations, thus generating a range of different diploid strains.

### **Towards fully synthetic cells**

The HR machinery of *S. cerevisiae* has already played a key role in the assembly of the first (fully synthetic) bacterial genomes (15-22). The research described in this thesis can contribute to an acceleration of the synthesis of novel synthetic chromosomes, which will not be constrained to yeast and might also be applied in the manufacture of Human Artificial Chromosomes (HACs). HACs can be used in fundamental research on chromosome and gene function as well as possibly applied for gene- and stem cell- therapy, although these disciplines might be considered as controversial (46,47).

Finally, this research might bring scientists closer to achieving the goal of synthetic biologists to *de novo* design and construct completely synthetic cell factories dedicated to specific functions or products (48). Achievements have been made in the three main aspects important for bottom-up synthetic cell construction: compartmentalization, metabolism and informational control. Mostly DNA synthetic genomes are considered for storage of information. Whether *S. cerevisiae* will have a role to play in this synthetic genome assembly remains to be seen. An ever-increasing amount of research is dedicated to writing and assembling error-free DNA as reviewed by Hughes *et al.* (49). Joining of DNA bases into oligonucleotides by phosphoramidite synthesis chemistry has expanded from column-based approaches to microarrays reducing cost and allowing multiplex synthesis. Synthons assembly from oligonucleotides is mainly based on polymerase cycle assembly (PCA) but can also be efficiently achieved by Gibson assembly. Moreover, error correction methods using MutS techniques or NextGen sequencing can efficiently sieve error-containing synthons from correct ones. Finally, synthons can be assembled into larger constructs by a variety of *in vitro* techniques which are becoming ever more efficient. What is now feasible can be seen from genes synthesis companies such as: CodexDNA which sells a cell-free DNA writing device which can from 60-mers make up to 7 kb long DNA fragments based on automatized Gibson assembly; and companies like GeneArt from which you can order up to 100 kb long DNA sequences. Despite these achievements in *in vitro* methods, ordering of genome scale DNA construct remains expensive. Moreover, *E. coli*, a preferred host for propagation and amplification of *in vitro* assembled constructs, often experiences toxicity from introduced bacterial DNA and has difficulty with replicating very large genome-sized DNA constructs (43). Another option is amplification by *in vitro* methods such as rolling circle amplification and *E. coli* chromosome-replication cycle, however these methods still have to improve themselves on delivering sufficient DNA yield and on routine and faithful amplification of genome-sized DNA (due to shearing) (50,51). Until these shortcomings of *in vitro* assembly and amplification can be improved, *S. cerevisiae* remains an attractive host for synthetic genome assembly.

An important research topic in bottom-up synthetic cell biology is how to kick-start a living cell from a life-less genome. However, the biggest challenge in the construction of a cell-like entity which can grow and divide, is merging the knowledge and accomplishment from the different disciplines into one working system (52). Several research consortia (*e.g.* BaSyC, Build-a-cell, MaxSynBio and the Synthetic cell initiative) aim to do just that (53). The managers of these research consortia predict that they will be able to construct living artificial cells in just over a decade (54).

## References

1. Keasling, J.D. (2010) Manufacturing molecules through metabolic engineering. *Science*, **330**, 1355-1358.
2. Cravens, A., Payne, J. and Smolke, C.D. (2019) Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat. Commun.*, **10**, 2142.
3. Ko, Y.-S., Kim, J.W., Lee, J.A., Han, T., Kim, G.B., Park, J.E. and Lee, S.Y. (2020) Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem. Soc. Rev.*, **49**, 4615-4636.
4. Kavcscek, M., Strazar, M., Curk, T., Natter, K. and Petrovic, U. (2015) Yeast as a cell factory: current state and perspectives. *Microb. Cell Fact.*, **14**, 94.
5. Nielsen, J. (2019) Yeast systems biology: model organism and cell factory. *Biotechnol. J.*, **14**, e1800421.
6. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J. and Daran, J.M. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **15**, 1-15.
7. DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.*, **41**, 4336-4343.
8. Adiego-Pérez, B., Randazzo, P., Daran, J.M., Verwaal, R., Roubos, J.A., Daran-Lapujade, P. and van der Oost, J. (2019) Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiol. Lett.*, **366**, fnz086.
9. Wijsman, M., Swiat, M.A., Marques, W.L., Hettinga, J.K., van den Broek, M., Torre Cortés, P., Mans, R., Pronk, J.T., Daran, J.M. and Daran-Lapujade, P. (2019) A toolkit for rapid CRISPR-*SpCas9* assisted construction of hexose-transport-deficient *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.*, **19**, foy107.
10. Nielsen, J. and Keasling, Jay D. (2016) Engineering cellular metabolism. *Cell*, **164**, 1185-1197.
11. Kuijpers, N.G., Solis-Escalante, D., Luttkik, M.A., Bisschops, M.M., Boonekamp, F.J., van den Broek, M., Pronk, J.T., Daran, J.-M. and Daran-Lapujade, P. (2016) Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 15060-15065.
12. Clarke, L. and Carbon, J. (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, **287**, 504-509.
13. Larionov, V., Kouprina, N., Graves, J., Chen, X.N., Korenberg, J.R. and Resnick, M.A. (1996) Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 491-496.
14. Kouprina, N. and Larionov, V. (2006) TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat. Rev. Genet.*, **7**, 805-812.
15. Venetz, J.E., Del Medico, L., Wölflé, A., Schächle, P., Bucher, Y., Appert, D., Tschan, F., Flores-Tinoco, C.E., van Kooten, M., Guennoun, R. *et al.* (2019) Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8070-8079.
16. Benders, G.A., Noskov, V.N., Denisova, E.A., Lartigue, C., Gibson, D.G., Assad-Garcia, N., Chuang, R.-Y., Carrera, W., Moodie, M., Algire, M.A. *et al.* (2010) Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.*, **38**, 2558-2569.
17. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215-1220.

18. Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O. and Hutchison, C.A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 20404-20409.
19. Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M. *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**, 52-56.
20. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L. *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253.
21. Tagwerker, C., Dupont, C.L., Karas, B.J., Ma, L., Chuang, R.-Y., Benders, G.A., Ramon, A., Novotny, M., Montague, M.G., Venepally, P. *et al.* (2012) Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res.*, **40**, 10375-10383.
22. Karas, B.J., Molparia, B., Jablanovic, J., Hermann, W.J., Lin, Y.C., Dupont, C.L., Tagwerker, C., Yonemoto, I.T., Noskov, V.N., Chuang, R.Y. *et al.* (2013) Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng.*, **7**, 30.
23. Pretorius, I.S. and Boeke, J.D. (2018) Yeast 2.0-connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Res.*, **18**, foy032.
24. DiCarlo, J.E., Conley, A.J., Penttila, M., Jantti, J., Wang, H.H. and Church, G.M. (2013) Yeast oligo-mediated genome engineering (YOGE). *ACS Synth. Biol.*, **2**, 741-749.
25. Yu, S.C., Kuemmel, F., Skoufou-Papoutsaki, M.N. and Spanu, P.D. (2019) Yeast transformation efficiency is enhanced by TORC1- and eisosome-dependent signaling. *MicrobiologyOpen*, **8**, e00730.
26. Ji, Q., Mai, J., Ding, Y., Wei, Y., Ledesma-Amaro, R. and Ji, X.J. (2020) Improving the homologous recombination efficiency of *Yarrowia lipolytica* by grafting heterologous component from *Saccharomyces cerevisiae*. *Metab. Eng. Commun.*, **11**, e00152.
27. Shao, S., Ren, C., Liu, Z., Bai, Y., Chen, Z., Wei, Z., Wang, X., Zhang, Z. and Xu, K. (2017) Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing *Saccharomyces cerevisiae* Rad52. *Int. J. Biochem. Cell Biol.*, **92**, 43-52.
28. Shaked, H., Melamed-Bessudo, C. and Levy, A.A. (2005) High-frequency gene targeting in *Arabidopsis* plants expressing the yeast *RAD54* gene. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12265-12269.
29. Sánchez-Pascuala, A., Fernández-Cabezón, L., de Lorenzo, V. and Nikel, P.I. (2019) Functional implementation of a linear glycolysis for sugar catabolism in *Pseudomonas putida*. *Metab. Eng.*, **54**, 200-211.
30. Liu, Q., Yu, T., Li, X., Chen, Y., Campbell, K., Nielsen, J. and Chen, Y. (2019) Rewiring carbon metabolism in yeast for high level production of aromatic chemicals. *Nat. Commun.*, **10**, 4976.
31. Meadows, A.L., Hawkins, K.M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., Dahl, R.H., Tai, A., Mahatdejkul-Meadows, T., Xu, L. *et al.* (2016) Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature*, **537**, 694-697.
32. Aslan, S., Noor, E. and Bar-Even, A. (2017) Holistic bioengineering: rewiring central metabolism for enhanced bioproduction. *Biochem J.*, **474**, 3935-3950.
33. François, J.M., Lachaux, C. and Morin, N. (2020) Synthetic biology applied to carbon conservative and carbon dioxide recycling pathways. *Front. Bioeng. Biotechnol.*, **7**, 446.
34. Papagianni, M. (2012) Recent advances in engineering the central carbon metabolism of industrially important bacteria. *Microb. Cell Fact.*, **11**, 50.
35. van Rossum, H.M., Kozak, B.U., Pronk, J.T. and van Maris, A.J.A. (2016) Engineering cytosolic acetyl-coenzyme A supply in *Saccharomyces cerevisiae*: Pathway stoichiometry, free-energy conservation and redox-cofactor balancing. *Metab. Eng.*, **36**, 99-115.

36. Postma, E.D., Dashko, S., van Breemen, L., Taylor Parkins, S.K., van den Broek, M., Daran, J.M. and Daran-Lapujade, P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*
37. Putnam, C.D. and Kolodner, R.D. (2017) Pathways and mechanisms that prevent genome instability in *Saccharomyces cerevisiae*. *Genetics*, **206**, 1187-1225.
38. Boonekamp, F.J., Dashko, S., van den Broek, M., Gehrman, T., Daran, J.-M. and Daran-Lapujade, P. (2018) The genetic makeup and expression of the glycolytic and fermentative pathways are highly conserved within the *Saccharomyces* genus. *Front. Genet.*, **9**, 504.
39. Chen, B., Lee, H.L., Heng, Y.C., Chua, N., Teo, W.S., Choi, W.J., Leong, S.S.J., Foo, J.L. and Chang, M.W. (2018) Synthetic biology toolkits and applications in *Saccharomyces cerevisiae*. *Biotechnol. Adv.*, **36**, 1870-1881.
40. Decoene, T., De Maeseeneire, S.L. and De Mey, M. (2019) Modulating transcription through development of semi-synthetic yeast core promoters. *PLoS one*, **14**, e0224476.
41. Curran, K.A., Crook, N.C., Karim, A.S., Gupta, A., Wagman, A.M. and Alper, H.S. (2014) Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat. Commun.*, **5**, 4002.
42. Wang, J., Zhai, H., Rexida, R., Shen, Y., Hou, J. and Bao, X. (2018) Developing synthetic hybrid promoters to increase constitutive or diauxic shift-induced expression in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **18**, foy098.
43. Karas, B.J., Suzuki, Y. and Weyman, P.D. (2015) Strategies for cloning and manipulating natural and synthetic chromosomes. *Chromosome Res.*, **23**, 57-68.
44. Dhar, M.K., Sehgal, S. and Kaul, S. (2012) Structure, replication efficiency and fragility of yeast ARS elements. *Res. Microbiol.*, **163**, 243-253.
45. Naesby, M., Nielsen, S.V., Nielsen, C.A., Green, T., Tange, T.O., Simon, E., Knechtle, P., Hansson, A., Schwab, M.S., Titiz, O. *et al.* (2009) Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. *Microb. Cell Fact.*, **8**, 45.
46. Dance, A. (2017) Core Concept: Human artificial chromosomes offer insights, therapeutic possibilities, and challenges. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 9752-9754.
47. Ikeno, M. and Hasegawa, Y. (2020) Applications of bottom-up human artificial chromosomes in cell research and cell engineering. *Exp. Cell Res.*, **390**, 111793.
48. Xu, C., Hu, S. and Chen, X. (2016) Artificial cells: from basic science to applications. *Mater. Today*, **19**, 516-532.
49. Hughes, R.A. and Ellington, A.D. (2017) Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.*, **9**, a023812.
50. Su'etsugu, M., Takada, H., Katayama, T. and Tsujimoto, H. (2017) Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. *Nucleic Acids Res.*, **45**, 11525-11534.
51. Dean, F.B., Nelson, J.R., Giesler, T.L. and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.*, **11**, 1095-1099.
52. Jia, H. and Schwillie, P. (2019) Bottom-up synthetic biology: reconstitution in space and time. *Curr. Opin. Biotechnol.*, **60**, 179-187.
53. Mutschler, H., Robinson, T., Tang, T.D. and Wegner, S. (2019) Special issue on bottom-up synthetic biology. *Chembiochem*, **20**, 2533-2534.
54. Powell, K. (2018) How biologists are creating life-like cells from scratch. *Nature*, **563**, 172-175.



## Acknowledgements

This PhD in the industrial microbiology group (IMB), helped me challenge myself and discover my talents. I am therefore really happy that I embarked on this adventure four years ago. Due to the challenging nature of the PhD, I would not have been able to finish without the help of these incredible people.

**Pascale**, as my promoter and daily supervisor, I would like to thank you first and foremost. You are a very caring supervisor, who really pays attention to the well-being of your students. This shows in the intensive guidance at the very beginning of my PhD, which eased my integration into the industrial microbiology group, which I joined as quite an outsider. You were also the perfect supervisor to help me deal with my insecurity issues by providing me with a perfect balance of autonomy and support. Thank you for all the interesting and inspiring meetings on science and non-science related topics, you taught me a lot!

I also had the privilege to have another inspiring promoter. **Jack**, thank you for all your interesting (and sometimes crazy) ideas during the fun and fruitful brainstorm meetings in the first year of my PhD. Although, these meetings sadly came to an end after you became occupied with your role as head of the Biotechnology department, you still showed interest in my well-being and scientific progress. I appreciate that I could walk in unannounced and without appointment to discuss pressing issues. Finally, I am grateful that as former head of the IMB group, you created such a fun, caring and engaging working environment. The very valuable employee meetings and lab retreats, would not have been organized without your encouragements.

I would also like to thank the other PIs of the IMB group, **Jean-Marc, Robert, Rinke, Walter** and **Aljoscha** for their support and creation of an inspiring atmosphere. **Jean-Marc**, I really profited from your in-depth molecular knowledge at the ERC meetings and for the feedback on the manuscripts that we wrote together. I also admire how you kept IMB running during the difficult corona times. **Robert**, thank you for all the moments that I could walk into your office unannounced to ask yeast physiology questions.

I would not have been accepted as a PhD without the teaching of my former supervisors during my bachelor and master projects. **Antonella, Mariana, Paulien, Igor** and **Gilles van Wezel**, thank you for shaping me into the scientist that I am today.

I had the luck to be part of the (extended) ERC team, and without the critical input from its members this thesis would not have had the same outline as it has now. **Francine**, I admire your perseverance and like to thank you for your guidance with enzyme activity assays and involving me in the watermarking project. I also really enjoyed the

walks in het Haagse bos with you, where I could discuss the ups and downs of the PhD. Moreover, I feel privileged that you chose me as your paranymph. **Sofia**, thank you for paving the way and subsequently collaborating on many of my projects. You were my guide when I first came to IMB and I really appreciate the patience and thoroughness by which you taught me my lab skills. **Ewout**, I really enjoyed our collaboration on the puzzle of the lower growth rate of the SwYG strain. You always stayed calm and collected even when we thought we exhausted all possible explanations. Furthermore, I had great fun piecing together a cabaret with you. Therefore, I am really grateful that you also accepted to be my paranymph! **Jordi**, I could not have delivered this thesis without all your hard work, enthusiasm and ideas. **Melanie**, thank you for your guidance in the lab, all your very thorough work in my research lines and the nice chats outside the lab. **Charlotte**, I admire how you can simply explain difficult science, definitely something I can learn from. Thank you for your input during the ERC meetings, our collaboration on the HR review and organizing and participating in fun activities (art night!) in the group. Finally, I want to thank you for accepting to be my paranymph! **Mark**, thank you so much for your clear and patient teaching on the FACS, physiology calculations, statistics and much more. You are a born teacher and I am really grateful that you were part of IMB when I did my PhD. **Paola**, I could always go to you for any questions concerning CRISPR and I enjoyed our time together in the Molbio team. **Michal**, I appreciate all your input in the ERC meetings.

**Jasmijn**, I am thankful that you were willing to join me in the challenging anthocyanin project. Without your hard work, vast knowledge on aromatic compounds and dedication, this research line would not have reached its end. Furthermore, there would not have been as much *gezelligheid* in the group without you organizing fun stuff. **Flip**, I learned a lot about fermentations from you. It was fun running overnight bioreactors with you. I am so grateful that you took care of me and finished the fermentations when I fell ill in the morning! **Susan**, thank you for driving me home that day. **Koen, Patricia** and **Martin**, I am very grateful for your help with the intracellular metabolite samples and MS analysis.

It was a privilege to have supervised eight bachelor and master students. In chronological order: **Lars, Ilse, Roderick, Shannara, Roel, Venda, Anne-Marijn** and **Luuk**, you all made valuable contributions to my thesis. I found it fun to teach you your way around the lab and I definitely also learned a lot from all of you.

IMB would definitely not run as smooth without its incredible staff technicians. **Erik, Marcel, Pilar** and **Marijke**, we would all be lost without you! Erik, you are one of the persons that forms the heart of IMB, organizing /participating in almost all of the activities. Also, I like to thank you for all your help and advice on my fermentations! **Marcel**, thank you for all your bioinformatics help in all of my projects. I appreciate

that you took the time to explain to me what you had done in *Jip en Janneke taal*. **Pilar**, you are so kind and caring! Thank you for listening and giving advice when I did not immediately see the answers. Moreover, as queen bee of the Molbio lab you reigned with a kind but firm hand, keeping the lab running where I spent most time. **Marijke**, you are always there when needed for problems in and outside the lab. Thank you for your help with enzyme assays, flow cytometry but also for our private coaching sessions. **Tracey** and **Michelle**, the planning of the work discussions and my various meetings would not have gone as smoothly without your help.

This list would definitely not be complete without thanking the ladies from MSD: **Jannie, Apilena** and **Astrid**, thank you for thinking along, last-minute autoclaving and always making a nice chat when I came upstairs.

My time at IMB would not have been as enjoyable without the fun, cooperative and caring atmosphere in the group, to which all members contributed. **Anja, Nicolás, Anna** and **Wijb**, it was a pleasure to share an office with you. I loved the fun and jokes but also the scientific discussions and random questions. Moreover, Anna and Nicolás, I really enjoyed teaching MR together with you. I will never look at articles the same way again. **Erik** and **Sophie**, I think we organized a really nice lab retreat. I, at least, had lots of fun organizing it with you and will never forget cycling on a three-seat tandem. **Sanne, Jasmijn, Ewout** and **Wijb**, our trip to Sweden for the MESB19 course was unforgettable, full of learning, fun but also meaningful discussions on work and career. I am also very grateful to all the other IMB people that I did not mention up until now, **Aafke, Angelica, Aurin, Arthur, Christiaan, Hannes, Ioannis, Jasmine, Jasper, Jolanda, Jonna, Laura, Maaike, Maarten, Maria, Mario, Matthijs, Nicole, Nick, Raul, Thomas, Wesley** and **Xavier** for organizing and participating in all the fun events like: parties, art night, movie night, Sinterklaas, game night and the borrels at the botanical garden.

Tijdens mijn PhD ben ik niet alleen meer te weten gekomen over gist, DNA en chromosomen maar ook meer over mijzelf. **Louise van Swaij** ik wil je bedanken voor jouw aandeel hierin.

Als laatste wil ik graag mijn vrienden en familie bedanken die altijd een welkome afleiding vormden als ik iets te veel in mijn PhD werd opgeslokt en lieten merken dat er ook nog een wereld daarbuiten was. Ik kan niet iedereen bedanken, sorry daarvoor, maar ik wil een paar specifieke mensen noemen.

Al zijn jullie te veel om op te noemen, ik wil iedereen van de **Reddingsbrigade Westerschouwen** bedanken voor de leuke tijd op strand en daarbuiten. Als je op het strand bent denk je werkelijk aan niets anders meer en daarom de perfecte manier om even loslaten. **Dennis, Jolie, Martijn, Dagmar** en **Laurens**, jullie hebben me echt in de

groep opgenomen. Ik heb erg genoten van de borrels, vakanties, golven en vele ander activiteiten die we samen hebben gedaan. De CollegeBeesten: **Laurens, Darlyne, Evelyne, Huub, Maarten, Meike, Reinoud, Sebastiaan, Simone, Tim** en **Vera**, al zien we elkaar een stuk minder sinds we allemaal zijn uitgewaaierd uit Groningen, geniet ik des te meer van de uitjes die we eens per jaar organiseren. Meiden van de groep, wij zien elkaar een stuk meer, zeker sinds het begin van Corona hebben de Skype sessies voor veel plezier bij mij gezorgd. **Erica** en **Eva**, al 23 jaar kan ik op jullie terugvallen als er iets is. Ik hoop dat we nog heel lang samen blijven shoppen, kletsen, theetjes doen *etc.*

**Tante Greet** en **ome Henny**, bedankt dat jullie als peetouders bij alle belangrijke gebeurtenissen in mijn leven aanwezig zijn geweest en voor alle gezellige diners. **Ome Theo** en **tante Judith**, over de jaren heen heb ik genoten van de leuke sinterklaasvieringen. **Marjon** en **Sander**, ik kan jullie niet genoeg bedanken dat jullie me in huis hebben genomen toen ik geen kamer kon vinden in Leiden. Dit samenwonen heeft een mooie vriendschap opgeleverd met veel spelletjes en filmavonden als gevolg. **Johan, Petra, Menno, Pake** en **Beppe** bedankt voor een thuis nu papa en mama zo ver in Frankrijk zitten.

Lieve **papa** en **mama**, jullie hebben me gevormd tot de persoon die ik vandaag ben. Ik heb zoveel van jullie geleerd en dat zal ik ongetwijfeld blijven doen. Ik denk dat door jullie liefde voor de wetenschap Rolf en ik daar toch uiteindelijk in terecht gekomen zijn en daar ben ik persoonlijk erg blij door. Mama, ik wil jou bedanken voor de prachtige cover die je hebt gemaakt voor mijn proefschrift. Ik bewonder dat jij nooit bang bent om een nieuwe uitdaging aan te gaan. Ook hebben jouw passie voor discussiëren me een bredere blik op de wereld gegeven die altijd van pas zal komen. Papa, het is raar hoe dingen kunnen lopen, en dat we uiteindelijk in dezelfde groep zijn gepromoveerd. Wel heel leuk dat we daarom samen lekker kunnen discussiëren over gist en genetische modificatie. Bedankt voor je advies, interessante vragen en wilde ideeën, dat deed me soms weer heel anders naar mijn onderzoek kijken. Jij bent echt iemand die out-of-the-box kan denken. **Rolf**, als mijn grote broer wil ik je bedanken voor alles wat je voor me hebt gedaan tijdens mijn gehele leertijd. Ik weet niet hoeveel uur je wel niet hebt gespendeerd aan dingen uitleggen en stukken nakijken, zelfs al werd dat soms niet altijd even lief ontvangen van mijn kant. Ik ben blij dat je **Maria** hebt uitgekozen als je partner en het is altijd gezellig als we elkaar weer zien.

Lieve **Wessel**, als laatste wil ik jou bedanken. Zonder jou zou ik deze promotie niet kunnen hebben afgerond. Jij was er altijd voor me als het tegen zat en kon de ernst van problemen goed voor mij relativeren. Ook had je een luisterend oor als ik (soms onsamenhangend) over mijn onderzoek, collega's en manuscripten aan het vertellen was. Ik hou van je en kijk uit naar al onze nieuwe avonturen samen!

## Curriculum Vitae

Eline Daphne Postma was born on the 30<sup>th</sup> of December 1993 in Hoorn, The Netherlands. She performed her primary education in Berkel and Rodenrijs and attended the first year of high school at the Stanislascollege in Pijnacker. She subsequently moved to Lyon, France where she completed the International General Certificate of Secondary Education (IGCSE) and the International Baccalaureate (IB) at the Ecole et Lycée Bilingues de Lyon Ombrosa. Next, in 2011 Eline enrolled in the BSc programme Life Sciences and Technology at Groningen University and performed her BSc thesis project in the Molecular Microbiology group of Prof. dr. Arnold Driessen under the supervision of dr. Antonella Caforio. During her studies, she was active in the congress committee of the study association GLV Idun and commenced as a Lifeguard at Reddingsbrigade Westerschouwen. After obtaining her BSc degree (cum laude) in 2014, Eline started a MSc study in Biology at Leiden University. Specializing in Microbial Biotechnology and Health, she performed two nine-month internships. The first, academic research project focused on genes responsible for biosynthesis of volatile organic antibiotic compounds in *Streptomyces* and was performed in the Microbial Sciences group, under supervision of Prof. dr. Gilles van Wezel and dr. Mariana Avalos Garcia. A second, industrial project at Dupont in Leiden focused on increasing transformation efficiency of the filamentous fungus *Trichoderma reesei* and was supervised by Paulien Kruithof and dr. Igor Nikolaev. After obtaining her MSc in 2016 (summa cum laude), Eline commenced a PhD project at the Department of Biotechnology of the Delft University of Technology, under supervision of Prof. dr. Pascale Daran-Lapujade and Prof. dr. Jack Pronk, of which the results can be found in this thesis. As from May 2021 Eline is working as an Analytical Writer at ICON plc in Assen.

## List of publications

- Boonekamp, F. J., Dashko, S., Duiker, D., Gehrman, T., van den Broek, M., den Ridder, M., Pabst, M., Robert, V., Abeel, T., **Postma, E.D.**, Daran, J.-M., Daran-Lapujade, P. (2020) Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences. *ACS Synthetic biology*, **9**, 1361–1375.
- **Postma, E.D.**, Dashko, S., van Breemen, L., Taylor Parkins, S.K., van den Broek, M., Daran, J.-M. and Daran-Lapujade, P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, **49**, 1769-1783.
- **Postma, E.D.**, Couwenberg, L.G.F., van Roosmalen, R.N., Geelhoed, J., de Groot, P.A., Daran-Lapujade, P. (2021) Top-down, knowledge-based genetic reduction of yeast central carbon metabolism. *Submitted for publication in mBio*.
- **Postma, E.D.**, Hassing, E.-J., Mangkusaputra, V., Geelhoed, J., de la Torre, P., van den Broek, M., Mooiman, C., Pabst, M., Daran, J.-M. and Daran-Lapujade, P. (2021) Modular, synthetic chromosomes as new tools for large scale engineering of metabolism. *Submitted for publication in Metabolic Engineering*.
- **Postma, E.D.**, Knibbe, E., Boonekamp, F. J., Dashko, S., Geelhoed, J., Maat, A.-M., van den Broek, M., Daran-Lapujade, P. (2021) What's wrong with SwYG? Reaching the limits of *Saccharomyces cerevisiae* molecular and analytical toolbox. *Manuscript in preparation for submission*.
- Koster, C.C., **Postma, E.D.**, Knibbe, E., Cleij, C., Daran-Lapujade, P. (2021) Synthetic genomics from a yeast perspective. *Manuscript in preparation for submission*.

