Delft University of Technology

Modular engineering of synthetic glycolytic pathways in Saccharomyces cerevisiae

Boonekamp, F.J.

DOI
10.4233/uuid:4cd6d858-5f09-4567-86e8-f9f61ca7941f

Publication date
2020

Document Version
Final published version

Citation (APA)
Boonekamp, F. J. (2020). *Modular engineering of synthetic glycolytic pathways in Saccharomyces cerevisiae*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:4cd6d858-5f09-4567-86e8-f9f61ca7941f

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.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Modular engineering of synthetic glycolytic pathways in *Saccharomyces cerevisiae*

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
Woensdag 21 oktober 2020 om 12:30 uur

door

Francine Judith BOONEKAMP

Master of Science in Environmental Biology, Universiteit Utrecht
geboren te Leidschendam, Nederland

Dit proefschrift is goedgekeurd door de promotoren

Prof. dr. P.A.S. Daran-Lapujade and 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
Prof. dr. B.M. Bakker	Universitair Medisch Centrum Groningen
Onafhankelijke leden:	
Prof. dr. M. Dogterom	Technische Universiteit Delft
Prof. dr. R. A. Weusthuis	Universiteit Wageningen
Prof. dr. J.H. de Winde	Universiteit Leiden
Dr. P.I. Nikel	DTU Biosustain, Denemarken
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 Science, Delft University of Technology, The Netherlands and was funded by a consolidator grant AdLibYeast from the European Research Council (ERC).

Cover: Marianne Boonekamp
Layout: Francine Boonekamp
Printed by: Ipskamp Printing B.V.
ISBN: 978-94-6421-020-0

© 2020 Francine J. Boonekamp
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

Samenvatting……5

Summary……9

Chapter 1: Introduction……13

Chapter 2: Pathway swapping: towards modular engineering of essential cellular processes……39

Chapter 3: The genetic makeup and expression of the glycolytic and fermentative pathways are highly conserved within the *Saccharomyces* genus……73

Chapter 4: Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences……109

Chapter 5: A yeast with muscle doesn't run faster: full humanization of the glycolytic pathway in *Saccharomyces cerevisiae*……153

Outlook……231

Acknowledgements……233

Curriculum vitae……236

List of publications……237

Samenvatting

Microbiële fermentatieprocessen worden al eeuwen gebruikt voor de productie van zuivelproducten, alcoholische dranken en brood. In de afgelopen decennia heeft de biotechnologie een enorme ontwikkeling doorgemaakt en tegenwoordig wordt een breed scala aan producten, variërend van biobrandstoffen tot chemicaliën en geneesmiddelen geproduceerd met behulp van micro-organismen. De ontwikkeling van technieken voor genetische modificatie heeft hier in grote mate aan bijgedragen. Doordat micro-organismen hernieuwbare grondstoffen kunnen gebruiken voor het produceren van brandstoffen en chemicaliën, biedt de inzet van micro-organismen een duurzaam alternatief voor productie van deze stoffen op basis van fossiele bronnen. De gist *Saccharomyces cerevisiae,* ook wel bakkersgist genoemd speelt een belangrijke rol in industriële biotechnologie. De populariteit van deze gist in zowel toegepast onderzoek als in de industrie is toe te schrijven aan een aantal belangrijke eigenschappen. Zo kan deze gist met grote snelheid suikers fermenteren, heeft hij een hoge tolerantie voor lage pH, kan tegen hoge suiker- en alcoholconcentraties en is genetisch makkelijk hanteerbaar. Voor *S. cerevisiae* is een uitgebreide set tools aanwezig om het genoom te modificeren. Dit maakt het mogelijk om, door het tot expressie brengen van genen (van andere organismen), nieuwe eigenschappen aan gist toe te voegen zoals recentelijk werd geïllustreerd met de succesvolle biosynthese van opioïden in gist. Ondanks deze enorme vooruitgang blijft het echter nog steeds lastig en tijdrovend om op grote schaal modificaties aan te brengen aan de van nature aanwezige metabole routes in gist. Dit kan voor een groot deel worden verklaard door het grote aantal 'dubbele' genen dat aanwezig is in het genoom die coderen voor iso-enzymen die dezelfde reactie katalyseren. Daarnaast liggen de genen die betrokken zijn bij een bepaalde metabole route verspreid over het gehele genoom, wat het moeilijk maakt al deze genen in één keer te bewerken en wat het zeer arbeidsintensief en tijdrovend maakt. Het doel van dit onderzoek was om een strategie te ontwerpen en te testen om het op grote schaal modificeren van (essentiële) metabole routes makkelijker te maken, door middel van het simplificeren en reorganiseren van het genoom. Het uitgangspunt van dit onderzoek is het centrale koolstofmetabolisme, en in het bijzonder de glycolyse.

Met behulp van de glycolyse als paradigma, onderzoekt **hoofdstuk 2** een strategie die bestaat uit het verwijderen van overtollige genen en het verplaatsen van de resterende genen naar één enkel chromosomaal locus, wat het mogelijk maakt om een hele (stofwisselings)pathway in een paar eenvoudige stappen te vervangen. Glycolyse, een van de meest intensief bestudeerde metabole pathways in gist, is de centrale route voor suikermetabolisme. De glycolyse vormt een set van 12 reacties die worden gekatalyseerd door 26 iso-enzymen die gecodeerd worden door een set van 26 genen.

Het uitgangspunt van dit onderzoek is een stam met een 'geminimaliseerde glycolyse' (de MG-stam genoemd) waarin de set van 26 genen is teruggebracht naar slechts 13, door alle overtollige genen te verwijderen. **Hoofdstuk 2** introduceert het concept van 'pathway swapping', gebaseerd op de verplaatsing van de minimale set glycolyse genen naar één enkel chromosomaal locus. Om dit te bereiken, werd de minimale set van 13 genen geclusterd in één chromosomaal locus, gevolgd door de verwijdering van de overeenkomstige 13 genen van hun oorspronkelijke locaties op de verschillende chromosomen. De resulterende stam vertoonde, afgezien van een iets lagere groeisnelheid, een zeer vergelijkbare fysiologie in vergelijking met de MG-stam. In deze stam met een uitwisselbare gist glycolyse kan de gehele glycolyse route in twee eenvoudige stappen worden verwisseld met een andere (heterologe) variant. Om het potentieel van deze stam te testen, werd de gehele glycolyse van *S. cerevisiae* verwisseld met de glycolyse van een verwante gistsoort *Saccharomyces kudriavzevii* en met een pathway bestaande uit een mix van genen van *S. cerevisiae, S. kudriavzevii* en menselijke genen. Verrassend genoeg hadden deze glycolyse-wissels nauwelijks invloed op de fysiologie van de stammen. De resultaten van de modulaire benadering die in deze studie wordt gebruikt, zijn veelbelovend voor het op grotere schaal herorganiseren van het gistgenoom. Bovendien is dit glycolyse-wisselplatform een uitstekend hulpmiddel om een beter inzicht te krijgen in glycolyse en de complexe regulatie ervan.

Een van de huidige beperkingen in het op grote schaal aanpassen van het metabole netwerk in gist is de beschikbaarheid van goed gekarakteriseerde, sterke, constitutieve promotors. Het doel van **hoofdstuk 3** was om de beschikbare set moleculaire tools voor *S. cerevisiae* te verrijken met dergelijke promotoren door gebruik te maken van de biodiversiteit van andere gistsoorten. De promotors die de glycolyse genen tot expressie brengen in *S. cerevisiae* behoren tot de sterkste promotors in de cel en zijn daarom populair voor stamconstructie. In deze studie werd onderzocht of de glycolyse promotors van de verwante gistsoorten *S. kudriavzevii* en *Saccharomyces eubayanus* ook sterke en constitutieve genexpressie in *S. cerevisiae* aan kunnen drijven. Omdat in tegenstelling tot *S. cerevisiae*, vrijwel niets bekend was over de genetische samenstelling en expressieniveaus van de glycolysegenen van *S. kudriavzevii* en *S. eubayanus*, werd de glycolyse route eerst gekarakteriseerd in zijn oorspronkelijke context en vergeleken met die van *S. cerevisiae.* Het sequencen van het genoom van deze gisten onthulde een opmerkelijk sterk geconserveerde genetische samenstelling van de glycolyse routes in de drie soorten wat betreft het aantal aanwezige paraloge genen. Hoewel de promotorsequenties minder goed geconserveerd waren dan coderende sequenties, waren de bindingslocaties voor de belangrijkste glycolyse regulatoren Rap1, Gcr1 en Abf1 in hoge mate geconserveerd tussen de drie *Saccharomyces* gisten.

Ondanks fysiologische verschillen tussen de drie soorten, bleken de expressieniveaus van de glycolysegenen gedurende aerobe batchfermentaties in chemisch gedefinieerd medium met glucose als enige koolstofbron opmerkelijk vergelijkbaar te zijn tussen de drie soorten. Vervolgens werden de promotors van de belangrijkste paralogen van *S. kudriavzevii* en *S. eubayanus* getransplanteerd naar *S. cerevisiae* en hun activiteit werd gevolgd onder verschillende groeiomstandigheden met behulp van het fluorescerende eiwit mRuby2. De resultaten van deze studie lieten zien dat de activiteit van de promoters van *S. kudriavzevii* en *S. eubayanus* sterk en constitutief was en opmerkelijk vergelijkbaar met hun tegenhangers in *S. cerevisiae.* In combinatie met de relatief lage homologie ten opzichte van de *S. cerevisiae*-promotors, zijn de promotors van *S. kudriavzevii* en *S. eubayanus* een zeer aantrekkelijk alternatief voor de constructie van stammen in *S. cerevisiae,* waardoor de set beschikbare moleculaire tools voor *S. cerevisiae* wordt uitgebreid.

Bij grootschalige reorganisatieprojecten van het genoom zoals beschreven in hoofdstuk 2, is er een toenemende behoefte aan mogelijkheden om onderscheid te kunnen maken tussen synthetische en natuurlijke kopieën van een gen in de cel. Gecombineerd met bio-informatica-tools, is het watermerken van DNA, de introductie van stille mutaties in een gen, een methode die onderscheid kan maken tussen natuurlijke en synthetische (watermerk) allelen van een gen op DNA- en mRNA-niveau. Hoewel het watermerken van genen op grote schaal wordt gebruikt en goed is onderzocht dat in eukaryoten het gebruik van alternatieve codons de translatie van een eiwit en hoogstwaarschijnlijk de mRNA-stabiliteit kan beïnvloeden, zijn er opmerkelijk weinig kwantitatieve studies die de impact van watermerken op transcriptie, eiwitexpressie en fysiologie in *S. cerevisiae* onderzoeken. **In hoofdstuk 4** werd een strategie ontwikkeld om genen systematisch te watermerken met als doel de fysiologie van de gist minimaal te beïnvloeden en werd deze strategie vervolgens geïmplementeerd en experimenteel gevalideerd. De 13 genen die coderen voor eiwitten die betrokken zijn bij glycolyse, werden gelijktijdig van 10-12 watermerken voorzien en tot expressie gebracht in *S. cerevisiae* met behulp van het glycolyse-'pathway swapping' concept beschreven in hoofdstuk 2. Het introduceren van watermerken in de glycolysegenen die van nature sterk tot expressie komen en die gebruik maken van de meest optimale codons, had geen invloed op transcriptie, enzymactiviteit en gistfysiologie, met uitzondering van het gen *GPM1*. De *markerQuant* bio-informatica-methode kon betrouwbaar natuurlijke van synthetische (watermerk) genen en transcripten onderscheiden. Verder werd aangetoond dat de watermerken ook selectieve CRISPR/Cas9 genetische modificatie mogelijk maakten, door modificatie alleen op de natuurlijke kopie te richten terwijl de synthetische, van een watermerk

voorziene variant intact bleef. Deze studie biedt een eenvoudige en gevalideerde watermerkstrategie die kan worden toegepast in *S. cerevisiae*.

Naast zijn belangrijke rol in de biotechnologie, is *S. cerevisiae* ook een populair modelorganisme voor hogere eukaryoten. Het tot expressie brengen van menselijke genen in gist is een veel gebruikte strategie om de functionaliteit van een gen te onderzoeken en om medicijnen te testen. Het hoge aantal 'dubbele' en daarmee overtollige genen in het genoom van eukaryoten en het gebrek aan moleculaire tools om het genoom op grote schaal te kunnen herorganiseren hebben ertoe geleid dat studies tot nu toe vooral gefocust waren op het vervangen en testen van één enkel gen. Met de ontwikkeling van stammen zoals de MG- en SwYG-stam zoals beschreven in hoofdstuk 2 komt het vermenselijken van volledige pathways of processen binnen de mogelijkheden te liggen. Als bewijs hiervan wordt in **hoofdstuk 5** de volledige vermenselijking van de glycolyse pathway beschreven. Door de combinatie van het afzonderlijk testen van genen, het vermenselijken van de volledige pathway en laboratoriumevolutie werd de functionaliteit van 25 menselijke enzymen in *S. cerevisiae* onderzocht. Het resultaat hiervan was dat behalve de hexokinase enzymen *Hs*HK1, *Hs*HK2 en *Hs*HK3 alle 25 geteste menselijke genen in staat waren de katalytische functie van hun gist-ortholoog over te nemen. De aldolase en enolase enzymen waren daarnaast ook in staat de secundaire functie over te nemen. Resultaten van laboratoriumevolutie suggereerden een opmerkelijke verscheidenheid aan cellulaire mechanismen die werden ingezet om de groei van stammen met volledig vermenselijkte glycolyse te optimaliseren, zoals de afgifte van aan actine gebonden aldolase. Uit vergelijking met menselijke spiercellen bleek tenslotte dat voor de meeste geteste menselijke enzymen transplantatie in gist hun activiteit (k_{cat}) niet beïnvloedde. Giststammen met volledig vermenselijkte glycolyse pathways waarin de enzymen in een natuurgetrouwere context bestudeerd kunnen worden zijn veelbelovende modellen om meer te leren over menselijke cellen.

Abstract

Already for millennia, microbial fermentation is used for the production of dairy products, alcoholic beverages and bread. In the last decades, the field of biotechnology has tremendously expanded and nowadays, a wide range of compounds ranging from biofuels to chemicals and pharmaceuticals is produced using microbial cell factories. The development of genetic engineering tools has greatly contributed to this rapid development. Catalysing the conversion of renewable carbohydrate feedstocks into fuels and chemicals, microbial cell factories offer a sustainable alternative to fossil resources-based production, and thereby contribute to reduce greenhouse gas emissions. The yeast *Saccharomyces cerevisiae* plays an important role in industrial biotechnology. Its popularity for applied research and industrial production can be attributed to several factors as its fast fermentative metabolism, its tolerance to low pH, high sugar and alcohol concentrations and its genetic tractability. *S. cerevisiae* possesses one of the best furbished molecular toolboxes, which makes it possible to assemble complex heterologous pathways, as was recently illustrated by the successful biosynthesis of opioids in yeast. Despite this great progress, extensive genetic remodelling of native pathways remains challenging. This can largely be explained by the high genetic redundancy present in the yeast genome, in which multiple genes encode proteins with redundant functions, and by the fact that the genes belonging to a pathway are scattered over the entire genome. The goal of this thesis was to design, set up and validate a strategy aiming at facilitating the remodelling of (essential) pathways, based on simplifying and reorganizing the yeast genome. The starting point of this research is the central carbon metabolism and in particular, as proof of concept, the glycolytic pathway.

Using the glycolytic and fermentative pathways as paradigm, **Chapter 2** explores a strategy consisting in removing redundant genes and relocating the remaining glycolytic and fermentative genes to a single chromosomal location, which would make it possible to replace a whole pathway in a few simple steps. Glycolysis, one of the most intensively studied pathways in yeast, is the central pathway for sugar metabolism. Together with the fermentative pathway, glycolysis forms a set of 12 reactions catalysed by 26 isoenzymes encoded by a set of 26 paralogs. The starting point of this research is a strain with Minimal Glycolysis and fermentation pathway (called the MG strain) in which the set of 26 paralogous genes has been reduced from 26 to 13. Chapter 2 introduces the pathway swapping concept, based on the relocalization of the minimal set of glycolytic and fermentative genes to a single chromosomal locus. To achieve this, the minimal set of 13 genes was assembled in a single chromosomal locus, followed by

the removal of the corresponding 13 genes from their native locations across the different chromosomes. Based on the MG strain, the newly constructed strain harboured a single locus glycolysis and displayed, apart from a slightly lower growth rate, a very similar physiology as compared to the MG strain. In this strain with Switchable Yeast Glycolysis (SwYG) the entire glycolytic pathway can be swapped with any another (heterologous) variant in two simple steps. To test the potential of this strain, the entire glycolytic and fermentative pathways of *S. cerevisiae* were swapped with the pathways of a related yeast species *Saccharomyces kudriavzevii* and with pathways consisting of a mix of genes from *S. cerevisiae, S. kudriavzevii* and *Homo sapiens*. Surprisingly, these glycolysis swaps hardly affected the physiology of the strains. The results of the modular engineering approach used in this study are very promising for remodelling of the yeast genome at a larger scale. In addition, this glycolysis swapping platform is an excellent tool to study glycolysis and its regulation.

One of the current limitations in large scale metabolic engineering is the availability of well characterized, strong, constitutive promoters. The aim of **Chapter 3** was to enrich the *S. cerevisiae* molecular toolbox with such promoters by exploring biodiversity. In *S. cerevisiae,* glycolytic and fermentative promoters are amongst the strongest promoters and are therefore popular for strain construction. This study explored the potential of glycolytic and fermentative promoters of the related yeast species *S. kudriavzevii* and *Saccharomyces eubayanus* to drive strong and constitutive gene expression in *S. cerevisiae* as well. As, in contrast to *S. cerevisiae,* virtually nothing was known about the genetic makeup and expression of the fermentative and glycolytic pathways of *S. kudriavzevii* and *S. eubayanus*, the pathways were first characterized in their native context and compared to that of *S. cerevisiae*. Sequencing data revealed a remarkably highly conserved genetic makeup of the glycolytic and fermentative pathways in the three species in terms of number of paralogous genes. Although the promoter sequences were less well conserved than coding regions, the binding sites for the main glycolytic regulators Rap1, Gcr1 and Abf1 were highly conserved between the three *Saccharomyces* species. Despite physiological differences between the three species, transcriptome analysis from aerobic batch fermentations in chemically defined medium with glucose as sole carbon source revealed a remarkably similar expression of the glycolytic and fermentative genes across species. Subsequently, the promoters of the major paralogs of *S. kudriavzevii* and *S. eubayanus* were transplanted to *S. cerevisiae* and their activity was monitored using the fluorescent protein mRuby2 under an array of growth conditions. This study revealed that the activity of the *S. kudriavzevii* and *S. eubayanus* promoters was strong and constitutive, and remarkably similar to their *S. cerevisiae* counterparts. In combination with the relatively low homology to their *S.*

cerevisiae promoters, the promoters of *S. kudriavzevii* and *S. eubayanus* are a very attractive alternative for strain construction in *S. cerevisiae*, thereby expanding the *S. cerevisiae* molecular toolbox.

In large scale remodelling projects, such as described in chapter 2, there is an increasing need for possibilities to distinguish between synthetic and native gene copies in a cell. Combined with bioinformatics tools, DNA watermarking, the introduction of silent mutations in a gene, is a method that can discriminate between native and watermarked alleles of a gene at DNA and mRNA level. Although watermarking is widely used and it is well documented that codon usage can affect translation, and most likely mRNA stability in eukaryotes, there are remarkably few quantitative studies that explore the impact of watermarking on transcription, protein expression and physiology in *Saccharomyces cerevisiae*. In **Chapter 4**, a systematic watermarking strategy, with the aim to minimally affect the yeast physiology, was designed, implemented and experimentally validated. The thirteen genes encoding proteins involved in glycolysis and alcoholic fermentation were simultaneously watermarked and expressed in *S. cerevisiae* using the glycolysis swapping concept described in Chapter 2. The codon changes which were introduced in the naturally codon optimized, highly expressed glycolytic genes, did not affect transcript abundance, enzyme activity and yeast physiology, with the notable exception of *GPM1*. The *markerQuant* bioinformatics method could reliably discriminate native from watermarked genes and transcripts. Furthermore, it was shown that the watermarks also enabled selective CRISPR/Cas9 genome editing, by targeting only the native gene copy while leaving the synthetic, watermarked variant intact. This study offers a simple and validated watermarking strategy which can be applied in *S. cerevisiae.*

Next to its important role in biotechnology, *S. cerevisiae* is also popular as a model for higher eukaryotes. Humanization of genes in yeast is a widely used strategy to explore gene functionality and test drugs, thereby improving yeast as metazoan model. Hindered by the high genetic redundancy of eukaryotic genomes and the lack of molecular tools for large scale genome remodelling, to date humanization studies have mostly focused on single gene complementation. With the development of strains such as the MG and SwYG strain as described in Chapter 2, these challenges can be overcome and bring humanization of full pathways or processes within reach. As proof of principle, **Chapter 5** demonstrates the full humanization of the glycolytic pathway. Combining single gene complementation, full pathway humanisation and laboratory evolution, the functionality of 25 human enzymes in *S. cerevisiae* was explored. Except for the hexokinases *Hs*HK1, *Hs*HK2 and *Hs*HK3, all 25 tested human genes were able to complement the catalytic function of their yeast orthologs and aldolase and enolases

also complemented their moonlighting functions. Laboratory evolution suggested a remarkable variety of cellular mechanisms deployed to optimize the growth of strains with fully humanized glycolysis, such as the release of actin-bound aldolase. Finally, comparison with skeletal muscle cells showed that, for most tested human enzymes, transplantation in yeast did not affect their turnover number (k_{cat}). Enabling to study the enzymes in a context closer to their native environment, yeast strains with fully humanized glycolytic pathways are promising models for metazoans.

Chapter 1

General Introduction

Saccharomyces cerevisiae and its role in industrial biotechnology

Microbial fermentation is a process which is already used for millennia to preserve the quality and safety of food and has been of great economic and social relevance throughout history (*1*). The earliest evidence of the use of fermentation processes for the production of fermented beverages dates already from 7000 BC (*2*). Although humans were until quite recently unaware of its existence, one of the major organisms responsible for alcoholic fermentation is *Saccharomyces cerevisiae,* also called baker's yeast. In the 17th century, yeast was already observed under the microscope by Antoni van Leeuwenhoek, but only in 1857 was alcoholic fermentation connected to Baker's yeast by Louis Pasteur (*3, 4*). After the discovery of yeast's prominent role in wine, beer and bread making throughout human history, in the 1970s, the development of genetic engineering tools led to a rapid development of new biotechnological applications for yeast (*5*). Genetically engineered yeast is currently used for the industrial production of a wide range of products, which range from fuels, chemicals, pharmaceuticals to nutraceuticals. Some examples of products which are made by *S. cerevisiae* as microbial cell factory are bioethanol, organic acids such as lactic acid which is used for the production of the biodegradable plastic poly-lactic acid, and pharmaceutical products like insulin, the antimalarial drug precursor artemisinic acid and vaccines (*6-9*). In view of current challenges to reduce greenhouse gas emissions there is an increasing interest in the use of microbial cell factories for the production of chemicals and fuels, since microbial production from renewable carbohydrate feedstocks offers a sustainable alternative to fossil resources-based production.

Saccharomyces cerevisiae's fermentative life-style

In *S. cerevisiae* two modes of sugar metabolism can be identified: respiration and alcoholic fermentation. An important trait of baker's yeast which has played a key role in its domestication, is its strong preference for fermentative dissimilation of sugars, even in the presence of oxygen. Under aerobic conditions *S. cerevisiae* exhibits at high specific growth rates a mixed, respirofermentative metabolism, irrespective of the mode of cultivation (i.e. growth with excess sugar or sugar-limited). This phenomenon is referred to as the Crabtree effect (*10*). Alcoholic fermentation, the conversion of glucose to two molecules of ethanol and CO_2, yields only two moles of ATP per mole of glucose, resulting from substrate level phosphorylation. To obtain sufficient ATP for biomass formation and maintenance, *S. cerevisiae* and other Crabtree positive yeasts maintain high glycolytic fluxes, and consequently high alcohol production rates (Christen 2010).

In batch cultivation mode, when sugars are depleted, *S. cerevisiae* can switch under aerobic conditions to a fully respiratory metabolism and consume the ethanol and organic acids accumulated during the respirofermentative growth phase. This switch in metabolism is referred to as diauxic shift (*11*). It is speculated that this 'make-accumulate-consume' strategy of *S. cerevisiae* has a selective advantage over other microorganisms which have in most cases a lower ethanol tolerance than *S. cerevisiae* (*12*). This strategy is shared among the naturally occurring species belonging to the *Saccharomyces* genus, which all evolved towards optimal performance in their different niches (*13, 14*) (Fig. 1). This has led to different physiological characteristics, as for example the increased cold tolerance of *S. kudriavzevii* and *S. eubayanus* as compared to *S. cerevisiae* (*15*). Hybridization between different species of the *Saccharomyces* genus has further increased the biodiversity and a number of *Saccharomyces* species and hybrids play an important role in wine and beer brewing, such as the lager brewing yeast *S. pastorianus*, (*16*).

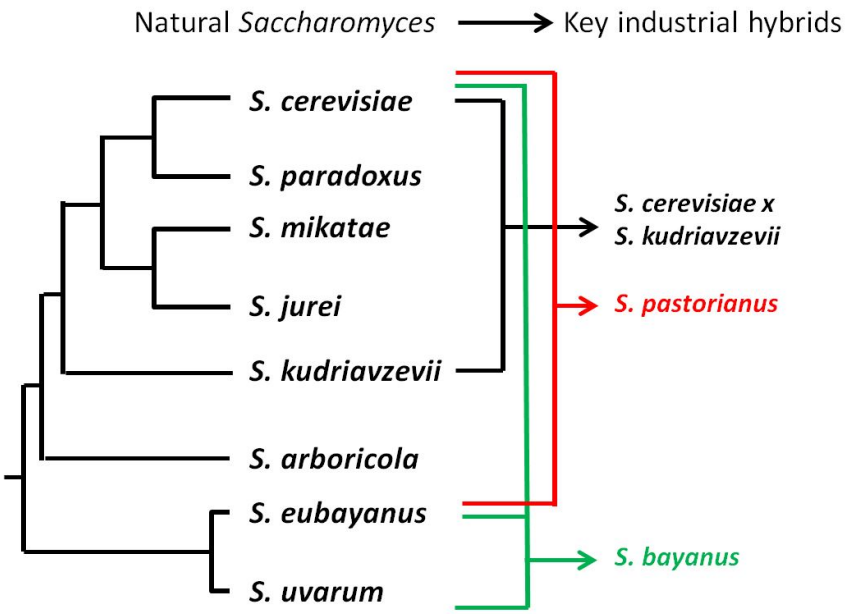


Figure 1 – Phylogenetic tree of the *Saccharomyces* genus.
Industrial hybrids resulting from hybridization events between natural *Saccharomyces* species are shown on the right side. Figure adapted from (*14*) and (*17*).

The glycolytic pathway of *Saccharomyces cerevisiae*

Structure and function of the glycolytic pathway

Glycolysis is the most common pathway for sugar catabolism across kingdoms of life. It has, due to its central role in most organisms and its economical relevance, extensively been studied. Next to its role in supply of energy and redox equivalents (ATP, NADH) in the cell, glycolysis plays an important role in precursor supply to biosynthetic pathways as the pentose phosphate pathway, glycerol metabolism (precursor for lipid synthesis) and amino acid biosynthesis routes (*18*) (Fig. 4). In addition, glycolysis provides precursors for metabolic stress protectants (trehalose and glycerol) (*19, 20*) and carbohydrate storage metabolism (trehalose and glycogen) (*20, 21*).

The predominant variant of glycolysis in nature is the Embden-Meyerhof -Parnas (EMP) pathway, which is near ubiquitous in eukaryotes (*22, 23*). Especially in prokaryotes a wide variety of glycolytic variants exist, of which the Entner-Doudoroff (ED) pathway is the most common. The main difference between the EMP and ED pathway is that the ED pathway yields only half of the ATP as compared to the EMP pathway. Often prokaryotes contain both the EMP and ED pathway (*23*).

Within the EMP pathway, from now on referred to as 'glycolysis', one molecule of glucose is oxidized in ten steps to two molecules of pyruvate. This process, called substrate level phosphorylation, yields net two ATP and two NADH molecules per glucose molecule. In *S. cerevisiae*, if pyruvate is not respired, it is reduced via a two-steps fermentative pathway in which pyruvate is first decarboxylated into acetaldehyde and CO_2 by the enzyme pyruvate decarboxylase, followed by the reduction of acetaldehyde into ethanol by the enzyme alcohol dehydrogenase. Reoxidation of NADH into NAD^+ in this last conversion step enables the glycolytic break down of carbon sources to remain redox neutral in the absence of oxygen (Fig. 2).

S. cerevisiae is characterized by the presence of a wide array of proteins that can transport hexoses (i.e. glucose, fructose, galactose and mannose) across the plasma membrane. At least 21 proteins have been characterized as hexose transporters (*24*). These transporters have a broad range of affinities for hexoses (K_m ranging from 1 to 100 mM) and a variety of transcriptional responses which enables yeast to grow under a wide range of sugar concentrations (*25, 26*).

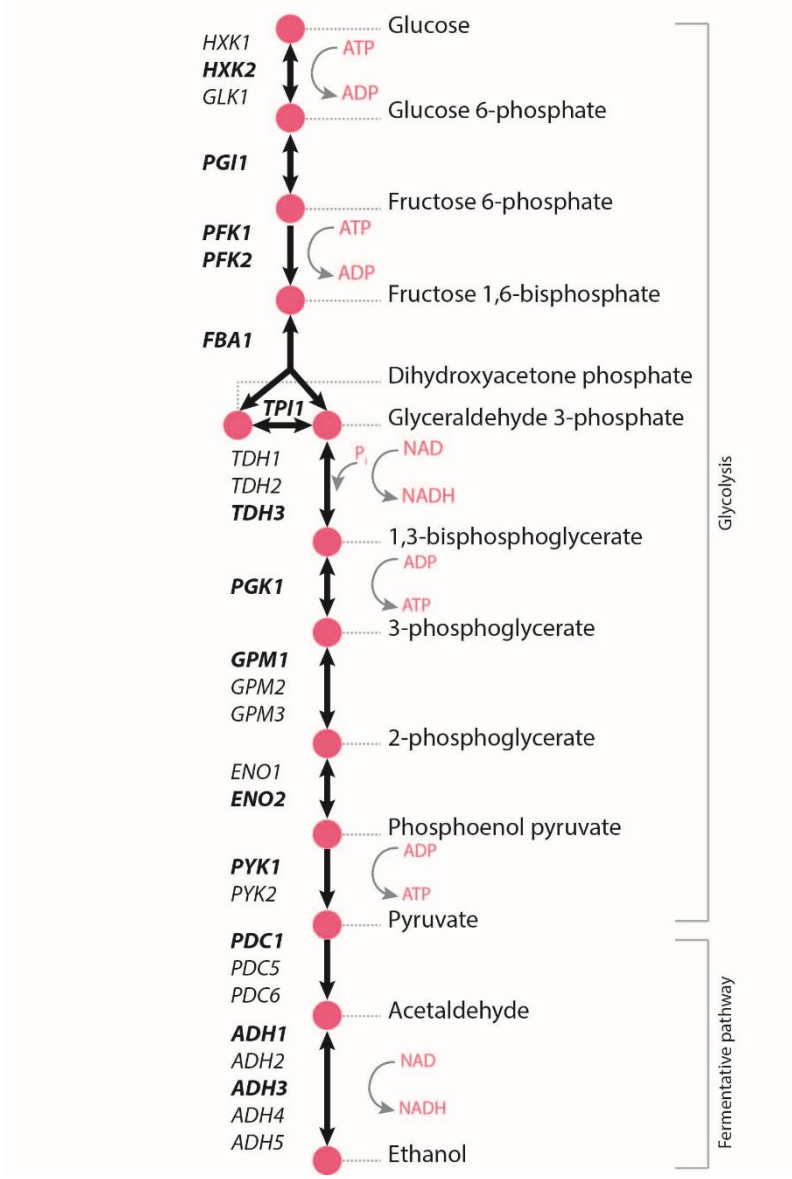


Figure 2 – The glycolytic and fermentative pathway in *S. cerevisiae.*

The paralogs encoding the different isoenzymes catalysing each reaction of the pathways are shown on the left side, the major paralogs are shown in bold. *HXK,* hexokinase; *GLK,* glucokinase; *PGI,* phosphoglucose isomerase; *PFK,* phosphofructokinase; *FBA,* fructose-bisphosphate aldolase; *TPI,* triosephosphate isomerase; *TDH,* glyceraldehyde-3-phosphate dehydrogenase; *PGK,* phosphoglycerate kinase; *GPM,* phosphoglycerate mutase; *ENO,* enolase; *PYK,* pyruvate kinase; *PDC,* pyruvate decarboxylase; *ADH,* alcohol dehydrogenase.

Genetic redundancy

Eukaryotic genomes are characterized by a high degree of genetic redundancy, for example, at least 15% of the genes in the human genome are duplicates (*27*). Metabolic functions are particularly rich in paralogs, a feature reflected in *S. cerevisiae* glycolysis (*28, 29*). The 12 steps of the glycolytic and fermentative pathway of *S. cerevisiae* are catalysed by a set of 26 enzymes. For eight out of the twelve steps, two to seven

paralogous genes are involved (Fig. 2), encoding highly similar isoenzymes with 70% to 99% identity at the protein level. Exceptions are the Gpm and Pfk isoenzymes which only share protein identities of 36%-66%, and glucokinase (Glk1) which only shares 37% protein identity with hexokinase 1 and 2 (Hxk1 and Hxk2). This genetic redundancy originates from two types of events. About 150 million years ago the ancestor of *S. cerevisiae* duplicated its number of chromosomes. Whether this duplication resulted from a duplication from its own genome or from ancient hybridization of two yeast lineages is under debate (*30-32*). Most paralogs in the glycolytic and fermentative pathways originate from this whole Genome Duplication (WGD), with the exception of *GLK1, PFK1, PFK2, TDH1, GPM1, PDC1,5,6* and *ADH2,3,4* which appeared post-WGD, from small scale duplications (*33, 34*).

There are many theories regarding the fate of duplicated genes. It is generally assumed that, after a gene duplication, the two copies will only be retained if they provide a fitness benefit to the cell or if one of the copies undergoes neo-functionalization leading to fulfilment of a new role in the cell (*27*). During evolution, duplication of the glycolytic and fermentative genes has probably provided a selective advantage by increasing the glycolytic flux (*33*). However, not all paralogs contribute equally to the glycolytic or fermentative activity. Based on gene expression and deletion studies, for all reactions besides phosphofructokinase a predominant paralog is the major catalyst during growth on excess glucose, while the other paralogs, considered "minor", hardly contribute to the catalytic activity (*35*) (Fig. 2). The phosphofructokinase activity requires two equally important subunits encoded by *PFK1* and *PFK2* that have similar expression levels and operate as hetero-octamers (*36-38*). Redundant genes can also be retained due to neofunctionalization. This is probably the case for some glycolytic paralogs that have been shown to provide increased fitness under specific conditions. For example *PDC6* is a minor paralog encoding a pyruvate decarboxylase variant that is characterized by a substantially lower sulfur amino acids content than its isoenzymes Pdc1 and Pdc5. Accordingly, *PDC6* expression is strongly upregulated in response to sulfur-limited conditions (*39, 40*).

To obtain more experimental evidence about the function of glycolytic and fermentative paralogs, a few years ago a 'Minimal Glycolysis' strain was constructed from which all minor paralogs were removed resulting in a strain with only 13 glycolytic genes, one for each step with the exception of *PFK1* and *PFK2* (Fig. 2). Challenging the aforementioned evolutionary theories for genetic redundancy, no effect on the physiology of the strain was observed under a wide range of conditions (*35*).

Moonlighting properties of Hxk2, Eno1/2 and Fba1

Several of the glycolytic genes acquired, next to their catalytic function in glycolysis, a secondary molecular function, also referred to as moonlighting function. Within *S. cerevisiae* glycolysis, at least three enzymes are known to have moonlighting properties, hexokinase 2 (Hxk2), the aldolase Fba1 and the 99% identical enolases Eno1 and Eno2 (*41*).

The most investigated and best characterized moonlighting glycolytic enzyme is Hxk2. Next to its glucose phosphorylating role in glycolysis, Hxk2 is involved in a cellular process known as glucose repression or carbon catabolite repression (*42*). Glucose is the preferred carbon source for *S. cerevisiae*, and its presence in excess leads to the transcriptional repression of a broad array of genes, such as those involved in the utilization of alternative carbon sources. For example, the expression of genes involved in galactose metabolism (*GAL* genes) or sucrose metabolism (*SUC2*) is repressed when *S. cerevisiae* is growing in media containing glucose in excess (*43, 44*). *S. cerevisiae* preference for fermentative metabolism when glucose is present in excess, the Crabtree effect, is reflected in the repression of genes involved in respiration under these conditions (*45, 46*). Hxk2 plays an important role in this process. In glucose excess conditions about 15% of Hxk2 localizes to the nucleus where it directly interacts with the transcriptional repressor Mig1 and forms a repressor complex which binds to the promoters of most glucose repressible genes (*47, 48*). This interaction and the nuclear localisation of Hxk2 is dependent on the Hxk2 Lys^{6}-Met^{15} decapeptide (*47*). When glucose is low or absent, Mig1 is phosphorylated by the active form of the protein kinase Snf1, which reduces Mig1 repressing capacity (*49, 50*). During glucose excess conditions, the transcription of *SUC2,* but also the *HXK1* and *GLK1* paralogs is repressed via Hxk2 (*48, 51*). Deletion of *HXK2* leads to a strain displaying in excess glucose conditions a Crabtree-negative phenotype with fully respiratory glucose dissimilation, reflected in a high biomass yield (*52*).

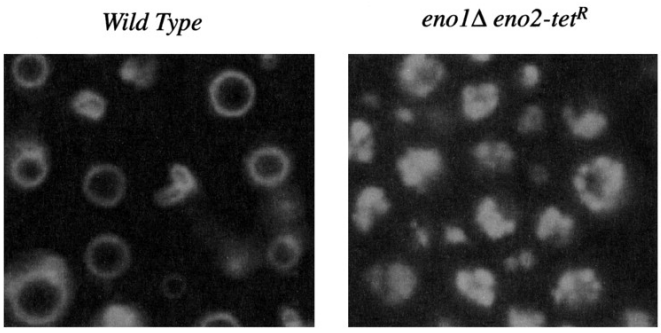


Figure 3 – Vacuoles from wild type and *eno1Δeno2-tet*^{R} strains.
Cells were incubated for 12h in YPD medium without tetracycline and visualized using FM4-64. Deletion of *ENO1* and repressed expression of *ENO2* led to a fragmented vacuoles phenotype (right panel). Figure from Decker *et al. 2006 (53).*

Eno1 and Eno2 are both involved in vacuole fusion by enabling specific protein trafficking to the vacuole (*53*). Vacuoles are important for several processes in the cell such as ion and pH homeostasis, protein turnover and as storage compartment for ions (*54*). *ENO1* deletion combined with diminished expression of *ENO2* leads to a phenotype with fragmented vacuoles (*53*). Decker *et al*. observed that activity of a single isoenzyme causes a milder phenotype in which only 35%-40% of the cells showed vacuole fragmentation. This result however contradicts the phenotype obtained for the Minimal Glycolysis strain, in which the expression of *ENO2* only does not affect vacuolar structures (*35*). These conflicting observations might result from differences in strain background or experimental conditions between the two studies. Next to its vacuolar function, yeast enolase has a second moonlighting function and is reported to be involved in the import of $tRNA^{Lys}$(CUU) (called tRK1) into mitochondria (*55*). In general, all tRNAs required in mitochondria are directly synthetized in mitochondria. However, in growth conditions above 37°C, cytosolic tRK1 translocation to the mitochondria is required (*55*). Yeast enolases bind to tRK1 and this complex then moves to the surface of mitochondria where the tRNA is transferred to the precursor lysyl-tRNA synthetase. Subsequently, this complex is imported in the mitochondria via the TOM and TIM complexes (*56*).

Fba1 plays a role in the association of the subunits of the highly conserved vacuolar proton- translocating ATPases (V-ATPases). In yeast, V-ATPases are mainly found in the vacuolar membrane and they couple ATP hydrolysis to proton transport out of the cytosol into the vacuole (*57*). Fba1 physically interacts with the V-ATPase, an essential interaction for the assembly of the subunits and the activity of the complex (*58*). Assembly of the V-ATPase subunits is strongly dependent on the presence of glucose, suggesting a glucose-dependent regulation of the acidification of intracellular

compartments (*59, 60*). Inactivity of the V-ATPase leads to yeast inability to grow in media buffered at alkaline pH (*61*). Interestingly, this moonlighting property is conserved between yeast and mammalian cells, even though mammals and yeast have radically different types of fructose bisphosphate aldolase that do not share homology (class I in mammalian cells and class II in yeast) (*58, 62, 63*).

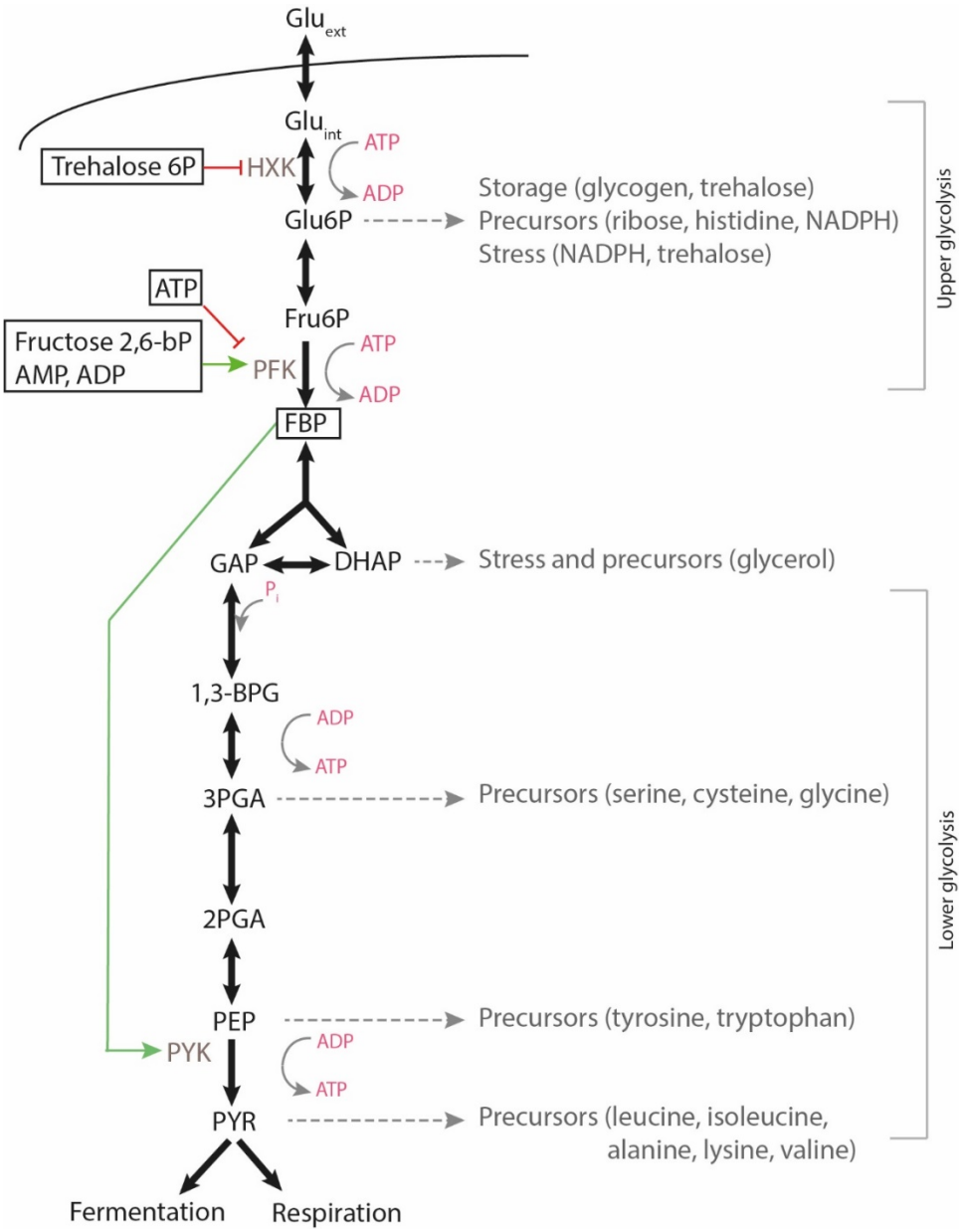


Figure 4 – Regulation of the glycolytic pathway in *S. cerevisiae.*

The three main points of allosteric regulation in *S. cerevisiae* are the hexokinase (HXK), phosphofructokinase (PFK) and pyruvate kinase (PYK) reactions. Green arrows indicate activation, red lines inhibition and black arrows enzymatic reactions. Trehalose 6-phosphate is produced from glucose 6-phosphate by the trehalose 6-phosphate synthase (encoded by *TPS1*). As indicated, the glycolytic pathway is important for supply of precursors, redox equivalents and stress related metabolites, energy conservation and storage of carbohydrates. Glu_{ext}, glucose extracellular; Glu_{int}, glucose intracellular; Glu6P, glucose 6-phosphate; Fru6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; 1,3-BPG, 1,3-bisphosphoglycerate; 3PGA, 3-phosphoglycerate; 2PGA, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate. ADP, adenosine diphosphate; ATP, adenosine triphosphate; AMP, adenosine monophosphate. Adapted from (*18*).

Regulation of glycolysis

The glycolytic pathway is one of the highest expressed pathways in the cell, which is reflected by the high concentration of glycolytic proteins which can reach 20% of the total amount of soluble protein in the cell (*64*). The capacity of the glycolytic enzymes estimated from *in vitro* assays, with the notable exception of phosphofructokinase, exceeds the *in vivo* glycolytic flux, depicting an overcapacity (*65-67*). This overcapacity confers metabolic flexibility, allowing for fast adaption to changes in environmental conditions (*66, 68*). To optimize performance in response to its environment, *S. cerevisiae* regulates the individual glycolytic steps and the overall glycolytic flux via multi-layered responses (*69, 70*).

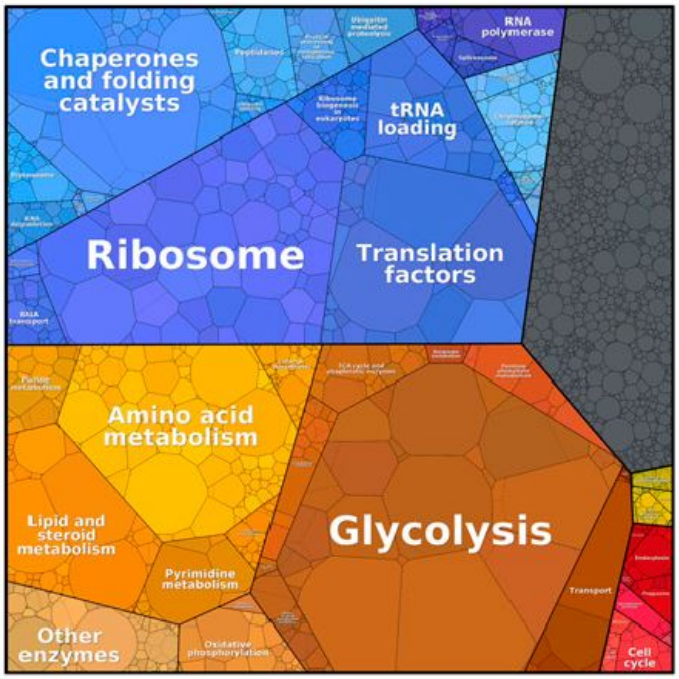


Figure 5 – Proteomap of the *S. cerevisiae* proteome.
The proteome of yeast cells cultivated in YPD medium was measured using mass spectrometry (*71*). Each shape represents the mass fraction of a protein within the proteome. Different colours represent proteins belonging to different cellular functions. Proteins that do not map to any category are shown in grey. Figure from Liebermeister *et al.* 2014 (*64*).

At the transcriptional level, the major glycolytic and fermentative paralogs of *S. cerevisiae* have high basal expression levels and are considered constitutively expressed, although their expression can be condition-dependent with variations up to five-folds (*35, 72*). These properties make glycolytic promoters very popular for heterologous gene expression. The expression levels of the different glycolytic and fermentative major paralogs are not homogeneous and can differ by up to an order of magnitude, with *TDH3* and *ADH1* being amongst the most highly expressed and *PFK1*,

PFK2 and *HXK2*, encoding kinases, being the least expressed genes (*35*). Most glycolytic and fermentative genes are activated via the specific Gcr1/Gcr2 and general Rap1 transcription factors (*73*). The minor paralogs generally display lower transcript levels than their major paralogs and their expression can be condition-dependent, as discussed above for *PDC6*. *GLK1, HXK1, PYK2* and *ADH2* are four minor paralogs that are particularly sensitive to glucose availability, as their transcription is strongly repressed in media with excess glucose (*51, 74-76*). Transcription is not considered as a key step in the regulation of the glycolytic pathway, as revealed by the lack of correlation between changes in transcript levels and in glycolytic flux under a range of conditions (*69, 77, 78*) and the failure to increase the glycolytic flux by overexpression of glycolytic genes (*79-81*). Regulation of the glycolytic flux is therefore predominantly regulated by post-transcriptional mechanisms.

The activity of enzymes *in vivo* can be modulated by a broad range of post-translational modifications, such as phosphorylation, methylation, acetylation, etc. These modifications can alter the enzyme structure, playing as an on/off switch and tuning the amount of active enzyme or more subtly by altering kinetic properties. Post-translational modifications can also tune the total amount of enzyme by affecting the balance between protein synthesis and degradation (*70*). Phosphorylation is one of the most frequent reversible post-translational modifications in yeast and about two third of the metabolic enzymes, including several glycolytic enzymes, are targets of the kinase and phosphatase signalling network (*82*). While it is long known that hexokinase 1 and hexokinase 2 are phosphoproteins, whether phosphorylation affects the glycolytic function of these proteins remains unknown (*83*). It has however been characterized that phosphorylation of Hxk2 affects its nuclear localization and activity as transcriptional regulator (*47*). Next to phosphorylation, most of the glycolytic and fermentative enzymes are ubiquitinated and acetylated (*70, 84, 85*). However, while several large scale studies have enabled to map all these post-translational modifications, their functional relevance has not been identified yet and their role in regulating the glycolytic flux remains to be elucidated.

The best characterized level of regulation of the glycolytic flux is the regulation of the enzyme's catalytic activity *in vivo* by low molecular weight compounds (i.e. metabolic regulation). Next to the expected regulation by substrate and product concentrations, several key glycolytic enzymes are activated or inhibited by allosteric effectors. This is the case for the three phosphorylation steps and more particularly Hxk1, Hxk2, Pfk1, Pfk2 and Pyk1 that are considered to be the key regulation points for fine-tuning the

glycolytic flux (*70*) (Fig. 4). Hxk1 and Hxk2 are both inhibited by trehalose-6-phosphate, product of the side reaction catalysed by the trehalose-6-phosphate synthase using glucose-6-phosphate as substrate (*86*). Pfk1 and Pfk2 are sensitive to a broad range of metabolites, but the major regulations are inhibition by ATP and activation by fructose-2,6-bisphosphate (synthetized from fructose-6-phosphate by the 6-phosphofructo-2-kinase) (*87*). Pyk1 is the target of a feed-forward activation by the glycolytic intermediate fructose-1,6-bisphosphate, product of the phosphofructokinase (*76*).

While studies by prestigious biochemists have very convincingly demonstrated in pioneering studies these regulatory mechanisms *in vitro*, up to date mutant studies on these individual key glycolytic steps have failed to demonstrate a role for these regulations *in vivo*, during steady-state growth on glucose media. More recent studies suggest that metabolic regulations are important during transitioning, and more particularly between conditions that lead to imbalance between top and bottom glycolysis (*88, 89*). While kinetic models are available to predict the impact of these allosteric regulations on the glycolytic flux, remarkably there are to date no experimental studies that explore the synergetic impact of simultaneous modifications of these regulations.

As the glycolytic enzymes are present at overcapacity and simultaneous overexpression of the glycolytic enzymes does not lead to increased glycolytic flux, the overall glycolytic flux is likely to be regulated by factors outside glycolysis. Van Heerden *et al.* (*18*) proposed that glycolytic flux at excess glucose conditions mainly depends on demand of ATP. Surplus of ATP would lead to inhibition of glycolysis and therewith ATP production. As discussed above, the main point of inhibition of glycolysis by ATP is PFK. Next to ATP demand, also glucose transport plays a role in tuning the glycolytic flux under conditions where transport is limited relative to the rest of the system such as during growth in chemostat or in strains with reduced glucose transport capacity (*18, 90*).

Beyond yeast: diversity in allosteric regulation

In the majority of organisms which have been studied, ranging from bacteria to human, regulation of the kinase reactions (e.g. HXK, PFK and PYK) plays an important role in glycolytic flux control (*91-93*), which makes sense in view of the supply and demand theory discussed above. Although the glycolytic enzymes have been strongly conserved during evolution, the mechanisms involved in their metabolic regulation are variable and are most likely dependent on the (cellular) environment of the organism (*94*). This is well illustrated by the diversity in metabolic regulation of hexokinase across different organisms (*92*). Regulating the flux between the top (ATP consumption) and the bottom

(ATP production) of glycolysis is important to avoid metabolic imbalance, particularly when the cells are exposed to sudden changes in glucose supply (*88, 89*). Organisms have evolved several mechanisms to avoid this imbalance, some of which involve the metabolic regulation of hexokinase. As mentioned above, yeast Hxk2, the main glucose phosphorylating enzyme during growth on excess glucose, is inhibited by trehalose-6P (*86*). In human and other mammalian cells several other tissue-dependent mechanisms are present: the muscle type hexokinase (HK2) is allosterically inhibited by its product, glucose-6P. On the other hand, the human HK4 enzyme (glucokinase) which is the predominant form in liver and pancreas, is not regulated by any effector molecules. The expression of *HK4* in pancreatic cells is however so low that feedback inhibition is not required to prevent the cells from reaching a state of glycolytic imbalance. In the liver, an extra layer of regulation is present, since inhibition is mediated by the glucokinase regulatory protein (GKRP) (*95*). In many bacteria, as for example lactic acid bacteria and *Escherichia coli,* the situation is different again, since glucose and other sugars are taken up by a PhosphoTransfer System (PTS) which couples glucose uptake to its phosphorylation, and therefore shortcuts the need for a hexokinase step (*96*). Finally, an interesting mechanism is present in *Trypanosoma brucei,* a parasite that causes the African sleeping Disease. Its hexokinase seems to be unregulated, but instead, the first part of glycolysis, till 3-phosphoglycerate is compartmentalized in specialized organelles called glycosomes (*97*), insulating the top of glycolysis to variations in sugar supply and from the bottom of glycolysis (*98, 99*).

PFK enzymes are sensitive to a wide range of effectors and have a very complex regulation. In most eukaryotic cells (animals, plants, yeasts) the metabolite fructose-2,6-bisphosphate (F-2,6-bP) plays a central role (*94*). While F-2,6-bP is a potent regulatory molecule in eukaryotes, it is absent from prokaryotes (*100*). In bacteria, phosphoenolpyruvate is the main inhibitor of PFK (*101*).

In most characterized eukaryotes, pyruvate kinase is activated by fructose-1,6-bisphosphate (*94, 102*). However, in most cases isoforms insensitive to fructose-1,6-bisphosphate are also present, such as the human muscle pyruvate kinase variant (PKM1) and yeast pyruvate kinase 2 (*76, 103*). In bacteria two classes of pyruvate kinases have been identified (*94*). Class I enzymes are dominant under fermentative conditions and are activated by fructose-1,6-bisP and inhibited by ATP. Class II pyruvate kinases play an important role in gluconeogenic conditions and are activated by AMP and by various sugar monophosphates as glucose-6P and ribose-5P (*94*). Most characterized bacteria harbour a single type of pyruvate kinase, however some bacteria

like *E. coli,* contain both types (*94, 104*). In bacteria pyruvate kinase activity is also coupled to carbohydrate uptake by a phosphotransferase system (PTS) for which the energy is provided by phosphoenolpyruvate (PEP) (*105*). Pyruvate kinases of parasitic protozoans such as trypanosomes are unique in that they are activated by fructose-2,6-bisP (*106, 107*).

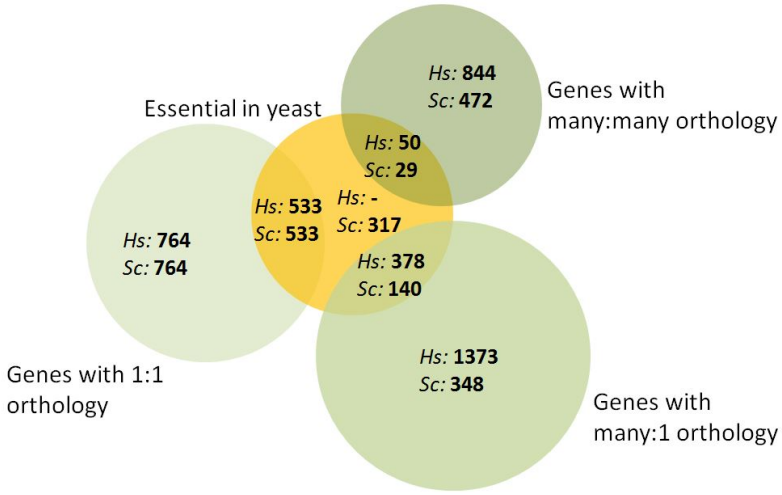


Figure 6 – Orthologs shared between the genomes of *S. cerevisiae* and *Homo sapiens.*
The green circles of the Venn diagram show the nature of the orthology (humans:yeast; 1:1, many:1 or many:many) and the number of orthologs in each group. The yellow circle reflects the genes which are essential in yeast under standard laboratory conditions with glucose as carbon source (*108*). Adapted from (*109*).

Yeast as model organism

Next to its applications in biotechnology, *S. cerevisiae* is intensively used as model for higher eukaryotes. *S. cerevisiae's* popularity as eukaryotic model can be largely explained by a high degree of conservation in pathways and cellular processes between human and yeast, while these two organisms are phylogenetically very distant (*110*). Indeed, the 12 Mb yeast genome, which contains roughly 6,000 protein-encoding genes, and the 3.3 Gb human genome containing roughly 20,000 protein-encoding genes share approximately 2,100 groups of orthologs involving 2,900 yeast genes and 3,900 human genes (*109*). Many of these genes are involved in essential cellular processes (Fig. 6). This high degree of conservation is not the only factor that propelled *S. cerevisiae* as favourite metazoan model. It is also very tractable, grows fast, has simple nutritional requirements, is highly genetically accessible largely due to the predominance of homologous recombination as mechanism for double strand DNA break repair and has a lower degree of genetic complexity than higher eukaryotes. One of the best known

examples of how *S. cerevisiae* has contributed as a model organism, is the elucidation of a large part of the molecular machinery underlying the cell cycle, which is, as was later discovered, very conserved from yeast to humans (*111, 112*). Also, the metabolic reactions of the glycolytic pathway were for a large part elucidated in yeast in the first part of the 20th century (*113*). In 1996 *S. cerevisiae* was the first eukaryote with a fully sequenced genome (*114*).

S. cerevisiae most certainly possesses the best furbished molecular toolbox and is at the forefront of CRISPR/Cas-based techniques developments (*115*). Particularly, yeast knockout collections have been available for many years and have been used in a wide array of screens with the aim to increase understanding of biological functions, response to stress, and mechanisms of drug action (*108, 116*). Because of the relatively high homology between human and yeast and the fact that about 30% of the human genes known to be involved in diseases have a homolog in yeast (*117*), yeast cell-based assays have led to the identification of multiple mechanisms contributing to a better understanding of human diseases and discovery of new targets for drug development (*118, 119*). For instance, screening of a yeast deletion library for reduced fitness on non-fermentable carbon sources led to the identification of genes involved in (human) mitochondrial disorders (*120*).

To improve yeast as a model organism and to learn more about the degree of conservation of cellular processes in human and yeast, there is a growing interest in the 'humanization' of yeast. A rapidly increasing number of studies is dedicated to systematically testing the ability of single human genes to replace (i.e. complement) their yeast orthologs (*121-128*). The fact that the *Saccharomyces* genome database (SGD) started a section with information about gene complementation also underlines the growing interest in this field (*129*). The largest study to date investigated the functional replaceability of 424 essential yeast genes by their human ortholog (*122*). Remarkably, complementation was successful for close to half of the genes. These results revealed that in most cases the degree of protein conservation was not a good predictor for complementation, while proteins from the same pathway or complex, even with low conservation levels, were similarly replaceable (*122*). This study suggested that humanizing entire pathways or cellular processes should be feasible, thereby improving the yeast model since proteins could be studied in conditions closer to their native context.

To date, the examples of large scale humanization of yeast are scarce, which can mainly be attributed to several technical challenges that have to be overcome. To replace an

entire pathway, multiple genes spread over the different yeast chromosomes have to be targeted, which involves complex genome editing approaches and/or multiple rounds of transformations. The high genetic redundancy of the yeast and human genome makes this task even more challenging. To date, a single study reports the humanization of an entire metabolic pathway. Very recently, Agmon *et al.* fully humanized *S. cerevisiae* adenine *de novo* biosynthesis pathway (*130*). They showed that six out of seven human genes could fully complement and that one, *PPAT* (the human equivalent of *ADE4*) could only partially. Differences in *PPAT* regulatory properties between the human and yeast had a strong impact on the growth rate of the humanized yeast strain. As also illustrated in chapter 5 of this thesis, differences in human and yeast cellular environment affecting the *in vivo* activity of enzymes might be one of the largest challenges in yeast humanization projects. Next to this extensive study, there are few other noteworthy examples of large scale humanization project in *S. cerevisiae* such as the partial humanization of yeast core nucleosomes (*131*) and the humanization of N-glycosylation pathways in both in *Pichia pastoris* and *S. cerevisiae* with the aim to optimize yeast as host for the production therapeutic glycoproteins (*132*).

Expanding the molecular toolbox of *Saccharomyces cerevisiae*

To be able to change existing functionalities and to add new ones, both fundamental and applied research rely on extensive remodelling of microbial genomes. During the past decades the possibilities to do this have tremendously increased, but several aspects remain challenging. An ideal futuristic scenario consists of *in silico* design, *in vitro* synthesis and transfer to 'start-up' cell machinery of 'plug-and-play' synthetic, designer genomes. A few examples of synthetic genomes already exist with the complete synthesis and assembly of the *Mycoplasma genitalium* genome (*133*) and more recently the large Sc2.0 project consortium aiming at redesigning and synthesising the entire yeast genome (*134*). Utilization of *S. cerevisiae* has been pivotal for the assembly of *M. genitalium* genome. Where *in vitro* stitching and assembly in *E. coli* failed, the remarkably efficient and faithful homologous recombination (HR) machinery of *S. cerevisiae* succeeded in assembling four very large DNA fragments into the full, 583 kb bacterial genome (*133*). It is therefore not surprizing that *S. cerevisiae* was chosen as host for the first, complete synthesis of a eukaryotic genome. This trait also makes *S. cerevisiae* an attractive host for the expression of large pathways, as exemplified by the biosynthesis of opioids, which involved the heterologous expression of a large number of genes from plants, bacteria and mammals (*135*). The decreasing costs of DNA synthesis have contributed to the increased possibilities to synthetize and implement entire codon-optimized heterologous pathways or even entire

chromosomes in *S. cerevisiae.* In addition, advancements in whole genome sequencing techniques have led to an increasing availability of genome sequences from a wide range of species, which serve as a source of diversity for addition of new functionalities to yeast.

Despite great advances in DNA assembly, large-scale remodelling of existing pathways and cellular processes remains challenging, even in the highly genetically accessible *S. cerevisiae*. This has two main reasons. Firstly, the yeast genome is characterized by a high genetic redundancy, and a small number of reactions can require the modification of dozens of genes. Secondly, unlike prokaryotic genomes, eukaryotic genomes are fragmented and genes belonging to the same pathway or cellular process are therefore scattered over *S. cerevisiae's* 16 different chromosomes. This genetic redundancy and scattered localization therefore renders replacing or modifying entire pathways or processes a daunting task. The discovery and development of CRISPR genome editing tools have strongly increased the speed and efficiency of genome editing and progress with multiplexing to target multiple sites in one transformation will in the future even further improve this (*115*). In chapter 2 of this thesis, HR and CRISPR are combined for the development of a new modular approach to remodel essential metabolic pathways in yeast.

In contrast to the simpler organisation of expression units in operons in bacteria, gene expression in yeast, as in all eukaryotes, requires the presence of a promoter directly upstream each open reading frame (ORF). Whereas for ORFs, functionality in yeast can be explored for a large pool of heterologous genes across kingdoms, for promoters the options are much more limited because they have to be compatible with the host transcription machinery. This strongly limits the range of usable promoters, even in *S. cerevisiae* which has one of the largest molecular toolboxes available. There is especially a need for strong and constitutive promoters which are preferred for metabolic engineering, and tuneable promoters which are active in a condition-dependent manner, as for example galactose inducible *GAL1* promoter (*136*). A lot of effort is invested in the development of synthetic promoters, not only to expand the number of promoters which can be used in *S. cerevisiae*, but also to create shorter variants to reduce the amount of DNA needed for expression of a gene (*137-140*). An alternative approach is to use heterologous promoter sequences of related yeast species for gene expression in *S. cerevisiae* (*141, 142*). Chapter 3 of this thesis explores the potential of using yeast species from the *Saccharomyces* genus as source of promoters for *S. cerevisiae*.

Scope of this thesis

This PhD project was part of a larger research project which was funded by the ERC consolidator grant AdLibYeast awarded to Prof. Pascale Daran-Lapujade. Central carbon metabolism (CCM) is a set of reactions involved in the production of all industrially-relevant biochemicals, via supply of precursors, energy-rich moieties and redox equivalents. The construction of microbial cell factories with optimal productivity, product yield and robustness requires the ability to remodel and fine tune CCM. Large-scale modifications of CCM are time consuming and laborious since CCM involves many biochemical reactions, is characterized by a high level of genetic redundancy and therefore encompasses many genes that are scattered all over the genome. To tackle this limited genetic accessibility, the overarching goal of AdLibYeast was to create a platform for *ad libitum* remodelling of CCM in *Saccharomyces cerevisiae*. The strategy employed to reach this goal consists in removing all redundant genes from the pathways involved in central carbon metabolism and to re-localize all the remaining genes to a single location in the genome. In such a microbial platform, new versions of the CCM can be implemented while simply removing the minimalized and relocalized, native version. This approach would therefore enable researchers to build and tune CCM at will, either for industrial purposes or to obtain a better understanding of regulatory mechanisms within the CCM.

In view of this large task, this project was divided over several research lines:

1. Development of new molecular tools in *S. cerevisiae* to make large scale modifications of the genome easier and more efficient.
2. Minimizing and relocalizing the whole CCM, starting with glycolysis as proof of concept.
3. Evaluation of synthetic, supernumerary chromosomes in *S. cerevisiae* as modular platforms to express and remodel central carbon metabolism and production pathways of interest.

The work presented in this thesis falls within the first two research lines.

Genetic redundancy is a major obstacle that has to be overcome to improve the genetic accessibility of eukaryotic cell factories. Before the start of this project, Solis-Escalante *et al.* showed that the set of genes encoding the glycolytic enzymes could be minimized from 26 to 13 without affecting yeast physiology (*35*). **Chapter 2** of this thesis shows how all these glycolytic genes could be re-localized to one single chromosomal locus, without major impact on the physiology of the strain. The concept of pathway swapping was explored by remodelling the entire glycolytic pathway of *S. cerevisiae* by

heterologous versions in two simple steps. This concept and the 'Switchable yeast glycolysis' strain (SwYG) are the fundament of the research described in this thesis.

One of the current bottlenecks in yeast strain construction is the limited availability of well- characterized, strong and constitutive promoters. In view of the ambitious strain construction plans in this project, **Chapter 3** aims at enriching *S. cerevisiae* molecular toolbox with such promoters by exploring biodiversity. This chapter evaluates the possibility to use glycolytic promoters from *Saccharomyces kudriavzevii* and *Saccharomyces eubayanus*, two yeast species closely related to *S. cerevisiae*, to drive strong and constitutive gene expression in *S. cerevisiae*. As virtually nothing was known about the glycolytic pathways of these *S. cerevisiae* relatives, the architecture and expression of the glycolytic genes were first analysed in their native context and compared to that of *S. cerevisiae*. Furthermore, the strength and context-dependency of *Saccharomyces kudriavzevii* and *Saccharomyces eubayanus* glycolytic promoters were tested in *S. cerevisiae* with fluorescent reporters.

Chapter 4 aims to add a thoroughly validated DNA and RNA watermarking strategy to *S. cerevisiae* molecular toolbox. A well-designed watermarking (i.e. silent base permutations) can be a powerful approach to discriminate between native and synthetic copies of yeast genes and transcripts. However, despite many studies, the impact of introducing synonymous nucleotide substitutions on gene expression and physiology remains poorly understood. In collaboration with colleagues from the Faculty of Electrical Engineering, Mathematics & Computer Science, a watermarking strategy, aiming at minimizing nucleotide substitutions while enabling DNA and RNA discrimination was designed. The pathway swapping strategy was used to construct a set of strains with native and watermarked glycolytic genes. These strains were used to test the ability of watermarks to discriminate between watermarked and native copies of the gene and transcript, and to test their impact on yeast physiology, using tightly controlled bioreactors.

In **Chapter 5**, brought to its full potential, the pathway swapping concept is used to fully humanize the glycolytic pathway in *S. cerevisiae*. The humanized yeast strains enable to address fundamental principles regarding pathway design and regulation, and to explore the potential of yeast with humanized pathways to serve as model for mammals. Despite several major, large-scale efforts, only 8 out of the 23 human glycolytic genes have been tested for complementation in *S. cerevisiae*. This knowledge gap is largely explained by the high genetic redundancy encountered in eukaryotes and the resulting technical challenge. In **Chapter 5**, the minimal glycolysis yeast strain

constructed by Solis-Escalante (*35*), in which the set of glycolytic genes has been minimalized, is used to test the functionality of 25 human glycolytic genes and splicing variants in *S. cerevisiae*. Using the SwYG strain, yeast glycolysis was swapped with the set of 10 genes described as highly expressed in human skeletal muscle, leading to a fully humanized glycolysis. The impact of pathway humanization on yeast physiology was investigated in depth, using bioreactors. To identify causes of slow growth, a combination of single gene complementation, full pathway humanization and laboratory evolution was used. To further evaluate the potential of humanized yeast strains as human model, in a collaboration with the University of Groningen, the kinetic properties of human glycolytic proteins expressed in their native (human skeletal muscle cells) and yeast context were compared.

References

1. R. Bud, *The uses of life: a history of biotechnology*. (Cambridge University Press, 1994).
2. P. E. McGovern *et al.*, Fermented beverages of pre-and proto-historic China. *Proceedings of the National Academy of Sciences* **101**, 17593-17598 (2004).
3. L. Pasteur, Mémoire sur la fermentation alcoolique. *Ann Chim Phys* **58**, 323-426 (1860).
4. L. Pasteur, Mémoire sur la fermentation alcoolique. *Compt Rend* **45**, 1032-1036 (1857).
5. P. J. Chambers, I. S. Pretorius, Fermenting knowledge: the history of winemaking, science and yeast research. *EMBO reports* **11**, 914-920 (2010).
6. I. Borodina, J. Nielsen, Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnology journal* **9**, 609-620 (2014).
7. J. Nielsen, M. C. Jewett, Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **8**, 122-131 (2008).
8. J. Nielsen, C. Larsson, A. van Maris, J. Pronk, Metabolic engineering of yeast for production of fuels and chemicals. *Curr. Opin. Biotechnol.* **24**, 398-404 (2013).
9. D.-K. Ro *et al.*, Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).
10. R. De Deken, The Crabtree effect: a regulatory system in yeast. *Microbiology* **44**, 149-156 (1966).
11. J. Monod, The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371-394 (1949).
12. J. Piškur, E. Rozpędowska, S. Polakova, A. Merico, C. Compagno, How did *Saccharomyces* evolve to become a good brewer? *Trends Genet.* **22**, 183-186 (2006).
13. S. Naseeb *et al.*, *Saccharomyces jurei* sp. nov., Isolation and genetic identification of a novel yeast species from *Quercus robur*. *Int. J. Syst. Evol. Microbiol.* **67**, 2046-2052 (2017).
14. C. T. Hittinger, *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* **29**, 309-317 (2013).
15. Z. Salvadó *et al.*, Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Appl. Environ. Microbiol.* **77**, 2292-2302 (2011).
16. D. Sicard, J.-L. Legras, Bread, beer and wine: yeast domestication in the *Saccharomyces* sensu stricto complex. *C. R. Biol.* **334**, 229-236 (2011).
17. S. Naseeb *et al.*, Whole genome sequencing, de novo assembly and phenotypic profiling for the new budding yeast species *Saccharomyces jurei*. *G3: Genes, Genomes, Genetics* **8**, 2967-2977 (2018).
18. J. H. Van Heerden, F. J. Bruggeman, B. Teusink, Multi-tasking of biosynthetic and energetic functions of glycolysis explained by supply and demand logic. *Bioessays* **37**, 34-45 (2015).
19. E. Nevoigt, U. Stahl, Osmoregulation and glycerol metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **21**, 231-241 (1997).
20. A. D. Elbein, Y. Pan, I. Pastuszak, D. Carroll, New insights on trehalose: a multifunctional molecule. *Glycobiology* **13**, 17R-27R (2003).

21. J. Francois, J. L. Parrou, Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **25**, 125-145 (2001).
22. A. Romano, T. Conway, Evolution of carbohydrate metabolic pathways. *Res. Microbiol.* **147**, 448-455 (1996).
23. A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, R. Milo, Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences* **110**, 10039-10044 (2013).
24. R. Wieczorke *et al.*, Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.* **464**, 123-128 (1999).
25. S. Özcan, M. Johnston, Function and regulation of yeast hexose transporters. *Microbiol. Mol. Biol. Rev.* **63**, 554-569 (1999).
26. E. Reifenberger, E. Boles, M. Ciriacy, Kinetic characterization of individual hexose transporters of *Saccharomyces cerevisiae* and their relation to the triggering mechanisms of glucose repression. *Eur. J. Biochem.* **245**, 324-333 (1997).
27. V. E. Prince, F. B. Pickett, Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics* **3**, 827-837 (2002).
28. M. Kellis, B. W. Birren, E. S. Lander, Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617-624 (2004).
29. L. Kuepfer, U. Sauer, L. M. Blank, Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* **15**, 1421-1430 (2005).
30. S. Ohno, Evolution by gene duplication. *Springer New York*, (1970).
31. K. H. Wolfe, D. C. Shields, Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708-713 (1997).
32. M. Marcet-Houben, T. Gabaldón, Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* **13**, (2015).
33. G. C. Conant, K. H. Wolfe, Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol. Syst. Biol.* **3**, 129 (2007).
34. J. L. Gordon, K. P. Byrne, K. H. Wolfe, Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet.* **5**, (2009).
35. D. Solis-Escalante *et al.*, A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell* **14**, 804-816 (2015).
36. A. Arvanitidis, J. J. Heinisch, Studies on the function of yeast phosphofructokinase subunits by *in vitro* mutagenesis. *J. Biol. Chem.* **269**, 8911-8918 (1994).
37. J. Heinisch, K. Vogelsang, C. P. Hollenberg, Transcriptional control of yeast phosphofructokinase gene expression. *FEBS Lett.* **289**, 77-82 (1991).
38. G. Kopperschläger, J. Bär, K. Nissler, E. Hofmann, Physicochemical parameters and subunit composition of yeast phosphofructokinase. *Eur. J. Biochem.* **81**, 317-325 (1977).
39. M. Fauchon *et al.*, Sulfur sparing in the yeast proteome in response to sulfur demand. *Mol. Cell* **9**, 713-723 (2002).
40. V. M. Boer, J. H. De Winde, J. T. Pronk, M. D. Piper, The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J. Biol. Chem.* **278**, 3265-3274 (2003).
41. C. Gancedo, C. L. Flores, Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.* **72**, 197-210 (2008).
42. J. M. Gancedo, Yeast carbon catabolite repression. *Microbiol. Mol. Biol. Rev.* **62**, 334-361 (1998).
43. D. Lohr, P. Venkov, J. Zlatanova, Transcriptional regulation in the yeast *GAL* gene family: a complex genetic network. *The FASEB Journal* **9**, 777-787 (1995).
44. S. Özcan, L. G. Vallier, J. S. Flick, M. Carlson, M. Johnston, Expression of the *SUC2* gene of *Saccharomyces cerevisiae* is induced by low levels of glucose. *Yeast* **13**, 127-137 (1997).
45. F. Rolland, J. Winderickx, J. M. Thevelein, Glucose-sensing and-signalling mechanisms in yeast. *FEMS Yeast Res.* **2**, 183-201 (2002).
46. C. J. Klein, L. Olsson, J. Nielsen, Glucose control in *Saccharomyces cerevisiae*: the role of Mig1 in metabolic functions. *Microbiology* **144**, 13-24 (1998).
47. D. Ahuatzi, P. Herrero, T. De La Cera, F. Moreno, The glucose-regulated nuclear localization of hexokinase 2 in *Saccharomyces cerevisiae* is Mig1-dependent. *J. Biol. Chem.* **279**, 14440-14446 (2004).

48. P. Herrero, C. Martínez-Campa, F. Moreno, The hexokinase 2 protein participates in regulatory DNA-protein complexes necessary for glucose repression of the *SUC2* gene in *Saccharomyces cerevisiae*. *FEBS Lett.* **434**, 71-76 (1998).
49. F. C. Smith, S. P. Davies, W. A. Wilson, D. Carling, D. G. Hardie, The SNF1 kinase complex from *Saccharomyces cerevisiae* phosphorylates the transcriptional repressor protein Mig1p in vitro at four sites within or near regulatory domain 1. *FEBS Lett.* **453**, 219-223 (1999).
50. M. A. Treitel, S. Kuchin, M. Carlson, Snf1 protein kinase regulates phosphorylation of the Mig1 repressor in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18**, 6273-6280 (1998).
51. A. Rodriguez, P. Herrero, F. Moreno, The hexokinase 2 protein regulates the expression of the *GLK1*, *HXK1* and *HXK2* genes of *Saccharomyces cerevisiae*. *Biochem. J.* **355**, 625-631 (2001).
52. J. A. Diderich, L. M. Raamsdonk, A. L. Kruckeberg, J. A. Berden, K. Van Dam, Physiological properties of *Saccharomyces cerevisiae* from which hexokinase II has been deleted. *Appl. Environ. Microbiol.* **67**, 1587-1593 (2001).
53. B. L. Decker, W. T. Wickner, Enolase activates homotypic vacuole fusion and protein transport to the vacuole in yeast. *J. Biol. Chem.* **281**, 14523-14528 (2006).
54. M. Thumm, Structure and function of the yeast vacuole and its role in autophagy. *Microsc. Res. Tech.* **51**, 563-572 (2000).
55. N. Entelis *et al.*, A glycolytic enzyme, enolase, is recruited as a cofactor of tRNA targeting toward mitochondria in Saccharomyces cerevisiae. *Genes Dev.* **20**, 1609-1620 (2006).
56. A. Schneider, Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annu Rev Biochem* **80**, 1033-1053 (2011).
57. P. M. Kane, The where, when, and how of organelle acidification by the yeast vacuolar H+-ATPase. *Microbiol. Mol. Biol. Rev.* **70**, 177-191 (2006).
58. M. Lu, D. Ammar, H. Ives, F. Albrecht, S. L. Gluck, Physical interaction between aldolase and vacuolar H+-ATPase is essential for the assembly and activity of the proton pump. *J. Biol. Chem.* **282**, 24495-24503 (2007).
59. Y. Y. Sautin, M. Lu, A. Gaugler, L. Zhang, S. L. Gluck, Phosphatidylinositol 3-kinase-mediated effects of glucose on vacuolar H+-ATPase assembly, translocation, and acidification of intracellular compartments in renal epithelial cells. *Mol. Cell. Biol.* **25**, 575-589 (2005).
60. P. M. Kane, Disassembly and reassembly of the yeast vacuolar H+-ATPase in vivo. *J. Biol. Chem.* **270**, 17025-17032 (1995).
61. M. Lu, Y. Y. Sautin, L. S. Holliday, S. L. Gluck, The glycolytic enzyme aldolase mediates assembly, expression, and activity of vacuolar H+-ATPase. *J. Biol. Chem.* **279**, 8732-8739 (2004).
62. M. Lu, L. S. Holliday, L. Zhang, W. A. Dunn, S. L. Gluck, Interaction between aldolase and vacuolar H+-ATPase evidence for direct coupling of glycolysis to the ATP-hydrolyzing proton pump. *J. Biol. Chem.* **276**, 30407-30413 (2001).
63. J. J. Marsh, H. G. Lebherz, Fructose-bisphosphate aldolases: an evolutionary history. *Trends Biochem. Sci.* **17**, 110-113 (1992).
64. W. Liebermeister *et al.*, Visual account of protein investment in cellular functions. *Proceedings of the National Academy of Sciences* **111**, 8488-8493 (2014).
65. S. L. Tai *et al.*, 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 (2007).
66. J. v. d. Brink *et al.*, Dynamics of glycolytic regulation during adaptation of *Saccharomyces cerevisiae* to fermentative metabolism. *Appl. Environ. Microbiol.* **74**, 5710-5723 (2008).
67. M. L. Jansen *et al.*, Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity. *Microbiology* **151**, 1657-1669 (2005).
68. M. Hebly *et al.*, Physiological and transcriptional responses of anaerobic chemostat cultures of *Saccharomyces cerevisiae* subjected to diurnal temperature cycles. *Appl. Environ. Microbiol.* **80**, 4433-4449 (2014).
69. S. R. Hackett *et al.*, Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, (2016).
70. F. Tripodi, R. Nicastro, V. Reghellin, P. Coccetti, Post-translational modifications on yeast carbon metabolism: regulatory mechanisms beyond transcriptional control. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1850**, 620-627 (2015).
71. N. Nagaraj *et al.*, System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Molecular & Cellular Proteomics* **11**, (2012).

72. D. G. Fraenkel, The top genes: on the distance from transcript to function in yeast glycolysis. *Curr. Opin. Microbiol.* **6**, 198-201 (2003).
73. A. Chambers, E. A. Packham, I. R. Graham, Control of glycolytic gene expression in the budding yeast (*Saccharomyces cerevisiae*). *Curr. Genet.* **29**, 1-9 (1995).
74. R. C. Vallari *et al.*, Glucose repression of the yeast *ADH2* gene occurs through multiple mechanisms, including control of the protein synthesis of its transcriptional activator, ADR1. *Mol. Cell. Biol.* **12**, 1663-1673 (1992).
75. P. Herrero, J. Galindez, N. Ruiz, C. Martinez-Campa, F. Moreno, Transcriptional regulation of the *Saccharomyces cerevisiae HXK1, HXK2* and *GLK1* genes. *Yeast* **11**, 137-144 (1995).
76. E. Boles *et al.*, Characterization of a glucose-repressed pyruvate kinase (Pyk2p) in *Saccharomyces cerevisiae* that is catalytically insensitive to fructose-1, 6-bisphosphate. *J. Bacteriol.* **179**, 2987-2993 (1997).
77. P. Daran-Lapujade *et al.*, The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proceedings of the National Academy of Sciences* **104**, 15753-15758 (2007).
78. P. Daran-Lapujade *et al.*, Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae* A chemostat culture study. *J. Biol. Chem.* **279**, 9125-9138 (2004).
79. H. Peter Smits *et al.*, Simultaneous overexpression of enzymes of the lower part of glycolysis can enhance the fermentative capacity of *Saccharomyces cerevisiae*. *Yeast* **16**, 1325-1334 (2000).
80. J. Hauf, F. K. Zimmermann, S. Müller, Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast *Saccharomyces cerevisiae*. *Enzyme Microb. Technol.* **26**, 688-698 (2000).
81. I. Schaaff, J. Heinisch, F. K. Zimmermann, Overproduction of glycolytic enzymes in yeast. *Yeast* **5**, 285-290 (1989).
82. A. P. Oliveira *et al.*, Regulation of yeast central metabolism by enzyme phosphorylation. *Mol. Syst. Biol.* **8**, (2012).
83. A. B. Vojtek, D. G. Fraenkel, Phosphorylation of yeast hexokinases. *Eur. J. Biochem.* **190**, 371-375 (1990).
84. Z. Liu *et al.*, CPLA 1.0: an integrated database of protein lysine acetylation. *Nucleic Acids Res.* **39**, D1029-D1034 (2011).
85. Z. Liu *et al.*, CPLM: a database of protein lysine modifications. *Nucleic Acids Res.* **42**, D531-D536 (2014).
86. M. A. Blazquez, R. Lagunas, C. Gancedo, J. M. Gancedo, Trehalose-6-phosphate, a new regulator of yeast glycolysis that inhibits hexokinases. *FEBS Lett.* **329**, 51-54 (1993).
87. G. Avigad, Stimulation of yeast phosphofructokinase activity by fructose 2, 6-bisphosphate. *Biochem. Biophys. Res. Commun.* **102**, 985-991 (1981).
88. J. H. van Heerden *et al.*, Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science* **343**, 1245114 (2014).
89. B. Teusink, M. C. Walsh, K. van Dam, H. V. Westerhoff, The danger of metabolic pathways with turbo design. *Trends Biochem. Sci.* **23**, 162-169 (1998).
90. K. Elbing *et al.*, Role of hexose transport in control of glycolytic flux in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **70**, 5323-5330 (2004).
91. L. B. Tanner *et al.*, Four key steps control glycolytic flux in mammalian cells. *Cell systems* **7**, 49-62. e48 (2018).
92. B. Teusink, H. V. Westerhoff, F. J. Bruggeman, Comparative systems biology: from bacteria to man. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2**, 518-532 (2010).
93. B. Teusink, D. Molenaar, Systems biology of lactic acid bacteria: For food and thought. *Current Opinion in Systems Biology* **6**, 7-13 (2017).
94. L. A. Fothergill-Gilmore, P. A. Michels, Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* **59**, 105-235 (1993).
95. L. Agius, Glucokinase and molecular aspects of liver glycogen metabolism. *Biochem. J.* **414**, 1-18 (2008).
96. J. Deutscher, C. Francke, P. W. Postma, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 939-1031 (2006).
97. F. R. Opperdoes, P. Borst, Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei:* the glycosome. *FEBS Lett.* **80**, 360-364 (1977).

98. B. M. Bakker *et al.*, Compartmentation protects trypanosomes from the dangerous design of glycolysis. *Proceedings of the National Academy of Sciences* **97**, 2087-2092 (2000).
99. J. R. Haanstra *et al.*, Compartmentation prevents a lethal turbo-explosion of glycolysis in trypanosomes. *Proceedings of the National Academy of Sciences* **105**, 17718-17723 (2008).
100. P. A. Michels, D. J. Rigden, Evolutionary analysis of fructose 2, 6-bisphosphate metabolism. *IUBMB life* **58**, 133-141 (2006).
101. D. Blangy, H. Buc, J. Monod, Kinetics of the allosteric interactions of phosphofructokinase from *Escherichia coli*. *J. Mol. Biol.* **31**, 13-35 (1968).
102. A. Mattevi, M. Bolognesi, G. Valentini, The allosteric regulation of pyruvate kinase. *FEBS Lett.* **389**, 15-19 (1996).
103. M. S. Jurica *et al.*, The allosteric regulation of pyruvate kinase by fructose-1, 6-bisphosphate. *Structure* **6**, 195-210 (1998).
104. E. Ponce, N. Flores, A. Martinez, F. Valle, F. Bolívar, Cloning of the two pyruvate kinase isoenzyme structural genes from *Escherichia coli*: the relative roles of these enzymes in pyruvate biosynthesis. *J. Bacteriol.* **177**, 5719-5722 (1995).
105. P. W. Postma, J. W. Lengeler, G. R. Jacobson, Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiol. Mol. Biol. Rev.* **57**, 543-594 (1993).
106. M. Callens, F. R. Opperdoes, Some kinetic properties of pyruvate kinase from Trypanosoma brucei. *Mol. Biochem. Parasitol.* **50**, 235-243 (1992).
107. A. Pontesucre *et al.*, Isolation of two pyruvate kinase activities in the parasitic protozoan Leishmania mexicana amazonensis. *Archives of biochemistry and biophysics* **300**, 466-471 (1993).
108. G. Giaever *et al.*, Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391 (2002).
109. J. M. Laurent, J. H. Young, A. H. Kachroo, E. M. Marcotte, Efforts to make and apply humanized yeast. *Briefings in functional genomics* **15**, 155-163 (2016).
110. E. J. Douzery, E. A. Snell, E. Bapteste, F. Delsuc, H. Philippe, The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proceedings of the National Academy of Sciences* **101**, 15386-15391 (2004).
111. J. A. Barnett, C. F. Robinow, A history of research on yeasts 4: cytology part II, 1950–1990. *Yeast* **19**, 745-772 (2002).
112. B. Pulverer, Trio united by division as cell cycle clinches centenary Nobel. *Nature* **413**, 553 (2001).
113. J. A. Barnett, A history of research on yeasts 5: the fermentation pathway. *Yeast* **20**, 509-543 (2003).
114. A. Goffeau *et al.*, Life with 6000 genes. *Science* **274**, 546-567 (1996).
115. B. Adiego-Perez *et al.*, Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiol. Lett.* **366**, (2019).
116. G. Giaever, C. Nislow, The yeast deletion collection: a decade of functional genomics. *Genetics* **197**, 451-465 (2014).
117. D. Botstein, S. A. Chervitz, M. Cherry, Yeast as a model organism. *Science* **277**, 1259-1260 (1997).
118. W. H. Mager, J. Winderickx, Yeast as a model for medical and medicinal research. *Trends Pharmacol. Sci.* **26**, 265-273 (2005).
119. L. Miller-Fleming, F. Giorgini, T. F. Outeiro, Yeast as a model for studying human neurodegenerative disorders. *Biotechnology Journal: Healthcare Nutrition Technology* **3**, 325-338 (2008).
120. L. M. Steinmetz *et al.*, Systematic screen for human disease genes in yeast. *Nat. Genet.* **31**, 400 (2002).
121. A. Hamza *et al.*, Complementation of yeast genes with human genes as an experimental platform for functional testing of human genetic variants. *Genetics* **201**, 1263-1274 (2015).
122. A. H. Kachroo *et al.*, Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **348**, 921-925 (2015).
123. R. K. Garge, J. M. Laurent, A. H. Kachroo, E. M. Marcotte, Systematic humanization of the yeast cytoskeleton discerns functionally replaceable from divergent human genes. *bioRxiv*, (2019).
124. F. Yang *et al.*, Identifying pathogenicity of human variants via paralog-based yeast complementation. *PLoS Genet.* **13**, e1006779 (2017).
125. S. Sun *et al.*, An extended set of yeast-based functional assays accurately identifies human disease mutations. *Genome Res.* **26**, 670-680 (2016).

126. J. M. Laurent *et al.*, Humanization of yeast genes with multiple human orthologs reveals principles of functional divergence between paralogs. *bioRxiv*, 668335 (2019).
127. N. Zhang *et al.*, Using yeast to place human genes in functional categories. *Gene* **303**, 121-129 (2003).
128. A. Hamza, M. R. Driessen, E. Tammpere, N. J. O'Neil, P. Hieter, Cross-species complementation of nonessential yeast genes establishes platforms for testing inhibitors of human proteins. *Genetics*, (2020).
129. M. S. Skrzypek *et al.*, *Saccharomyces* genome database informs human biology. *Nucleic Acids Res.* **46**, D736-D742 (2017).
130. N. Agmon *et al.*, Phylogenetic debugging of a complete human biosynthetic pathway transplanted into yeast. *Nucleic Acids Res.*, (2019).
131. D. M. Truong, J. D. Boeke, Resetting the yeast epigenome with human nucleosomes. *Cell* **171**, 1508-1519. e1513 (2017).
132. S. Wildt, T. U. Gerngross, The humanization of N-glycosylation pathways in yeast. *Nature Reviews Microbiology* **3**, 119 (2005).
133. D. G. Gibson *et al.*, Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215-1220 (2008).
134. S. M. Richardson *et al.*, Design of a synthetic yeast genome. *Science* **355**, 1040-1044 (2017).
135. S. Galanie, K. Thodey, I. J. Trenchard, M. F. Interrante, C. D. Smolke, Complete biosynthesis of opioids in yeast. *Science* **349**, 1095-1100 (2015).
136. U. Güldener, S. Heck, T. Fiedler, J. Beinhauer, J. H. Hegemann, A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24**, 2519-2524 (1996).
137. H. Redden, H. S. Alper, The development and characterization of synthetic minimal yeast promoters. *Nat. Common.* **6**, 7810 (2015).
138. G. Naseri *et al.*, Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **6**, 1742-1756 (2017).
139. R. M. Portela *et al.*, Synthetic core promoters as universal parts for fine-tuning expression in different yeast species. *ACS Synth. Biol.* **6**, 471-484 (2017).
140. J. Blazeck, R. Garg, B. Reed, H. S. Alper, Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnol. Bioeng.* **109**, 2884-2895 (2012).
141. C. J. Harvey *et al.*, HEx: a heterologous expression platform for the discovery of fungal natural products. *bioRxiv*, 247940 (2018).
142. M. Naesby *et al.*, Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* **8**, 45 (2009).

Chapter 2

Pathway swapping: towards modular engineering of essential cellular processes

Niels G.A. Kuijpers, Daniel Solis-Escalante, Marijke A.H. Luttik, Markus M.M. Bisschops, Francine J. Boonekamp, Marcel van den Broek, Jack T. Pronk, Jean-Marc Daran, Pascale Daran-Lapujade.

Published in Proceedings of the National Academy of Sciences of the United States of America (PNAS) **113**, 15060-15065 (2016)

Abstract

Recent developments in synthetic biology enable one-step implementation of entire metabolic pathways in industrial microorganisms. A similarly radical remodelling of central metabolism could greatly accelerate fundamental and applied research, but is impeded by the mosaic organization of microbial genomes. To eliminate this limitation, we propose and explore the concept of "pathway swapping," using yeast glycolysis as the experimental model. Construction of a "single-locus glycolysis" *Saccharomyces cerevisiae* platform enabled quick and easy replacement of this yeast's entire complement of 26 glycolytic isoenzymes by any alternative, functional glycolytic pathway configuration. The potential of this approach was demonstrated by the construction and characterization of *S. cerevisiae* strains whose growth depended on two non-native glycolytic pathways: a complete glycolysis from the related yeast *Saccharomyces kudriavzevii* and a mosaic glycolysis consisting of yeast and human enzymes. This work demonstrates the feasibility and potential of modular, combinatorial approaches to engineering and analysis of core cellular processes.

Introduction

Replacement of petrochemistry by bio-based processes is a key element for sustainable development and requires microbes equipped with novel-to-nature capabilities. Recent developments in synthetic biology enable introduction of entire metabolic path-ways and, thereby, new functionalities for product formation and substrate consumption, into microbial cells (*1*). However, industrial relevance of the resulting strains critically depends on optimal interaction of the newly introduced pathways with the core metabolism of the host cell. Central metabolic pathways such as glycolysis, tri-carboxylic acid cycle, and pentose phosphate pathways, are essential for synthesis of precursors, for providing free energy (ATP), and for redox-cofactor balancing. Optimization of productivity, product yield, and robustness therefore requires modifications in the configuration and/or regulation of these core metabolic functions. Engineering of central metabolism is in some respects more challenging than the functional expression of heterologous product pathways. Millions of years of evolution of microorganisms have endowed their metabolic and regulatory networks with a level of complexity that cannot be efficiently reengineered by iterative, single-gene modifications. Enzymes of central metabolism are encoded by hundreds of genes that, especially in eukaryotes, are scattered across microbial genomes. Moreover, inactivation and subsequent replacement of genes involved in central metabolism is complicated by functional redundancy of isoenzymes (*2, 3*) as well as by the essential role of many of the corresponding biochemical reactions. Microbial platforms in which the configuration of key pathways can be remodelled in a swift, combinatorial manner would provide an invaluable asset for fundamental research and engineering of central metabolism. Whereas rapid, cost-effective assembly of entire synthetic genomes is becoming a realistic perspective for small bacterial genomes (*4, 5*), routine synthesis and expression of entire eukaryotic genomes is unlikely to be implemented in the next few years. Here, we propose and experimentally explore a modular approach to the engineering of central metabolism that involves versatile, synthetic microbial strain platforms in which entire metabolic pathways can, in a few simple steps, be replaced by any functional, newly designed configuration. As a proof of principle, we set out to construct a platform that enables swapping of the entire Embden–Meyerhoff–Parnas pathway of glycolysis, a strongly conserved metabolic highway for sugar utilization in the model eukaryote and industrial yeast *Saccharomyces cerevisiae*. Including the reactions leading to the formation of ethanol, the main fermentation product of *S. cerevisiae,* yeast glycolysis encompasses a set of 12 reactions, catalysed by no fewer than 26 cytosolic isoenzymes. Several of these (e.g., Tpi1, Tdh3, and Adh1) are among the most abundant proteins in yeast cells. The genes encoding glycolytic enzymes are

scattered over 12 of the 16 yeast chromosomes. Construction of a platform for glycolysis swapping involved a two-step approach (Fig.1A). In the first step, described in a recent study by our group, the genetic complexity of yeast glycolysis was reduced by deleting the structural genes for 13 of the 26 glycolytic enzymes. Remarkably, a detailed systems analysis revealed that, under laboratory conditions, the phenotype of the resulting minimal glycolysis (MG) strain was virtually identical to that of the parental strain carrying a full complement of glycolytic genes (*3*). In a next step, the remaining 13 glycolytic genes in MG were expressed from a single chromosomal locus. Finally, the remaining scattered native genes were removed from their original loci, leading to switchable yeast glycolysis (SwYG), a yeast strain carrying a native, minimal single-locus glycolysis on chromosome IX. In SwYG, glycolysis can be swapped in two steps by integration of a new, heterologous, or synthetic glycolytic gene cluster, followed by removal of the minimal single-locus glycolysis that was initially integrated on chromosome IX (Fig. 1B).

Results

Engineering of a yeast platform for glycolysis swapping

A single-locus native glycolysis gene cluster was assembled from 'glycoblocks' (Fig. 1, SI Appendix, Fig. S1), 13 DNA cassettes each consisting of a *S. cerevisiae* glycolytic gene, including its native promoter and terminator, flanked by 60-bp Synthetic Homologous Recombination (SHR) sequences (*6*). SHR sequences share no homology with the *S. cerevisiae* genome and can be used for efficient *in vivo* assembly and integration, by homologous recombination, of the glycoblocks. Moreover, use of standardized SHR sequences enables flexible design and combinatorial assembly of different glycolytic pathway variants. The single-locus minimal glycolysis cluster was composed of 13 glycoblocks (corresponding to the 13 genes remaining in MG) and of a selectable marker (amdSYM (*7*)) flanked by SHR sequences (Fig. 1A).

The single-locus native glycolytic cluster was integrated into the yeast genome to promote stable, single-copy expression, using Combined Assembly and Targeted Integration (CATI, (*8*)) (SI Appendix, Fig. S1). To this end, recognition sequences for the I-SceI homing endonuclease were introduced at the *SGA1* locus on chromosome IX of MG (SI Appendix Fig. S1). The 13 glycoblocks and the amdSYM cassette were co-transformed to the modified MG strain, in which *SCEI* was induced by growth on galactose to introduce a double-strand DNA break at the *SGA1* locus, thereby promoting integration of the glycoblocks (Fig. 2a, Fig. S1). Four of five tested transformant colonies harboured the complete 35-kb SinLoG-IX (Single Locus native Glycolysis) integrated at the *SGA1* locus. In a selected transformant (strain IMX382) Next-Generation (Next-Gen)

Sequencing showed that the *in vivo*-assembled and integrated glycolytic gene cluster was virtually identical to its *in silico* blueprint (Fig. 2a). Six out of the nine deviations in nucleotide sequence were found at the HR loci linking the glycoblocks, which may either reveal recombinase-based errors or simply errors in the primers used to construct the HR sequences. Of the glycolytic genes only *ADH1* was found to contain a mutation which was synonymous (A180A).

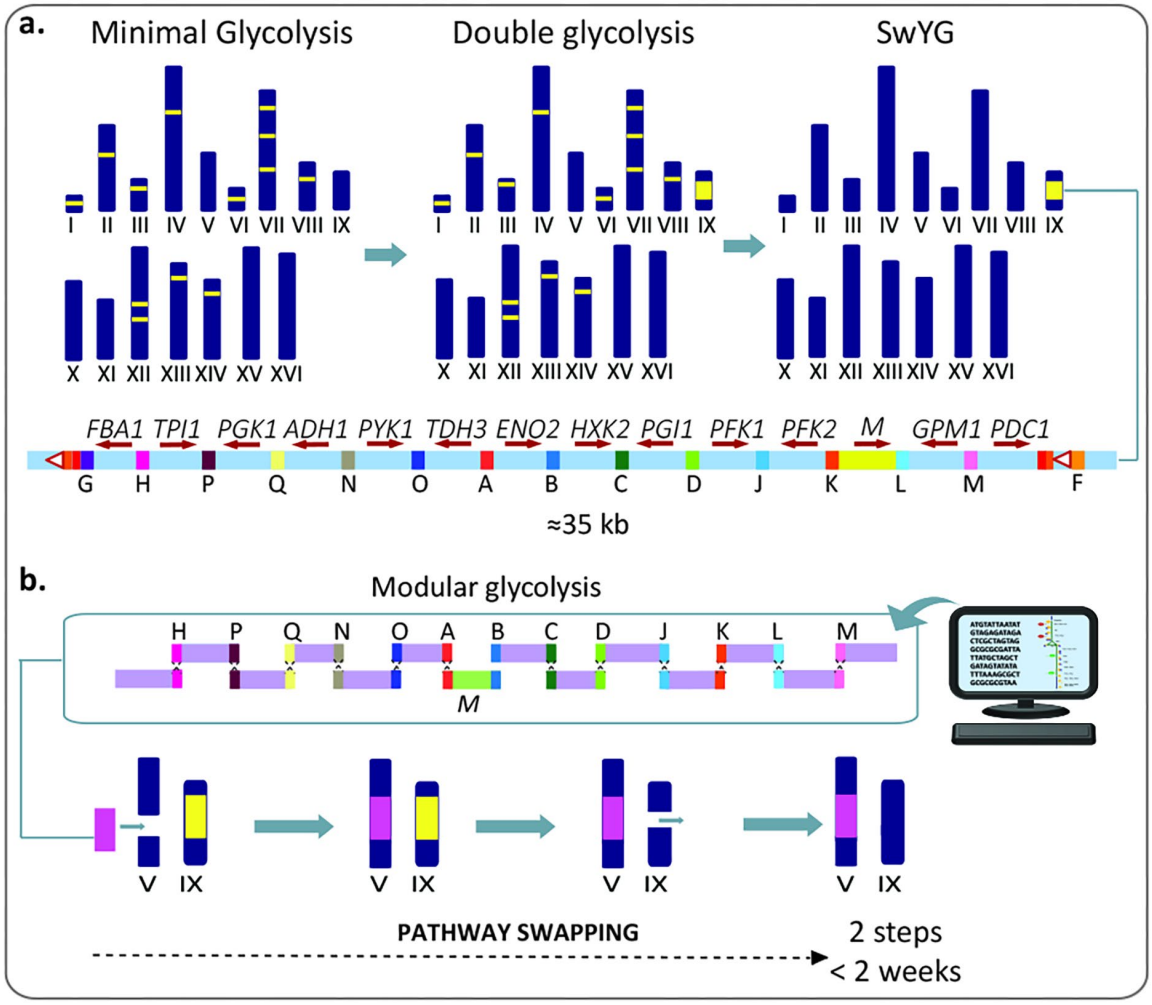


Figure 1 – Schematic overview of the glycolysis swapping approach. A) Construction of SwYG that contains a single locus endogenous glycolysis platform for pathway swapping. **B)** *In silico* design and *in vivo* assembly and integration of the glycolytic gene cluster on chromosome V, followed by the removal of the endogenous glycolysis on chromosome IX, leading to a strain with a redesigned glycolysis. M, selectable marker

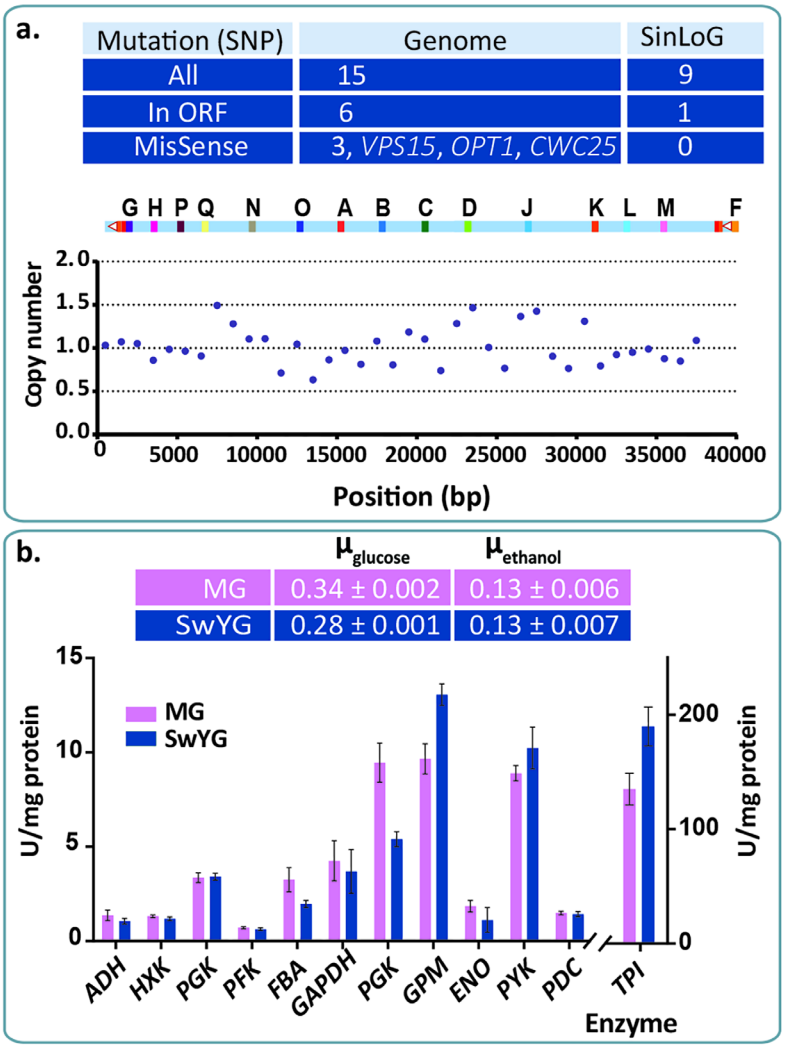


Figure 2 – Characterization of SwYG. A) Next-gen sequencing and copy number analysis of the auxotrophic SwYG (IMX589) containing a clustered set of glycolytic genes (SinLoG-IX). **B)** Physiological characterization in shake-flask culture using chemically defined medium with glucose as carbon source of IMX606 (prototrophic SwYG strain) and the MG strain (IMX370). Growth rates (hour^{-1}) and enzyme activity data represent the average and SEM of at least two independent culture replicates.

S. cerevisiae IMX382 was further engineered by deleting the 13 remaining glycolytic genes from their native loci (Fig. 1a, SI Appendix, Fig. S2). The first five deletions, targeting *PYK1*, *PGI1*, *TPI1*, *TDH3* and *PGK1*, were performed with standard deletion cassettes, using I-SceI-mediated marker removal to recycle multiple selection markers simultaneously without leaving scars in the genome (*9*) (SI Appendix, Fig. S3). For subsequent engineering, an expression cassette encoding the CRISPR endonuclease Cas9 was integrated at the *PFK2* locus, thereby deleting *PFK2* (SI Appendix, Fig. S4). The remaining glycolytic genes, *PFK1*, *GPM1*, *HXK2*, *FBA1*, *ADH1* and *PDC1*, were deleted from their native loci with the CRISPR/Cas9 system (*10*). Except for *ENO2*, all glycolytic expression cassettes harboured by the SinLoG-IX cluster were able to complement a

null mutation in the corresponding native gene. The native promoter of *ENO2*, designed to be 411-bp long to avoid expressing unwanted open reading frame from the glycolytic gene cluster, proved to be too short to drive expression of *ENO2* and was replaced by a longer promoter (1012 bp, SI Appendix, Table S9). This observation underlines the limited knowledge, even for glycolytic genes, on promotor structure and function in yeast and highlights the need for systematic design of synthetic promoter. This last genetic modification yielded SwYG (IMX589). Whole genome sequencing of this strain confirmed: i) the correct sequence of the single-locus native glycolysis (SI Appendix, Table S1) and its integration at the *SGA1* locus, ii) deletion of the native glycolytic genes from their original loci (SI Appendix, Fig. S5) and iii) absence of duplicated glycolytic genes in the single-locus native glycolysis and in the genome (SI Appendix, Fig. S6). Relative to the ancestor MG strain, only six open reading frames in the genome of SwYG contained a nucleotide difference. Three of these caused an amino acid substitution in the encoded protein (Fig. 2b, SI Appendix, Table S2). None affected glycolytic genes or genes that are known to be otherwise associated with glycolysis. The specific growth rate of the prototrophic SwYG (IMX606) measured in aerobic batch in shake-flask culture, on chemically defined medium using glucose as carbon source, was slightly lower than that of the parental MG strain (17% lower, Fig. 2b). However, more accurate quantification of specific growth rates in tightly controlled bioreactors revealed a stronger impact of the relocalization of the glycolytic gene cluster. The specific growth rate of SwYG was decreased by 28% as compared to its parent MG (Fig. 3). Concomitantly the major metabolic fluxes, *i.e.* the specific glucose uptake and ethanol production rates, were decreased in SwYG. Remarkably, the biomass and product yields on glucose remained unaffected by the relocalization of the glycolytic genes to chromosome IX (Fig. 3), showing that the growth stoichiometry was conserved and that only the magnitude of fluxes was attenuated in SwYG. These decreased rates of SwYG are unlikely to result from differences in glycolytic capacity as the activities of glycolytic enzymes in cell extracts of these two strains were highly similar (Fig. 2b).

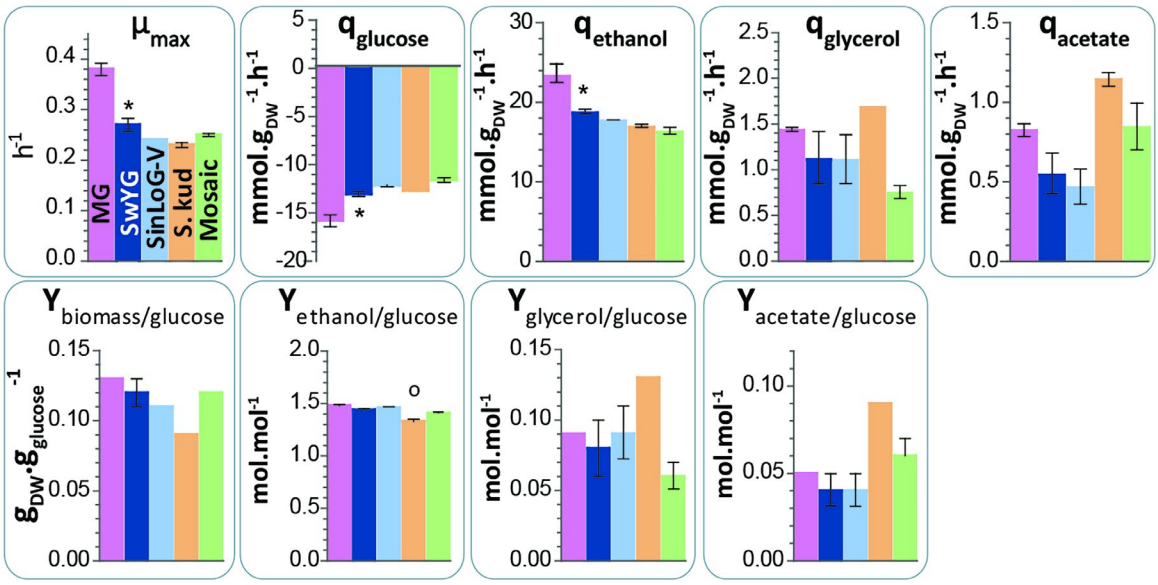


Figure 3 – Physiological characterization during aerobic batches in bioreactors of MG (IMX370), SwYG (IMX606), SinLoG-V (IMX605), Sk-SinLoG-V (S. kud, IMX652), and mosaic SinLoG-V (Mosaic, IMX645). The strains were cultivated in chemically defined medium with glucose as carbon source. Bars and error bars represent the average and SEM of independent duplicate cultures. Stars indicate that the data from SwYG significantly differ from MG data; empty dots indicate significant differences between SwYG and the SwYG-based strains (Pvalue<0.05, two-tailed t test, samples with equal variance). DW, biomass dry weight.

Chromosome hopping of a yeast glycolysis gene cluster

To test the feasibility of glycolysis swapping, we attempted to exchange the single-locus native glycolysis integrated on chromosome IX by a nearly identical copy integrated on chromosome V (SI Appendix, Fig. S7). SwYG was transformed with a complete set of glycoblocks (Fig. 4a), and with a CRISPR plasmid carrying a guide RNA designed to target Cas9 to the *CAN1* locus on Chromosome V, thus enabling *in vivo* assembly and integration of a second glycolytic genes cluster called SinLoG-V. Colony PCR showed that at least two of 12 G418-resistant transformants carried the complete set of genes in this SinLoG-V, correctly inserted at the *CAN1* locus. After curing of the *URA3*-carrying CRISPR plasmid, a selected clone was transformed with a 120-bp repair fragment and a new CRISPR plasmid carrying gRNAs targeting Cas9 to sequences positioned at each end of the SinLoG-IX, thereby excising the entire cluster from the genome. All three tested transformants were shown to lack the SinLoG-IX and to have retained the newly inserted single-locus native glycolysis on chromosome V. Whole-genome sequencing of one clone (IMX605) confirmed the successful relocalization of the entire glycolytic gene cluster and that no recombination had occurred between the glycoblocks or the excised single-locus native glycolysis gene cluster and the genome during glycolysis swapping (SI Appendix, Fig. S6). IMX605 grew as fast as SwYG in chemically defined medium (Fig.

4c) and displayed the same activity of the glycolytic enzymes in cell extracts (Fig. 4d). In IMX605, the position of the selectable marker and the *ENO2* glycoblock were reversed as compared to the SinLoG-IX cluster in SwYG. However, this different organization did not affect *ENO2* expression (Fig. 4d). The similarity in specific growth rate between SwYG and IMX605 was also observed during growth in bioreactors (Fig. 3). These cultures also revealed identical metabolic rates and yields in these two strains (Fig. 3). This lack of locus-specific expression demonstrated that the *CAN1* locus on chromosome V was a suitable 'landing pad' for further testing of the pathway swapping concept.

***Saccharomyces cerevisiae* expressing a heterologous glycolytic pathway**

Demonstration of the technical feasibility of pathway swapping opened up the way to test whether it is possible to integrally replace yeast glycolysis, an essential, tightly controlled metabolic pathway, by heterologous or synthetic variants. For this purpose, we selected a donor of glycolytic genes from within the *Saccharomyces* genus. *S. kudriavzevii* is a cold-tolerant close relative of *S. cerevisiae*, recently identified as an important contributor to wine making in cool climates (*11, 12*). While glycolytic genes and enzymes of *S. kudriavzevii* have not been characterized in detail, its genome sequence is available (*13*). The complement of putative glycolytic genes in *S. kudriavzevii* and their sequences differ from those of the established *S. cerevisiae* glycolytic genes. However, putative *S. kudriavzevii* glycolytic genes with substantial identity (above 89% at the protein level) with their *S. cerevisiae* orthologs were easily identified by sequence comparison (SI Appendix, Table S3). Major glycolytic isoenzymes in *S. kudriavzevii* were selected based on high transcript levels of their structural genes (*14*). *S. kudriavzevii* does not contain a homolog of *ScTDH3*, the most highly expressed glycolytic gene in *S. cerevisiae* (*15*), but harbours two other putative *TDH* genes. *SkTDH1* closely resembles *ScTDH1*, while *SkTDH2* is more similar to *ScTDH2* and *ScTDH3*. Based on its high expression level during wine fermentation, the gene homologous to *ScTDH1* was selected for construction of a synthetic *S. kudriavzevii* glycolysis cluster. Since promoters within the *Saccharomyces* genus are functional in different species belonging to this group (*16, 17*) and promoters of *S. kudriavzevii* and *S. cerevisiae* are highly homologous (44 to 80% identity with an average of 72% for glycolytic promoters), *S. kudriavzevii* glycoblocks were constructed with their native promoters and terminators. Following the approach described above, the single-locus native glycolysis in SwYG was replaced by the *S. kudriavzevii* glycolysis integrated on chromosome V (*sk*-SinLoG-V, Fig. 4a). Transformants that only retained the *S. kudriavzevii* glycolysis, including a selected clone, IMX637, showed a strongly reduced specific growth rate on glucose, suggesting that the glycolytic function of the integrated

set of *S. kudriavzevii* genes was suboptimal (Fig. 4c). Since *SkTDH1* was not an ortholog of *ScTDH3*, we hypothesized that insufficient glyceraldehyde-3-phosphate dehydrogenase (GAPDH) activity caused the suboptimal specific growth rate. Indeed, the specific growth rate was almost completely restored to the level of SinLoG-V (IMX605) when *SkTDH1* was overexpressed in the *Sk*-SinLoG-V strain IMX637 (strain IMX652, Fig. 4c). Accordingly GAPDH activity, very low in the *Sk*-SinLoG-V strain, was boosted by overexpressing *skTDH1* (Fig. 4d). Enzyme activity assays also revealed a very low activity of phosphofructokinase in cell extracts of the *Sk*-SinLoG-V strain. Remarkably, this activity was fully restored upon overexpression of *skTDH1* (Fig. 4d). During growth in bioreactor IMX652, carrying an overexpression of *SkTDH1*, grew nearly as fast as SwYG, but seemed to display a perturbation of its metabolic network (Fig. 3). Most of the observed increase in specific glycerol and acetate production rates and yields in IMX652 were however not deemed statistically significant as compared to SwYG (Fig 3). While these variations fell within measurement error, they would be consistent with a perturbation of flux distribution at the glyceraldehyde-3-phosphate branch point in favour of glycerol synthesis. Altogether these results illustrate how pathway swapping can identify interesting regulatory phenomena and identifies interesting leads for follow-up studies on the largely characterized *S. kudriavzevii* glycolytic enzymes.

***Saccharomyces cerevisiae* expressing a mosaic glycolysis**

To further test the pathway-swapping concept, we constructed a mosaic SinLoG composed of a combination of five *S. cerevisiae*, five *S. kudriavzevii* and two *Homo sapiens* genes. *HsTPI1* and *HsPGK1* can complement null mutations in their *S. cerevisiae* orthologs (*18, 19*). The most abundant splicing variant of *HsTPI1 (20)* and the single splicing variant of *HsPGK1* from muscle were codon-optimized for expression in *S. cerevisiae* (SI Appendix, Table S4) and each stitched to the promoter and terminator of their respective yeast orthologs. The resulting human glycoblocks were pooled with *HXK2*, *TDH3*, *PYK1*, *FBA1* and *PDC1* glycoblocks from *S. cerevisiae* and *PGI1*, *PFK1*, *PFK2*, *ENO2*, *GPM1* and *ADH1* from *S. kudriavzevii* and transformed to SwYG, resulting in the integration of a mosaic glycolysis gene cluster in chromosome V (Mosaic-SinLoG-V gene cluster). Subsequent removal of the native SinLoG-IX gene cluster yielded strain IMX645, carrying a mosaic single locus glycolytic gene cluster encoding a set of enzymes capable of supporting the entire glycolytic flux. Whole-genome sequencing confirmed the absence of the SinLoG-IX and the presence of the complete Mosaic-SinLoG-V. Just a single, silent nucleotide variation was detected within an ORF of the mosaic SinLoG IMX645 strain and its mosaic-SinLoG-V cluster was identical to the *in silico* design (Fig.4b, SI Appendix, Fig. S5 and Table S1). Although *HsTPI1* and *HsPGK1* expression

was driven by the *ScTPI1* and *ScPGK1* promoters, *in vitro* enzyme activity of both encoded enzymes was ca. 50 % lower in the mosaic glycolysis strain than in the native SinLoG-V IMX605 strain (Fig. 4c, t-test p-value < 0.01). Also the activity of *Sk*ADH was ca. 50% lower in IMX645 than in strain IMX605 (Fig. 4d, t-test p-value < 0.01). Despite these lower enzyme activities, the strain carrying the mosaic-SinLoG-V grew as fast as SinLoG-V IMX605 both in shake-flask and bioreactor, and its metabolic fluxes were undisturbed (Fig. 3 and 4), consistent with the notion that most glycolytic enzymes in *S. cerevisiae* have an overcapacity under standard laboratory growth conditions (*15*).

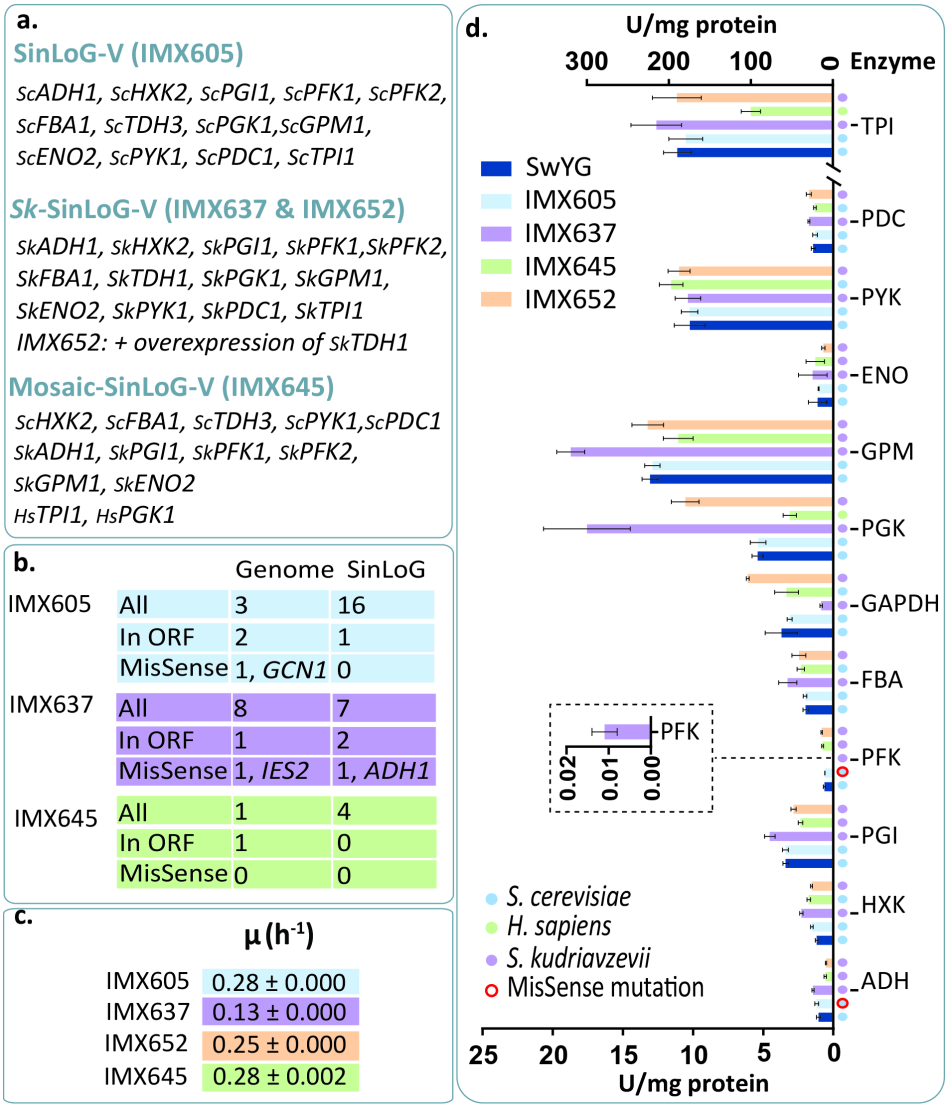


Figure 4 – Characterization of yeast strains with a remodelled glycolysis. A,B) SinLoG-V (IMX605), Sk-SinLoG-V (IMX637), and mosaic SinLoG-V (IMX645) were analyzed by next-gen sequencing to identify mutations compared with SwYG. **C)** The maximum specific growth rate on chemically defined medium with glucose as carbon source was measured, **D)** as well as *in vitro* enzymatic activity of the glycolytic enzymes from cell extracts. Data represent the average and SEM of at least two independent culture replicates.

Discussion

This study demonstrates how modern genome-editing techniques can make essential biological processes, the partially redundant genetic information for which is scattered over an entire eukaryotic genome, accessible to fast, combinatorial analysis and optimization.

The high efficiency of the pathway swapping approach exceeded our expectations. Pathway swapping involves the transient, simultaneous presence in the yeast nucleus of two SinLoGs sharing high, and in one of the experiments even near-complete, sequence identity. One of these is integrated in chromosome V while a second, 35-Kb SinLoG is excised from chromosome IX. DNA ends are highly recombinogenic and can interact directly with homologous sequences (*21*) and homologous recombination is the main double-strand break repair mechanism in growing *S. cerevisiae* (*22, 23*). The excised SinLoG-IX, or fragments generated by unspecific nuclease activity, might recombine with the newly integrated SinLoG-V. In practice, unintended genome rearrangements caused by homologous recombination, either between the engineered genetic elements themselves or between engineered genetic elements and the native yeast genome, were not observed. These results highlight the amazing efficiency and versatility of *in vivo* assembly and CRISPR/Cas9-facilitated genome editing in *S. cerevisiae (6, 24)*.

While genetic reduction of the glycolytic pathway did not lead to a detectable phenotype (*3*), clustering of the entire glycolytic gene set on a single locus resulted in decreased growth rate and metabolic fluxes. A logical explanation for this phenotype could be a decreased glycolytic capacity resulting from reduced expression of the clustered and relocalized glycolytic genes, however the activity of the glycolytic enzymes measured *in vitro* remained remarkably similar between the SwYG and MG strains. Relocalization of the entire glycolytic gene cluster from chromosome IX to V further substantiated the insensitivity of gene expression to large scale targeted genome remodelling in *S. cerevisiae*. While it is well documented that the genetic context can strongly influence gene expression, current knowledge does not allow for predictions on how the genomic site at which large synthetic gene clusters are integrated will affect transcription of genes harboured by such clusters. Using single-gene reporter systems, several studies have shown that localization in the vicinity of centromeres and telomeres leads to gene silencing while proximity to autonomously replicating sequences (ARS) tends to enhance transcription (*25, 26*). Although not completely understood at a mechanistic level, nucleosome positioning also modulates transcription (*27*). In the present study, glycolytic genes were concatenated and

integrated at two different loci, without noticeable impact on expression. While these rearrangements appear drastic, our design aimed at limiting epigenetic effects by integrating the glycolytic gene clusters in regions distant (> 30 Kb) from telomeres and centromeres, and from active ARS (> 3 Kb). Furthermore, nucleosome positioning in promoter regions is important for transcription and is strongly influenced by promoter sequences (*28*). In the SinLoGs, expression of the glycolytic genes was driven by *S. cerevisiae* promoters or by highly homologous *S. kudriavzevii* promoters. The genetic design of the SinLoGs may therefore explain the remarkable insensitivity of gene expression to relocalization.

Promoters of glycolytic genes are among the strongest in yeast and, consistent with the essentiality of the encoded proteins, glycolytic genes are constitutively expressed (*29*). Co-localization, in a single locus, of 13 genes that are heavily loaded with RNA polymerase II, might affect the conformation of DNA and thereby locally affect transcription. Moreover, Pol II disruption of chromatin has been proposed to increase sensitivity to several stresses (*30*), and to affect binding of proteins such as cohesin that play an important role in genetic stability (*31*). This co-localization of glycolytic genes into a 35-Kb transcriptional hotspot had remarkably little impact on the expression of the glycolytic enzymes. Similar to the recently developed "telomerator" (*26*), SwYG offers an attractive experimental model to systematically explore the impact of broader genomic context and, thereby, to guide the *de novo* design of synthetic yeast chromosomes.

The present data suggest that the reduced growth phenotype of the SwYG strain lineage is unlikely to result from a reduced glycolytic activity. Several other mechanisms, directly or indirectly related to the pathway swapping concept, could contribute this slower growth phenotype. A potential epigenetic factor is DNA replication and the requirement for regularly spaced ARS along chromosomes. While ARS are typically spaced by 30-40 Kb in *S. cerevisiae (32)*, insertion of the 35 Kb sequence carrying the glycolytic genes resulted in a spacing between adjacent confirmed ARS (ARS504.2 and ARS507, and ARS912 and ARS913) of 82 and 74 Kb for chromosome V and IX respectively (*33*). In their design for a synthetic chromosome III, Annaluru and co-workers kept a conservative approach by maintaining 12 out of the 19 native ARS, with a maximum spacing between ARS of ca. 50 Kb (*34*). Although it has been proposed that 120 to 300 Kb of chromosomal DNA could be replicated from a single replication origin (*35*), a quantitative evaluation of the impact on the physiology of *S. cerevisiae* of increasing spacing between ARS would facilitate the design of large chromosomal constructs and synthetic chromosomes. Another factor potentially involved in the slow growth phenotype of SwYG is the secondary function, unrelated to their catalytic role

in glycolysis, of three glycolytic enzymes (*36*), which may be affected by the genetic relocalization. However, involvement of these secondary functions in the slow growth phenotype of SwYG does not appear very likely as the vacuolar role of Fba1 and Eno2 is not expected to lead to visible growth defects under acidic environments (the pH used was six or lower, (*37, 38*)) and the physiological characterization of SwYG did not suggest an altered regulatory activity of Hxk2 (*39*). Alternatively, we cannot rule out that factors external to the clustering and relocalization of glycolytic genes are responsible for the slower growth of SwYG. For instance three genes, *VPS15*, *CWC25* and *OPT1*, of which the first two are essential, have a missense mutation in SwYG as compared to its parent MG. These mutations may be deleterious for the growth of *S. cerevisiae*. Further research is ongoing to evaluate contribution of these factor in the slow growth phenotype of the SwYG and SwYG-derived strains.

Functional replacement of the entire *S. cerevisiae* glycolysis by that of its close relative, the cold tolerant yeast *S. kudriavzevii*, provided a proof of principle that pathway swapping can be used to rapidly express and study entire metabolic pathways in a heterologous context. *S. kudriavzevii* and *S. cerevisiae* are sympatric and both show fast, fermentative sugar dissimilation in glucose-rich media (*40*). Pathway swapping demonstrated that a set of *S. kudriavzevii* glycolytic enzymes can support glycolysis and growth of *S. cerevisiae*. Co-evolution in the same ecological niches may have led to similar optima in term of expression level and transcriptional regulation and explain the highly similar activities observed for most glycolytic enzymes upon replacement of all *S. cerevisiae* glycolytic genes by their *S. kudriavzevii* counterparts, controlled by their native promoters. Functional replacement of the full complement of 26 glycolytic genes in *S. cerevisiae*, deletion of a significant number of which is lethal in wild-type strains, by a set of 13 heterologous variants, demonstrates the potential of the pathway-swapping concept for studying essential metabolic pathways. Moreover, it paves the way for modifications and, indeed, complete redesign of other multi-gene, essential cellular processes.

*C*a. 60% of the *S. cerevisiae* genes share significant homology with human genes (*41*) and, moreover, an estimated 30% (*42*) of human genes connected with specific diseases have a yeast ortholog. The popularity of *S. cerevisiae* as a model eukaryote is further boosted by its experimental tractability and by the availability of a wide range of tools and technologies, that make this yeast particularly well-suited for high throughput studies (*43*). 'Humanized' yeast strains provide powerful models to explore effects of therapeutics, gene dosage and of wild-type or disease-causing variants of human genes on protein function (*43*). Recent examples include breakthroughs in research on cell-autonomous mechanisms of neurodegeneration and identification of drug candidates

against neurodegenerative diseases (*44, 45*). Recent large-scale studies on the ability of human genes to complement native genes in *S. cerevisiae* demonstrated that complementation of haploid yeast gene knockouts is a reliable approach for functional characterization of human gene variants (*43*). However elegant, many such studies are limited to single-gene complementation and require the generation of multiple yeast strains that each contain only a single ortholog of a studied heterologous gene. Pathway swapping enables the systematic analysis of heterologous complementation of entire pathways and should enable humanization of, for example, the complete glycolytic pathway. Availability of strains containing a fully or partially humanized glycolytic pathway will enable to test the impact of mutations or drugs on human proteins in their natural glycolytic context, and thereby to identify potential synergetic effects between native human proteins.

The modular pathway swapping approach opens up unprecedented possibilities. Applications ranging from functional analysis of heterologous proteins, testing of kinetic models (now hindered by the multiplicity of paralogs (*46*)) or screening drugs, to more technical aspects such as exploring the effect of genomic location of highly expressed native pathways, are now within reach. Continued improvements in CRISPR-mediated removal of scattered genes (*24*) should even further facilitate functional clustering and fast, modular swapping of key pathways/processes.

A worldwide research effort has already led to the first synthetic yeast chromosome and is progressing towards the synthesis of an entire yeast genome (*34*). The present study demonstrates that a modular design of such synthetic yeast genomes, in which the genetic information for key processes is functionally clustered, offers unprecedented possibilities for fast, combinatorial exploration and optimization of the biological function of multi-gene, essential cellular functions.

Acknowledgements

We thank M. Niemeijer, R. Brinkman and M. Zaidi for their valuable contribution to the construction and confirmation of the *S. kudriavzevii* and *H. sapiens* glycoblocks. We thank Eladio Barrio Esparducer for kindly providing *S. kudriavzevii* CR85 and for advice on selecting the genes to design *S. kudriavzevii* minimal glycolysis. This work was supported by the Technology Foundation STW (Vidi grant 10776) and by the European Research Council (CoG-648141-AdLibYeast).

Material and Methods

Strains and media

Saccharomyces kudriavzevii CR85 (supplied by Prof. Querol, Universitat de València, València, Spain) a wild isolate from oak bark (Ciudad Real, Spain) (*11*) (Table S5), was grown at 16 ˚C on YPD medium containing (10 $g \cdot l^{-1}$ Bacto Yeast extract, 20 $g \cdot l^{-1}$ Bacto Peptone and 20 $g \cdot l^{-1}$ glucose). All *S. cerevisiae* strains used in this study belong to the CEN.PK family (Table S5) (*47-49*) and were grown at 30 ˚C. Cultures for transformation were grown in YPD medium. For galactose induction of *SCEI*, overnight cultures were transferred to and grown for 4 h on YPGal medium (containing 20 $g \cdot l^{-1}$ galactose instead of glucose) prior to transformation (*8*). Synthetic media (SM) contained, per liter of demineralized water, 5 g $(NH_4)_2SO_4$, 3 g KH_2PO_4, 0.5 g $MgSO_4 \cdot 7 \cdot H_2O$, and trace elements (*50*). The pH was set at 6.0 by 1M KOH and filter-sterilized vitamin solution (5) was added after autoclaving the medium at 120 °C for 20 min. Glucose was separately sterilized at 110 °C and added to a final concentration of 20 $g \cdot l^{-1}$. Uracil auxotrophic strains were grown on SM supplemented with 150 $mg \cdot L^{-1}$ uracil (*51*). SM without nitrogen source (SMwn) was prepared by replacing $(NH_4)_2SO_4$ with 6.6 $g \cdot l^{-1}$ K_2SO_4. SMU, which was used for growth-rate determinations, was prepared by supplementing SMwn with 2.3 $g \cdot l^{-1}$ urea (filter sterilized). Use of urea as the nitrogen source prevents the strong acidification which occurs in ammonium-based SM (*52*). For transformation experiments with the amdSYM marker cassette (*7*), 1.8 $g \cdot l^{-1}$ acetamide was added to SMwn. For counter-selection of the *URA3* and *KlURA3* marker gene, SMwn was supplemented with 3.53 $g \cdot l^{-1}$ proline, 0.010 $g \cdot l^{-1}$ uracil and 0.20 $g \cdot l^{-1}$ 5-fluoroorotic acid (5-FOA) (Sigma Aldrich, St. Louis, MO). For selection of transformants carrying the marker genes kanMX (*53*), natNT1 or hphNT1 (*54*), 200 $mg \cdot l^{-1}$ G418, 100 $mg \cdot l^{-1}$ nourseothricin or 200 $mg \cdot l^{-1}$ hygromycin (*55*), respectively, were added to complex media. Solid media were prepared by adding 2% (w/v) agar prior to autoclaving.

Plasmid propagation and isolation were performed with chemically competent *Escherichia coli* DH5α (Z-competent transformation kit; Zymo Research, Orange, CA) cultivated in lysogeny broth (LB) medium supplemented with 100 mg liter.1^{-1} ampicillin when required (*56, 57*).

Frozen glycerol stocks were prepared by addition of glycerol (30% v/v) to exponentially growing shake-flask cultures of *S. cerevisiae* and overnight cultures of *E. coli* and stored aseptically in 1 mL aliquots at –80˚C.

Molecular Biology techniques

PCR amplification for cloning purposes was performed with Phusion® Hot Start II High Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA). To improve PCR efficiency, conditions in the PCR reaction as recommended by the supplier were modified by decreasing the primer concentration from 500 nM to 200 nM and increasing the polymerase concentration from 0.02 U μl^{-1} to 0.03 U μl^{-1}. Diagnostic PCR was performed with the DreamTaq PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's recommendations. Template genomic DNA for amplification of glycoblocks was isolated from *S. cerevisiae* CEN.PK113-7D and *S. kudriavzevii* CR85 using the Qiagen 100/G kit (Qiagen, Hilden, Germany). Genomic DNA for, sequencing, PCR or restriction analysis was isolated with the YeaStar kit (Zymo Research, Irvine, CA). Plasmids maintained in *E. coli* DH5α were isolated with the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich). PCR products were separated in 1% (w/v) agarose (Sigma) gels in 1x TAE (40 mM Tris-acetate, pH 8.0 and 1 mM EDTA) or, when fragments were smaller than 500 bp, in 2% (w/v) agarose in 0.5x TBE (45 mM Tris-borate, pH 8.0, and 1 mM EDTA). Glycoblocks were isolated from gel using the Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine, CA). Prior to transformation, fragments were pooled, maintaining equimolar concentrations (150 fmol per fragment) with the DNA fragments containing the marker (*8*). Yeast transformation was performed with the LiAc/ssDNA method (*58*). Plasmids used in this study and primers used for their construction are described in Tables S6 and S7, respectively.

Construction of glycoblocks and marker cassettes

The glycolytic gene cassettes flanked by SHR-sequences (glycoblocks) were obtained by extension PCR. Genomic DNA of *S. cerevisiae* CEN.PK113-7D or *S. kudriavzevii* CR85 was used as PCR template for amplification of glycolytic genes, including their native promoter and terminator sequences. Promoters of *S. cerevisiae* glycolytic genes were tentatively defined as the 800-bp sequences upstream of their start codon. When these 800-bp sequence overlapped with another, upstream gene, the promoter size was shortened to eliminate overlap with the coding sequence of the neighbouring gene. Sequences of *S. kudriavzevii* CR85 glycolytic genes were kindly provided by Prof. Eladio Barrio (Universitat de València, València, Spain). DNA sequences of *S. cerevisiae* and *S. kudriavzevii* genes used to construct the glycoblocks and predicted sequences of the encoded proteins are shown in Table S2. For *S. kudriavzevii* genes, fragments of ca. 800 bp upstream of the genes were selected as promoter sequence. In all glycoblocks, terminator sequences comprised of the ca. 200 bp downstream of the respective stop codons. PCR primers used to construct the glycoblocks are described in Table S8 and

the length of *S. cerevisiae* and *S. kudriavzevii* promoters used in this study are reported in Table S9.

To add extra restriction sites for HO and I-CreI endonucleases, enabling later excision of the single locus glycolysis, the *PDC1* glycoblock was prepared differently. *PDC1* was obtained by PCR amplification from CEN.PK113-7D genomic DNA using primers PDC1 Fw+RES and PDC1 Rv+M (Table S8). The fragment SYN2 was obtained by fusion PCR of oligonucleotides Syn2 Fw and Syn2 Rv using primers FUS2 Fw and FUS2 Rv (Fig. S8). The resulting product was cloned in a pCR™4Blunt-TOPO® vector and verified by restriction/digestion, resulting in pUD336. The glycoblock PDC1-$SYN_{FM,}$ was obtained from pUD336, using primers FUS2 Fw and FUS2 Rv.

Coding sequences for *Homo sapiens* genes *TPI1* (muscle, splicing variant 1) and *PGK1* (muscle, splicing variant 1) were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/) (Table S3), codon optimized (*59*) and chemically synthesized and cloned in plasmids pSYN-TPI1 and pSYN-PGK1 (GeneArt, Life Technologies, Table S4). These plasmids were used as templates for PCR amplification of the codon-optimized ORFs with specific primers (Table S7). Promoters and terminators of the corresponding *S. cerevisiae* orthologues were amplified from CEN.PK113-7D genomic DNA using primers that added overlapping sequences to the codon-optimized human ORFs (Table S7). The promoter, ORF and terminator fragments were mixed in equimolar amounts, normalized to 100 ng of the ORF, and stitched by fusion-PCR. The resulting products were cloned in pCR™4Blunt-TOPO®vectors and verified by restriction/digestion, yielding pUD329 (p*TPI1-HsTPI1-tTPI1*) and pUD331 (p*PGK1-HsPGK1-tPGK1*). Plasmids pUD329 and pUD331 were used as templates for the human *TPI1* and *PGK1* glycoblocks, respectively. The amdSYM and kanMX marker cassettes were obtained by PCR with pUGamdSYM (*7*) and pUG6 (*53*) as templates, respectively, using specific oligonucleotide primers (Table S8). All cassettes were gel-purified prior to transformation and DNA concentrations were measured in a NanoDrop 2000 spectrophotometer (wavelenght 260 nm) (Thermo Fisher Scientific).

Construction of deletion cassettes and CRISPR/Cas9 plasmids

Native S. cerevisiae PYK1, PGI1, TPI1, TDH3 and *PGK1* genes were deleted using standard techniques and deletion cassettes were obtained as previously described (*9*) using the pDS-plasmid series (Table S6). Primers used for construction of deletion cassettes are given in Table S10. Cassettes were gel-purified and 500 ng of each cassette was used for yeast transformation. To enable CRISPR/Cas9 mediated genome editing, the gene encoding Cas9 (*10, 24*), driven by the constitutive *TEF1* promoter, was integrated in the genome of strain IMX511. Two fragments were constructed to replace the native locus

of the deleted *pfk2* gene with *cas9* (Fig. S4). A cassette containing *cas9* was obtained by PCR with p414-TEF1p-cas9-CYC1t (*10*) as template and primers CAS9 Fw+pfk2 and CAS9 Rv+link (Table S10). A second cassette containing the natNT1 marker gene was obtained by PCR on plasmid pUGnatNT1 with primers nat Fw+link and nat Rv+Rpt+pfk2 (Table S10). Both cassettes were gel purified and pooled in equimolar amounts. 500 ng of this mixture were used to transform strain IMX511, yielding strain IMX535 constitutively expressing Cas9. CRISPR-Cas9 editing was subsequently used to delete *HXK2*, *FBA1*, *ENO2*, *GPM1*, *PFK1*, *PDC1* and *ADH1 (24)* (Fig. S2 and Table S5). To rescue the double strand DNA break (DSB) introduced by Cas9, 120 bp marker-free deletion cassettes (repair fragments) were used. These dsDNA repair fragments were constructed by annealing two complementary single-stranded oligonucleotides listed in Table S10 (*24*). Expression cassettes for the guide RNAs (gRNAs) used to target Cas9 to *HXK2*, *FBA1*, *ENO2*, *GPM1* and *PFK1*, flanked by SHR-sequences (*8*), were chemically synthesized (GeneArt). Plasmids containing the synthesized gDNAs, as supplied by the manufacturer, were used as templates for construction of the gRNA expression cassettes including the SHR-sequences by PCR. Primers are given in (Table S7). To incorporate the gRNA cassettes in a yeast expression vector, p426-GPD (*9*) was linearized by PCR with primers adding SHR-sequences corresponding to the SHR-sequences of the gRNA cassettes (Table S7). The gRNA cassettes were assembled into the p426 backbone by Gibson assembly (New England Biolabs, Ipswich, MA) following the manufacturer's recommendations (Fig. S9a). Each plasmid contained a single gRNA. For each deletion 100 ng of the appropriate CRISPR-plasmid was co-transformed with 1.5 µg of the corresponding repair fragment. In its native locus, the *S. cerevisiae ENO2* gene is closely flanked by other genes. To avoid interference with the expression of these adjacent genes, deletion sites were chosen that were also present in the *ENO2* glycoblock. To prevent deletion of the *ENO2* gene on the single locus glycolysis, two different repair fragments were used (Fig. S10).

PDC1 and *ADH1* were simultaneously deleted using Cas9. The two plasmids carrying the gRNAs targeting *PDC1* and *ADH1* were constructed using *in vivo* assembly. Plasmid p426-SNR52p-gRNA.CAN1.Y-SUP4t (*10*) was linearized with primers p426-crispr Fw and p426-crispr Rv. The 120 bp targeting fragments (*crPDC1* and *crADH1*) were obtained by annealing complementary oligonucleotides as previously described (*24*) (Table S7). A mix consisting of 100 ng of the linearized CRISPR-backbone, 300 ng of *crPDC1*, 300 ng of *crADH1* and 1.8 µg of each of the appropriate repair fragments (obtained as described above, Table S10) was used for transformation.

Two additional CRISPR-plasmids, targeting the amdSYM cassette and the flanking regions of the single-locus glycolysis gene cluster, respectively, were constructed.

Plasmids were designed as previously described (*24*) (Fig. S9b). As described above, the linearized plasmid backbone was obtained with primers p426-crispr Fw and p426-crispr Rv from p426-SNR52p-gRNA.CAN1.Y-SUP4t (*10*) and the 120 bp targeting fragments (cramdSYM and crRECYCLE) were obtained by annealing complementing 120 bp oligo's (Table S7). The backbone and the desired targeting fragments were assembled into the CRISPR plasmids by Gibson assembly resulting in pUDE337 carrying cramdSYM and pUDE342 carrying crRECYCLE (Fig. S9b).

Construction of the SwYG strain

Fig. S2 provides an overview of the construction of the SwYG strain. A locus for chromosomal integration of the glycolytic gene cluster was prepared by introduction of an I-SceI restriction site at the *SGA1* locus on chromosome IX. Expression of *SGA1*, encoding a sporulation-specific glucoamylase, is induced by Ime1p in diploid cells during late sporulation but is repressed by Rme1p during vegetative growth of *S. cerevisiae (60, 61)*. The cassette carrying the I-SceI recognition site targeted to *SGA1* also carried the *SCEI* gene which encodes an intron-encoded homing endonuclease, under the control of the galactose inducible promoter *GAL1 (8)*, and the selection marker *KlURA3*. First the *SCEI/KlURA3* cassette was obtained by PCR using IMX221 genomic DNA as template (*8*) and the primers Tag G Fw and SGA1 Rv (Table S8, Fig. S11a). Fragment SYN1 (Fig. S11b), was obtained by mixing the oligonucleotides Syn1 Fw and Syn1 Rv. The resulting fragment SYN1 and the *SCEI/KlURA3* cassette were gel-purified and fused by fusion-PCR (*9*) using primers FUS1 Fw and FUS1 Rv (Table S8, Fig. S11b and S11c). The resulting product was cloned in a pCR™4Blunt-TOPO®vector (Invitrogen, Life Technologies), resulting in pUD335, which was checked by restriction analysis. The *KlURA3-SCEI* cassette was obtained by PCR from pUD335 using primers FUS1 Fw and FUS1 Rv (Table S8). *S. cerevisiae* strain IMX370 (*3*), which carries a minimal set of 13 glycolytic genes, was transformed with 100 ng of the *KlURA3-SCEI* cassette (Fig. S1a), resulting in IMX377. IMX377 also harbored additional restriction sites, recognized by the HO and I-CreI endonucleases, in its integrated *KlURA3-SCEI* cassette, and homologous flanking regions to promote recombination upon excision of the endogenous Single Locus Glycolysis (SinLoG) cassette (Fig. S1a).

The endogenous SinLoG cassette was assembled and integrated in IMX377 using the Combined *in vivo* Assembly and Targeted chromosomal Integration (CATI) approach (*8*). IMX377 was transformed with a mix consisting of the *S. cerevisiae* glycoblocks and the *amdS* marker cassette (*FBA1*$_{GH}$, *TPI1*$_{HP}$, *PGK1*$_{PQ}$, *ADH1*$_{QN}$, *PYK1*$_{NO}$, *TDH3*$_{OA}$, amdSYM$_{AB}$, *HXK2*$_{BC}$, *PGI*$_{CD}$, *PFK1*$_{DJ}$, *PFK2*$_{JK}$, *ENO2*$_{KL}$, *GPM1*$_{LM}$, *PDC1-SYN*$_{MF}$) (subscript letters indicate the SHR-sequences, Fig. S1b). The molar ratio of transformed fragments was 1:1 normalized to 150 ng of the amdSYM$_{AB}$ cassette. Transformants were selected on

medium containing acetamide as sole nitrogen source. Clones were analysed for presence of all junctions between glycoblocks and selection markers with primers given in (Table S11). One of the colonies that showed correct PCR patterns was selected and named IMX382. This strain was further analyzed by sequencing a set of 14 PCR products obtained with primer pairs 1 to 14 (Table S11 and Fig. S1d). All PCR products were pooled in a molar ratio of 1:1. From this set of 14 products a library of 300 bp insert was constructed and paired end sequenced (100bp paired end reads) using an Illumina HiSeq 2500 sequencer (BaseClear, Leiden, The Netherlands). Sequence reads were mapped onto the glycolytic genes cluster using Burrows-Wheeler Alignment tool (using "BWA mem" command; version 0.7.10-r789) and the resulting Alignment file (BAM file) was further processed by Pilon (version 1.10; using "--vcf --fix all,breaks" parameter (*62*)) for variant detection which were stored in VCF (Variant Call Format) file.

To construct the SwYG strain, the 13 genes made redundant by the newly added glycolytic cluster were removed from IMX382 in the following order: *PYK1, PGI1, TPI1, TDH3, PFK2, PGK1, GPM1, FBA1, HXK2, PFK1, ADH1, PDC1, ENO2* (Fig. S2). The natNT1, kanMX and hphNT1 marker cassettes were used for the deletion of *PYK1, PGI1* and *TPI1* respectively. These marker cassettes were excised using I-SceI as previously described (*9*) by transforming strain IMX493 with plasmid pUDC073 carrying *SCEI* (Fig. S3). *PFK2* was deleted by a cassette containing *cas9* and the natNT1 marker cassette (Fig. S4). The *KlURA3* and kanMX markers used for the subsequent deletion of *TDH3* and *PGK1* were recycled by the same I-SceI facilitated marker removal, by transforming the *SCEI* expressing plasmid pUDE206 to IMX557. Deletion of *GPM1, FBA1, HXK2* and *PFK1* was performed by the CRISPR/Cas9 system by transforming the appropriate CRISPR-plasmid and accompanying repair fragment. *ADH1* and *PDC1* were simultaneously deleted using the CRISPR/Cas9 cloning-free deletion method (*24*). Transformants were selected on SM and the CRISPR-plasmids were recycled by growing transformants overnight on YPD medium followed by plating on SM medium with 5-FOA. In order to restore a functional *ENO2* glycoblock to the single locus glycolysis, a glycoblock containing *ENO2* with a longer promoter sequence (*ENO2-LONG*$_{AB}$) was introduced to the SinLoG-IX cluster by replacing the *amdSYM*$_{AB}$ marker cassette in IMX583 resulting in strain IMX586 (Fig. S10). This was achieved by transforming CRISPR-plasmid pUDE337 together with the *ENO2-LONG*$_{AB}$ glycoblock. Transformants were selected on SM. Subsequently, the endogenous *ENO2* gene could be deleted by co-transforming the CRISPR-plasmid pUDE326 and the corresponding repair fragments in IMX586 resulting in IMX587. Transformants were selected on SM. Finally the dysfunctional glycoblock ENO2$_{KL}$ was replaced by transforming 500 ng of marker cassette amdSYM$_{KL}$.

Transformants were selected on SMwn with acetamide and one clone displaying the correct PCR profile was plated on medium with 5-FOA to recycle the CRISPR-plasmid and stocked as IMX589 (auxotrophic SwYG strain). To be able to perform growth experiments on SM, the uracil auxotrophy was repaired by transforming IMX589 with pUDE325, resulting in the prototrophic SwYG strain IMX606.

Construction of glycolytic gene clusters in the *CAN1* locus

The SinLoG clusters introduced in the *CAN1* locus were obtained by transforming IMX589 with a mix of glycoblocks for assembly and targeted integration of the desired glycolytic design. To facilitate the targeted integration into the genome, a similar approach to the CATI approach was chosen, but the CRISPR/Cas9 system was used instead of I-SceI to promote the formation of a double strand break and therefore integration of the SinLoG clusters at the targeted locus. Therefore 300 ng of p426-SNR52p-gRNA.CAN1.Y-SUP4t plasmid coding for the gRNA targeting the *CAN1* locus (*10*) was co-transformed with the glycoblocks and with a cassette carrying the kanMX selection marker. The glycoblocks for the native SinLoGv were: $FBA1_{can1H}$, $TPI1_{HP}$, $PGK1_{PQ}$, $ADH1_{QN}$, $PYK1_{NO}$, $TDH3_{OA}$, $ENO2_{AB}$, $HXK2_{BC}$, PGI_{CD}, $PFK1_{DJ}$, $PFK2_{JK}$, $GPM1_{LM}$, $PDC1_{Mcan1}$. For the *S.k.* SinLoG were used: $skFBA1_{can1H}$, $skTPI1_{HP}$, $skPGK1_{PQ}$, $skADH1_{QN}$, $skPYK1_{NO}$, $skTDH1_{OA}$, $skHXK2_{BC}$, $skPGI_{CD}$, $skPFK1_{DJ}$, $skPFK2_{JK}$, $skENO2_{KL}$, $skGPM1_{LM}$, $skPDC1_{Mcan1}$. For the mosaic SinLoG the following mixture was transformed: $FBA1_{can1H}$, $hsTPI1_{HP}$, $hsPGK1_{PQ}$, $skADH1_{QN}$, $PYK1_{NO}$, $TDH3_{OA}$, $skHXK2_{BC}$, $skPGI_{CD}$, $skPFK1_{DJ}$, $skPFK2_{JK}$, $skENO2_{KL}$, $skGPM1_{LM}$, $PDC1_{Mcan1}$. Cassettes were mixed in a 1:1 molar ratio normalized to 140 ng of the kanMX cassette. Selection was on SM for presence of the CRISPR-plasmid, which contained the *URA3* marker. For each transformation eight clones were plated to medium selective for kanMX. Resistant clones were analyzed by PCR for presence of the full SinLoG clusters with primers given in Table S12. For each glycolytic variant, a correctly assembled strain was grown on complex medium and plated on SM proline with 5-FOA and uracil to recycle the CRISPR-plasmid. The resulting strains were stocked on SM acetamide supplemented with uracil (IMX591, IMX607, IMX633).

Excision of the native SinLoG cassette from chromosome IX

The native SinLoG, integrated at the *SGA1* locus, was removed from strains IMX591, IMX607 and IMX633 using CRISPR-Cas9. To this end, 100 ng of the CRISPR-plasmid pUDE342 was transformed into these strains, together with 1.5 µg of the recycle repair fragment (Table S10 and Fig. S12). Transformants were selection on SM glucose plates and analyzed for correct removal of the endogenous SinLoG by PCR with primers SGA1 Fw and SGA1 Rv (Table S12).

skTDH1 overexpression in IMX637

A plasmid backbone, PCR amplified with primers p426-rv+O and p426-fw+A (Table S5) and plasmid p426-GPD as the template, and the *skTDH1* glycoblock were assembled *in vitro* using Gibson assembly (New England Biolabs, Beverly, MA, USA), resulting in the plasmid pUDESkTDH1 (Table S4). IMX637 was plated on complex medium with 5-FOA to counterselect the pUDE342 plasmid. A selected colony was then transformed with 100 ng of the pUDEskTDH1 plasmid and transformants were selected on SM. One transformant was stocked as IMX652.

Sequencing

Genomic DNA of strains IMX589 (auxotrophic SwYG strain), IMX605 (endogenous SinLoG on chromosome V), IMX637 (*Sk*SinLoG) and IMX645 (mosaic SinLoG) was sequenced. Ilumina Nextera libraries (300-bp insert size) were constructed and paired-end sequenced (100 bp reads) using an Illumina HISeq 2500 sequencer at Baseclear BV (Leiden, The Netherlands). A minimum quantity of 750 Mb was generated, representing a minimum 60-fold coverage. Genome sequences were *de novo* assembled using the gsAssembler (version 2.6) software package, also known as the Newbler software package (454 Life Sciences, Branford, CT). To verify deletions in IMX589, all contigs were mapped to the *in silico* design after gene removal using Clustal X in Clone Manager 9 (Sci-Ed Software, Cary, NC). The sequences are accessible at NCBI (http://www.ncbi.nlm.nih.gov/) under the bioproject number PRJNA317665.

To verify correct integration of the different glycolytic gene clusters, contigs were mapped to the *in silico* design. To exclude possible duplications of glycolytic genes, a copy number variation analysis was performed with the Magnolya algorithm (*63*) (Fig. S6).

To identify any unintended changes at the nucleotide level, all sequence libraries of samples IMX372 (*3*), IMX589, IMX605, IMX637 and IMX645 were processed by an in-house pipeline hosted in Galaxy (https://galaxyproject.org/). Sequence data were mapped to the genome of *S. cerevisiae* CEN.PK113-7D (*47*) for whole genome comparison, as well as to the *in silico* design of the SinLoG present each sequenced strain. The Burrows-Wheeler Alignment tool (BWA, version 0.7.10-r789) was used and the resulting binary alignment file (BAM file) was further processed using SAMtoolsmpileup (version 0.1.18) and bcftools (from the SAMtools package) to compute the genotype likelihood and stores these likelihoods in Binary variant call format (BCF). The script vcfutils.pl was used, with parameter varFilter and maximum read depth 400, to filter and convert to variant call format (VCF). The resulting VCF files were annotated and effects of variants on genes were predicted by the snpEff package

(version 3.4). To compare IMX589 to IMX372, the called and annotated variants in both IMX589 and IMX372 samples were subtracted from sample IMX589 with the "subtract whole dataset from another dataset" tool in Galaxy. The same procedure was followed to compare IMX605, IMX637 and IMX645 to IMX589.

Determination of specific growth rates in shake-flask cultures

Glycerol stocks from strains IMX372 (MG), IMX606 (prototrophic SwYG strain), IMX605 (endogenous SinLoGv), IMX637 (*Sk*SinLoG), IMX652 (*Sk*SinLoG with *SkTDH1* overexpression), IMX645 (mosaic SinLoG) were inoculated in 100 ml SM urea + 2% glucose (w/v) in 500 ml shake flasks and grown to late exponential phase. Cells were harvested and immediately transferred to pre-warmed 500 ml flasks containing the same medium at an OD660 of 0.2. Biomass formation was followed by measuring OD660. Concentrations of extracellular metabolites in culture supernatants were measured by HPLC using a Aminex HPX-87H ion exchange column operated at 60 ˚C with 5 mM H_2SO_4 as mobile phase at an isocratic flow rate of 0.6 ml·min^{-1}. Data reported in the results section are calculated based on at least two independent culture replicates.

Determination of *in vitro* enzyme activities

Cell extracts were prepared as previously described (*64*) from culture samples (ca. 60 mg biomass dry weight) taken from mid-exponential phase shake-flask cultures. Spectrophotometric assays of glycolytic enzyme activities were done as previously described (*65*), except for phosphofructokinase, whose activity was determined as described by Cruz and co-workers (*66*). Enzyme activities are expressed as µmol substrate converted (mg protein)$^{-1}$ h^{-1}. Protein concentrations in the cell extracts were determined as described by Lowry and co-workers (*67*) with bovine serum albumin as a standard. Reported enzyme activities are based on measurements on at least two independent culture replicates, with at least two analytical replicates for each assay.

Quantitative physiological analyses of SwYG strains in aerobic batch cultures

For analysis of quantitative physiology, SwYG strains IMX605, IMX606, IMX645 and IMX652 were grown in duplicate aerobic batch cultures in bioreactors. SM was supplemented with 20 g.L^{-1} glucose as sole carbon-source and 0.2 g.L^{-1} antifoam Emulsion C (Sigma, St. Louise, USA). Batch cultures were inoculated at an initial OD660 of 0.3 with in water resuspended cells obtained from exponentially growing cultures on identical medium. Aerobic batch cultures were performed at a working volume of 1.4 L in 2 L bioreactors (Applikon, Schiedam, The Netherlands). Culture conditions were a temperature of 30 °C, constant agitation at 800 rpm, sparging of 700 mL.min^{-1} dried,

compressed air (Linde Gas Benelux, The Netherlands) and a pH of 5.0, maintained by automatic addition of 2 M KOH and 2 M H_2SO_4.

Biomass concentrations as culture dry biomass were measured by filtering samples of 10 mL culture through pre-dried filters (pore-size 0.45um, Whatman / GE Healthcare Life Sciences, United Kingdom) and drying in a microwave oven at 360W for 20 minutes, as adapted from (*64*). Additionally biomass concentrations were more frequently determined by measuring the optical density at 660 nm (OD660). Biomass concentrations measured as OD660 and culture dry weight were linearly correlated (coefficients of determination were at least 0.997). Based on these experimentally determined linear correlations, dry biomass concentrations were calculated using OD660 measurements and were used to estimate specific growth rates by simple linear regression of the natural logarithm of dry biomass concentrations as function of culture age.

Extracellular glucose, ethanol, glycerol and acetate concentrations were determined by high-performance liquid chromatography (HPLC) analysis of culture supernatants obtained by centrifugation of samples (3 min. at 20.000 g). HPLC analysis was performed using a Agilent HPLC equipped with a Bio-Rad Aminex HPX-87H column at 60 °C and 5 mM H_2SO_4 as a mobile phase at a flow rate of 0.6 $mL \cdot min^{-1}$ using, coupled to a UV and RI detector (Agilent, Santa Clara, USA).

CO_2 and O_2 concentrations in the exhaust gas were analysed using a Rosemount NGA 2000 analyser (Baar, Switzerland), after cooling by means of a condenser (2 °C) and drying using a PermaPure Dryer (model MD 110-8P-4; Inacom Instruments, Veenendaal, the Netherlands) of the gas. Independent culture duplicates were performed for each tested strain. Carbon balances for all cultures closed within 5%. Previously published data were used for the MG strain (*3*). The aerobic batch cultures of the MG strain were performed in conditions scrupulously identical to those described above.

Additional Material

Supplemental Figures and Table S4, S6, S7, S8, S10, S11 and S12 can be found at https://doi.org/10.1073/pnas.1606701113

Supporting Table S1 – Amino acid substitutions identified in the proteins encoded by the SinLoG genes of the constructed strains as compared to the *in silico* design.

Systematic name	Name	Type	Amino acid change
SwYG strain (IMX589)			
YOL086C	*ADH1*	Synonymous	A180A
SinLoG-V strain (IMX605)			
YOL086C	*ADH1*	Non-synonymous	R212G
YGR240C	*PFK1*	Non-synonymous	T118A
***Sk*-SinLoG-V strain (IMX637)**			
No syst. name	*SkPYK1*	Synonymous	A167G
Mosaic-SinLoG-V strain (IMX645)			
None detected			

Supporting Table S2 – Amino acid substitutions identified in the constructed strains as compared to the most relevant parental strains.

Systematic name	Name	Type	Amino acid change
SwYG strain (IMX589) vs. Minimal Glycolysis strain (IMX372)			
YBR079W	*VPS15*	Non-synonymous	E474K
YJL212C	*OPT1*	Non-synonymous	I463T
YNL245C	*CWC25*	Non-synonymous	P62L
YDL079C	*MRK1*	Synonymous	I190I
YLR180W	*SAM1*	Synonymous	V217V
YNL262W	*POL2*	Synonymous	F1536F
SinLoG-V strain (IMX605) vs SwYGstrain (IMX589)			
YNL215W	*IES2*	Non-synonymous	E160G
***Sk*-SinLoG-V strain (IMX637) vs SwYG strain (IMX589)**			
YGL195W	*GCN1*	Non-synonymous	G427C
Mosaic-SinLoG-V strain (IMX645) vs SwYG strain (IMX589)			
YDR539W	*FDC1*	Non-synonymous	P117S

Supporting Table S3 – Comparison between *S. cerevisiae* and *S. kudriavzevii or H. sapiens* of the DNA and protein sequence of the glycolytic genes used in the SinLoG gene cluster.

Prefix *Sc*, *Sk* and *Hs* indicate the gene origin, i.e. *S. cerevisiae*, *S. kudriavzevii* and human respectively. *S. cerevisiae* CEN.PK 113-7D sequences were compared to *S. kudriavzevii* IFO1802 (http://sss.genetics.wisc.edu/cgi-bin/s3.cgi) and *H. sapiens* (http://www.ncbi.nlm.nih.gov/, accession number NP_000356.1 for *HsTP1* and NP_000282.1 for *HsPGK1*) sequences by BLASTN and BLASTX analysis .

Gene		% sequence identity	
		Gene	Protein
S. kudravzevii vs S. cerevisiae			
ScHXK2	*SkHXK2*	90	96
ScPGI1	*SkPGI1*	91	98
ScPFK1	*SkPFK1*	89	98
ScPFK2	*SkPFK2*	90	98
ScFBA1	*SkFBA1*	95	95
ScTPI1	*SkTPI1*	95	97
ScTDH3	*SkTDH1*	88	89
ScPGK1	*SkPGK1*	97	99
ScGPM1	*SkGPM1*	96	97
ScENO2	*SkENO2*	97	98
ScPYK1	*SkPYK1*	95	97
ScPDC1	*SkPDC1*	95	98
ScADH1	*SkADH1*	95	96
H. sapiens vs S. cerevisiae			
ScTPI1	*HsTP1*	- [a]	53
ScPGK1	*HsPGK1*	- [a]	66

[a] The DNA sequence was not compared because the human gene was codon-optimized.

Table S5 – Strains used in this study.

Strain	Relevant genotype	Source
IMX372 (Minimal Glycolysis, MG)	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1::KlURA3 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4*	*(3)*
IMX221	*MATa ura3-52 MAL2-8c SUC2 spr3::(TagG-KlURA3- P_{GAL1}-SCEI-T_{cyc1}-TagF)*	*(8)* (*68*)
S. kudriavzevii CR85	*Wild isolate*	*(11)*
IMX370	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4*	*(3)*
IMX377	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(TagG-KlURA3- P_{GAL1}-SCEI-T_{cyc1}-TagF)*	This study
IMX382	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$)*	This study
IMX457	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$), pyk1::*natNT1	This study
IMX492	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc, pdc, adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1::*NatNT1 *pgi1::*kanMX	This study
IMX493	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$), pyk1::*NatNT1 *pgi1::*kanMX *tpi1::*hphNT1	This study
IMX509	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1::*NatNT1, *pgi1::*kanMX, *tpi1::*hphNT1 pUDC073*(CEN6/ARS4 ori URA3 GAL1pr-SCEI-CYC1ter)*	This study
IMX510	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1*	This study
IMX511	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2, tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1$_{MF}$) pyk1 pgi1 tpi1 tdh3::*kanMX	This study

IMX535	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3::*kanMX *pfk2::(pTEF-cas9-tCYC1* natNT1*)*	This study
IMX557	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3::*kanMX *pfk2::(pTEF-cas9-tCYC1*natNT1*) pgk1::KlURA3*	This study
IMX561	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1*	This study
IMX566	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1*	This study
IMX568	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFcas9-tCYC1* natNT1*) pgk1 gpm1 fba1*	This study
IMX570	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2*	This study
IMX571	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1*	This study
IMX583	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$* amdSYM$_{AB}$ *HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1*	This study
IMX586	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5, hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1*	This study
IMX587	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6, adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ ENO2$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*	This study
IMX589 (auxotrophic SwYG)	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*	This study
IMX606	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$*	This study

(prototrophic SwYG)	*AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$)pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2* pUDE325	
IMX591	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(FBA1$_{can1H}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ KanMX$_{KL}$ GPM1$_{LM}$ PDC1$_{Mcan1}$)*	This study
IMX607	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(SkFBA1$_{can1H}$ SkTPI1$_{HP}$ SkPGK1$_{PQ}$ SkADH1$_{QN}$ SkPYK1$_{NO}$ SkTDH1$_{OA}$ KanMX$_{AB}$ SkHXK2$_{BC}$ SkPGI1$_{CD}$ SkPFK1$_{DJ}$ SkPFK2$_{JK}$ SkENO2$_{KL}$ SkGPM1$_{LM}$ SkPDC1$_{Mcan1}$)*	This study
IMX633	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(FBA1$_{can1H}$ pTPI1-HsTPI1-tTPI1$_{HP}$ pPGK1-HsPGK1-tPGK1$_{PQ}$, SkADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ KanMX$_{AB}$ HXK2$_{BC}$ SkPGI1$_{CD}$ SkPFK1$_{DJ}$ SkPFK2$_{JK}$ SkENO2$_{KL}$ SkGPM1$_{LM}$ PDC1$_{Mcan1}$)*	This study
IMX605	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(FBA1$_{can1H}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ KanMX$_{KL}$ GPM1$_{LM}$ PDC1$_{Mcan1}$)* pUDE342	This study
IMX637	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*), pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(SkFBA1$_{can1H}$ SkTPI1$_{HP}$ SkPGK1$_{PQ}$ SkADH1$_{QN}$ SkPYK1$_{NO}$ SkTDH1$_{OA}$ KanMX$_{AB}$ SkHXK2$_{BC}$ SkPGI1$_{CD}$ SkPFK1$_{DJ}$ SkPFK2$_{JK}$ SkENO2$_{KL}$ SkGPM1$_{LM}$ SkPDC1$_{Mcan1}$) pUDE342*	This study
IMX645	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2, can1::(FBA1$_{can1H}$ pTPI1-HsTPI1-tTPI1$_{HP}$ pPGK1-HsPGK1-tPGK1$_{PQ}$, SkADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ KanMX$_{AB}$ HXK2$_{BC}$ SkPGI1$_{CD}$ SkPFK1$_{DJ}$ SkPFK2$_{JK}$ SkENO2$_{KL}$ SkGPM1$_{LM}$ PDC1$_{Mcan1}$)* pUDE342	This study
IMX652	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1* natNT1*) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(SkFBA1$_{can1H}$ SkTPI1$_{HP}$ SkPGK1$_{PQ}$ SkADH1$_{QN}$ SkPYK1$_{NO}$ SkTDH1$_{OA}$ KanMX$_{AB}$ SkHXK2$_{BC}$ SkPGI1$_{CD}$ SkPFK1$_{DJ}$ SkPFK2$_{JK}$ SkENO2$_{KL}$ SkGPM1$_{LM}$ SkPDC1$_{Mcan1}$)* pUDESkTDH1	This study

Supporting Table S9 – Length of *S. cerevisiae* and *S. kudriavzevii* promoters used in this study.

Gene	*S. cerevisiae* (bp)	*S. kudriavzevii* (bp)
FBA1	517	808
TPI1	513	819
PGK1	667	769
ADH1	964	992
PYK1	860	1099
TDH1		948
TDH3	632	
ENO2	(411[a]) 1012	840
HXK2	479	843
PGI1	697	850
PFK1	904	1028
PFK2	804	1026
GPM1	431	769
PDC1	864	908

[a] This promoter size did not result in functional expression of *ENO2*.

References

1. A. L. Meadows *et al.*, Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* **537**, 694-697 (2016).
2. R. Wieczorke *et al.*, Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.* **464**, 123-128 (1999).
3. D. Solis-Escalante *et al.*, A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell* **14**, 804-816 (2015).
4. L. Serrano, Synthetic biology: promises and challenges. *Mol. Syst. Biol.* **3**, (2007).
5. D. G. Gibson *et al.*, One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proceedings of the National Academy of Sciences* **105**, 20404-20409 (2008).
6. N. G. Kuijpers *et al.*, 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 (2013).
7. D. Solis-Escalante *et al.*, *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **13**, 126-139 (2013).
8. N. G. Kuijpers *et al.*, One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **13**, 769-781 (2013).
9. D. Solis-Escalante *et al.*, Efficient simultaneous excision of multiple selectable marker cassettes using I-SceI-induced double-strand DNA breaks in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **14**, 741-754 (2014).
10. J. E. DiCarlo *et al.*, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336-4343 (2013).
11. C. A. Lopes, E. Barrio, A. Querol, Natural hybrids of *S. cerevisiae x S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*. *FEMS Yeast Res.* **10**, 412-421 (2010).
12. F. N. Arroyo-López, S. Orlić, A. Querol, E. Barrio, Effects of temperature, pH and sugar concentration on the growth parameters of *Saccharomyces cerevisiae, S. kudriavzevii* and their interspecific hybrid. *Int. J. Food Microbiol.* **131**, 120-127 (2009).
13. D. R. Scannell *et al.*, The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces* sensu stricto genus. *G3: Genes, Genomes, Genetics* **1**, 11-25 (2011).
14. J. Tronchoni, V. Medina, J. M. Guillamón, A. Querol, R. Pérez-Torrado, Transcriptomics of cryophilic *Saccharomyces kudriavzevii* reveals the key role of gene translation efficiency in cold stress adaptations. *BMC Genomics* **15**, 432 (2014).
15. P. Daran-Lapujade *et al.*, The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proceedings of the National Academy of Sciences* **104**, 15753-15758 (2007).
16. B. M. Oliveira, E. Barrio, A. Querol, R. Perez-Torrado, Enhanced enzymatic activity of glycerol-3-phosphate dehydrogenase from the cryophilic *Saccharomyces kudriavzevii*. *PLoS One* **9**, (2014).
17. E. T. Young *et al.*, Evolution of a glucose-regulated *ADH* gene in the genus *Saccharomyces*. *Gene* **245**, 299-309 (2000).
18. N.-M. Grüning, D. Du, M. A. Keller, B. F. Luisi, M. Ralser, Inhibition of triosephosphate isomerase by phosphoenolpyruvate in the feedback-regulation of glycolysis. *Open biology* **4**, 130232 (2014).
19. A. H. Kachroo *et al.*, Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **348**, 921-925 (2015).
20. L. E. Maquat, R. Chilcote, P. M. Ryan, Human triosephosphate isomerase cDNA and protein structure. Studies of triosephosphate isomerase deficiency in man. *J. Biol. Chem.* **260**, 3748-3753 (1985).
21. T. L. Orr-Weaver, J. W. Szostak, R. J. Rothstein, Yeast transformation: a model system for the study of recombination. *Proceedings of the National Academy of Sciences* **78**, 6354-6358 (1981).
22. J. M. Daley, P. L. Palmbos, D. Wu, T. E. Wilson, Nonhomologous end joining in yeast. *Annu. Rev. Genet.* **39**, 431-451 (2005).
23. J. M. Daley, P. Sung, 53BP1, BRCA1, and the choice between recombination and end joining at DNA double-strand breaks. *Mol. Cell. Biol.* **34**, 1380-1388 (2014).
24. R. Mans *et al.*, CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **15**, (2015).

25. D. B. Flagfeldt, V. Siewers, L. Huang, J. Nielsen, Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* **26**, 545-551 (2009).
26. L. A. Mitchell, J. D. Boeke, Circular permutation of a synthetic eukaryotic chromosome with the telomerator. *Proceedings of the National Academy of Sciences* **111**, 17003-17010 (2014).
27. J. J. Wyrick *et al.*, Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **402**, 418-421 (1999).
28. V. R. Iyer, Nucleosome positioning: bringing order to the eukaryotic genome. *Trends Cell Biol.* **22**, 250-256 (2012).
29. J. Heinisch, R. Rodicio, Fructose-1, 6-bisphophate aldolase, triosephosphate isomerase, glyceraldhyde-3-phosphate dehydrogenases, and phosphoglycerate mutase. *Yeast Sugar Metabolism*, 119-140 (1997).
30. S. Ide, T. Miyazaki, H. Maki, T. Kobayashi, Abundance of ribosomal RNA gene copies maintains genome integrity. *Science* **327**, 693-696 (2010).
31. C. Bausch *et al.*, Transcription alters chromosomal locations of cohesin in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 8522-8532 (2007).
32. M. K. Dhar, S. Sehgal, S. Kaul, Structure, replication efficiency and fragility of yeast ARS elements. *Res. Microbiol.* **163**, 243-253 (2012).
33. C. C. Siow, S. R. Nieduszynska, C. A. Müller, C. A. Nieduszynski, OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* **40**, D682-D686 (2012).
34. N. Annaluru *et al.*, Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58 (2014).
35. D. Williamson, The timing of deoxyribonucleic acid synthesis in the cell cycle of *Saccharomyces cerevisiae*. *The Journal of cell biology* **25**, 517-528 (1965).
36. C. Gancedo, C. L. Flores, Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.* **72**, 197-210 (2008).
37. M. Lu, Y. Y. Sautin, L. S. Holliday, S. L. Gluck, The glycolytic enzyme aldolase mediates assembly, expression, and activity of vacuolar H+-ATPase. *J. Biol. Chem.* **279**, 8732-8739 (2004).
38. P. M. Kane, The where, when, and how of organelle acidification by the yeast vacuolar H+-ATPase. *Microbiol. Mol. Biol. Rev.* **70**, 177-191 (2006).
39. J. A. Diderich, L. M. Raamsdonk, A. L. Kruckeberg, J. A. Berden, K. Van Dam, Physiological properties of *Saccharomyces cerevisiae* from which hexokinase II has been deleted. *Appl. Environ. Microbiol.* **67**, 1587-1593 (2001).
40. P. Gonçalves, E. Valério, C. Correia, J. M. de Almeida, J. P. Sampaio, Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One* **6**, e20739 (2011).
41. D. Botstein, S. A. Chervitz, M. Cherry, Yeast as a model organism. *Science* **277**, 1259-1260 (1997).
42. F. Foury, Human genetic diseases: a cross-talk between man and yeast. *Gene* **195**, 1-10 (1997).
43. J. M. Laurent, J. H. Young, A. H. Kachroo, E. M. Marcotte, Efforts to make and apply humanized yeast. *Briefings in functional genomics* **15**, 155-163 (2016).
44. S. Lindquist, E. Craig, The heat-shock proteins. *Annu. Rev. Genet.* **22**, 631-677 (1988).
45. S.-K. Park *et al.*, Development and validation of a yeast high-throughput screen for inhibitors of Aβ42 oligomerization. *Disease models & mechanisms* **4**, 822-831 (2011).
46. T. Vavouri, J. I. Semple, B. Lehner, Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends Genet.* **24**, 485-488 (2008).
47. J. F. Nijkamp *et al.*, De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* **11**, 36 (2012).
48. K. D. Entian, P. Kötter, in *Yeast Gene Analysis,* I. Stansfield, M. J. R. Stark, Eds. (Academic Press, Elsevier, Amsterdam, 2007), vol. 36, chap. 25, pp. 629-666.
49. J. Van Dijken *et al.*, An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* **26**, 706-714 (2000).
50. C. Verduyn, E. Postma, W. A. Scheffers, J. P. Van Dijken, 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 (1992).
51. J. T. Pronk, Auxotrophic yeast strains in fundamental and applied research. *Appl. Environ. Microbiol.* **68**, 2095-2100 (2002).

52. M. Hensing, R. Rouwenhorst, J. Heijnen, J. Van Dijken, J. Pronk, Physiological and technological aspects of large-scale heterologous-protein production with yeasts. *Antonie Van Leeuwenhoek* **67**, 261-279 (1995).
53. U. Güldener, S. Heck, T. Fiedler, J. Beinhauer, J. H. Hegemann, A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24**, 2519-2524 (1996).
54. A. L. Goldstein, J. H. McCusker, Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**, 1541-1553 (1999).
55. L. Gritz, J. Davies, Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**, 179-188 (1983).
56. G. Bertani, Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* **186**, 595-600 (2004).
57. G. Bertani, Studies on lysogenis I.: The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293-300 (1951).
58. R. D. Gietz, R. A. Woods, Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87-96 (2002).
59. D. de Ridder, J. de Ridder, M. J. Reinders, Pattern recognition in bioinformatics. *Briefings in bioinformatics* **14**, 633-647 (2013).
60. K. Kihara, M. Nakamura, R. Akada, I. Yamashita, Positive and negative elements upstream of the meiosis-specific glucoamylase gene in *Saccharomyces cerevisiae*. *Molecular and General Genetics MGG* **226**, 383-392 (1991).
61. I. Yamashita, S. Fukui, Transcriptional control of the sporulation-specific glucoamylase gene in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**, 3069-3073 (1985).
62. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* **9**, e112963 (2014).
63. J. F. Nijkamp *et al.*, *De novo* detection of copy number variation by co-assembly. *Bioinformatics* **28**, 3195-3202 (2012).
64. E. Postma, C. Verduyn, W. A. Scheffers, J. P. Van Dijken, Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **55**, 468-477 (1989).
65. M. L. Jansen *et al.*, Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity. *Microbiology* **151**, 1657-1669 (2005).
66. L. A. Cruz *et al.*, Similar temperature dependencies of glycolytic enzymes: an evolutionary adaptation to temperature dynamics? *BMC systems biology* **6**, 151 (2012).
67. O. H. Lowry, H. J. Rosebrough, A. L. Farr, R. J. Randall, Protein measurement with the folin phenol reagent. *J. Biol. Chem* **193**, 265-275 (1951).

Chapter 3

The genetic makeup and expression of the glycolytic and fermentative pathways are highly conserved within the *Saccharomyces* genus

Francine J. Boonekamp, Sofia Dashko, Marcel van den Broek, Thies Gehrmann, Jean-Marc Daran, Pascale Daran-Lapujade

Published in *Frontiers in genetics* **9**, 504 (2018)

Abstract

The ability of the yeast *Saccharomyces cerevisiae* to convert glucose, even in the presence of oxygen, via glycolysis and the fermentative pathway to ethanol has played an important role in its domestication. Despite the extensive knowledge on these pathways in *S. cerevisiae,* relatively little is known about their genetic makeup in other industrially-relevant *Saccharomyces* yeast species. In this study we explore the diversity of the glycolytic and fermentative pathways within the *Saccharomyces* genus using *S. cerevisiae*, *S. kudriavzevii* and *S. eubayanus* as paradigms. Sequencing data revealed a highly conserved genetic makeup of the glycolytic and fermentative pathways in the three species in terms of number of paralogous genes. Although promoter regions were less conserved between the three species as compared to coding sequences, binding sites for Rap1, Gcr1 and Abf1, main transcriptional regulators of glycolytic and fermentative genes, were highly conserved. Transcriptome profiling of these three strains grown in aerobic batch cultivation in chemically defined medium with glucose as carbon source, revealed a remarkably similar expression of the glycolytic and fermentative genes across species, and the conserved classification of genes into major and minor paralogs. Furthermore, transplantation of the promoters of major paralogs of *S. kudriavzevii* and *S. eubayanus* into *S. cerevisiae* demonstrated not only the transferability of these promoters, but also the similarity of their strength and response to various environmental stimuli. The relatively low homology of *S. kudriavzevii* and *S. eubayanus* promoters to their *S. cerevisiae* relatives makes them very attractive alternatives for strain construction in *S. cerevisiae,* thereby expanding the *S. cerevisiae* molecular toolbox.

Introduction

The yeast *Saccharomyces cerevisiae* is known for its fast fermentative metabolism, which has played an important role in its domestication (*1*). *S. cerevisiae* converts glucose to ethanol via the Embden-Meyerhof-Parnas pathway of glycolysis and the fermentative pathway, encompassing a total of 12 enzymatic steps (*2, 3*). While *S. cerevisiae* can respire glucose, leading to an ATP yield of 16 moles of ATP per mole of glucose, it favours alcoholic fermentation. Indeed, even in the presence of oxygen, glucose excess triggers ethanol formation in *S. cerevisiae* and its relatives from the *Saccharomyces* genus, a phenomenon known as the Crabtree effect (*4, 5*). To sustain the energy demand for growth and maintenance despite the low ATP yield of alcoholic fermentation (2 moles of ATP per glucose molecule), the glycolytic flux in *S. cerevisiae* can easily reach fluxes of 20 to 25 mmoles ethanol per gram dry weight per hour (*4*). This high activity of the glycolytic pathway is reflected in the remarkably high concentration of glycolytic enzymes in the cell, which can represent up to 30% of the total amount of soluble protein (*5, 6*).

The genome of *S. cerevisiae* is characterized by a high genetic redundancy which can largely be attributed to a whole genome duplication event (*7, 8*). This redundancy is even more prominent among 'metabolic' genes and is remarkably elevated in the glycolytic and fermentative pathways of *S. cerevisiae* (*9-11*). These two pathways have been thoroughly investigated (and even established) in *S. cerevisiae* (*2, 12*). With the exception of three steps that are catalysed by single enzymes, i.e. phosphoglucose isomerase (Pgi1), fructose-bisphosphate aldolase (Fba1) and triosephosphate isomerase (Tpi1), the glycolytic and fermentative steps are catalysed by at least two and potentially up to seven isoenzymes for alcohol dehydrogenase (Adh). However, not all isoenzymes are equally important for the glycolytic and fermentative activity. With the notable exception of Pfk1 and Pfk2, two isoenzymes forming a heterooctamer that are equally important for the functionality of phosphofructokinase (*13, 14*), for each step, a single isoenzyme is responsible for the bulk of the glycolytic and fermentative flux. These so-called major isoenzymes are encoded by major paralogs, which expression is strong and constitutive (i.e. *HXK2*, *TDH3*, *GPM1*, *ENO2*, *PYK1*, *PDC1*, *ADH1*) (*4*). Because of these properties, glycolytic promoters are often used to drive gene expression in engineered strains (*15*). Conversely the expression of minor paralogs is, in most instances, far lower than the expression of the corresponding major paralogs and is condition-dependent (*4, 16-18*). Following duplication events, redundant genes can have different fates. If their presence brings additional benefits to the cell, either in their native form or via neo-functionalization, the gene and its duplicate will be retained

in the genome, otherwise the redundant copy will be lost (*9, 19*). The fact that the glycolytic and fermentative pathways still contain many paralogs that do not display obvious new functions suggests that they might increase fitness under specific conditions. For example, *PDC6* encoding a pyruvate decarboxylase with low sulfur amino acid content is specifically induced in sulfur limiting conditions (*17, 20*). However challenging this theory, it was recently shown that the simultaneous removal of all minor paralogs from the glycolytic and fermentative pathways had no detectable effect on *S. cerevisiae* physiology under a wide variety of conditions (*4*).

The *Saccharomyces* genus consists of at least eight naturally occurring species which all evolved toward optimal performance in their niche, leading to different physiological characteristics (*21-23*). For instance, *Saccharomyces kudriavzevii, Saccharomyces uvarum* and *Saccharomyces eubayanus* are cold-tolerant, and perform better than *S. cerevisiae* at temperatures below 20°C (*24-27*). Strains belonging to different *Saccharomyces* species can mate and form viable hybrids, some of which play an important role in the beverage industry. For instance *Saccharomyces pastorianus*, a hybrid of *S. cerevisiae* and *S. eubayanus*, is the main lager-brewing yeast (*28*) and hybrids of *S. cerevisiae* and *S. kudriavzevii* and of *S. uvarum* and *S. eubayanus* (known as *S. bayanus*) play an important role in beer and wine fermentation (*29-32*). The cold-tolerance of *S. pastorianus* and *S. eubayanus* has indubitably promoted the selection of their hybrids with *S. cerevisiae* in cold environments (*24, 33, 34*).

In a recent study, using a unique yeast platform enabling the swapping of entire essential pathways, it was shown that *S. kudriavzevii* glycolytic and fermentative pathways could be transplanted in *S. cerevisiae* and could efficiently complement the native pathways. Expression of the full set of *S. kudriavzevii* orthologs in *S. cerevisiae*, expressed from *S. kudriavzevii* promoters, resulted in enzyme activities and physiological responses remarkably similar to the parental strain carrying a full set of native *S. cerevisiae* genes. However, the impact of *S. kudriavzevii* promoters on transcriptional activity in *S. cerevisiae* was not explored (*35*). Despite *S. eubayanus* and *S. kudriavzevii* industrial importance and the availability of their full genome sequence, remarkably little is known about the genetic makeup and transcriptional regulation of the glycolytic and fermentative pathways.

To address this knowledge gap, the present study explores the diversity of the glycolytic and fermentative pathways within the genus *Saccharomyces*, using the industrially-relevant yeasts *S. cerevisiae*, *S. eubayanus* and *S. kudriavzevii* as paradigms. More precisely, the presence and sequence similarity between paralogs in these three yeasts were explored. Cultivation in bioreactors combined with transcriptome analysis was

used to evaluate the presence of dominant paralogs in *S. eubayanus* and *S. kudriavzevii* and to compare the expression levels of glycolytic and fermentative orthologs in their native context. Finally, we explored transferability of *S. kudriavzevii* and *S. eubayanus* promoters by monitoring their expression and context-dependency upon transplantation in *S. cerevisiae*.

Material and Methods

Strains and culture conditions

All yeast strains used in this study are derived from the CEN.PK background (*36*) and are listed in Table 1. Yeast cultures for transformation and genomic DNA isolation were grown in 500 mL shake flasks with 100 mL of complex, non-selective media (YPD) containing 10 g L^{-1} Bacto Yeast extract, 20 g L^{-1} Bacto Peptone and 20 g L^{-1} glucose. Promoter regions were obtained from the strains *S. cerevisiae* CEN.PK113-7D (*36-38*), *S. kudriavzevii* CR85 a wild isolate from oak bark (supplied by Prof. Querol and dr. Barrio, Universitat de València, Spain) (*39*) and *S. eubayanus* CBS12357 (*34*). The same strains were used for transcriptome analysis, with the exception of *S. cerevisiae* for which the diploid strain CEN.PK122 was used instead of the haploid CEN.PK113-7D (*36*). All *S. cerevisiae* strains were grown at 30˚C and *S. kudriavzevii* and *S. eubayanus* at 20 ˚C in shake flasks at 200 rpm, unless different conditions are mentioned.

All transformations were done in *S. cerevisiae* CEN.PK113-5D using the auxotrophic marker *URA3* for selection. Synthetic medium containing 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4 \cdot 7H_2O$, 5 g L^{-1} $(NH_4)_2SO_4$, 1 mL L^{-1} of a trace element solution, and 1 mL L^{-1} of a vitamin solution was used (*40*). Synthetic medium supplemented with 20 g L^{-1} glucose (SMG) or 2% (vol/vol) ethanol (SMEtOH) was used for culture propagation where specified. For solid media 20 g L^{-1} agar was added prior to heat sterilization. For storage and propagation of plasmids *Escherichia coli* XL1-Blue (Agilent Technologies, Santa Clara, CA) was used, and grown in lysogeny broth (LB) supplemented with ampicillin (100 mg L^{-1}) (*41, 42*). For the storage of yeast and *E. coli* strains 30% or 15% (v/v) glycerol was added to exponentially growing cultures respectively and aliquots were stored at -80 ˚C.

Table 1 – Strains used in this study.

Strain	Genotype	Plasmid for integration	Source
Saccharomyces cerevisiae			
CEN.PK113-5D	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2*	-	(*36*)
CEN.PK113-7D	*MATa URA3 HIS3 LEU2 TRP1 MAL2-8c SUC2*	-	(*36-38, 43*)
CEN.PK122	*MATa/Matα*	-	(*36, 37*)
IMX1042	*MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pHXK2sc-mRuby2-tENO2, URA3)*	pUDI098	This study
IMX1016	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pHXK2sk-mRuby2-tENO2, URA3)*	pUDI097	This study
IMX1102	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pHXK2se -mRuby2-tENO2, URA3)*	pUDI108	This study
IMX1068	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPGI1sc-mRuby2-tENO2, URA3)*	pUDI101	This study
IMX1017	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPGI1sk-mRuby2-tENO2, URA3)*	pUDI095	This study
IMX1103	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPGI1se -mRuby2-tENO2, URA3)*	pUDI109	This study
IMX1171	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPFK1sc-mRuby2-tENO2, URA3)*	pUDI121	This study
IMX1249	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPFK1sk-mRuby2-tENO2, URA3)*	pUDI126	This study
IMX1174	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPFK1se-mRuby2-tENO2, URA3)*	pUDI118	This study
IMX1175	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPFK2sc-mRuby2-tENO2, URA3)*	pUDI131	This study
IMX1176	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPFK2sk-mRuby2-tENO2, URA3)*	pUDI130	This study
IMX1177	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPFK2se-mRuby2-tENO2, URA3)*	pUDI132	This study
IMX1041	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pFBA1sc-mRuby2-tENO2, URA3)*	pUDI099	This study
IMX1070	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pFBA1sk-mRuby2-tENO2, URA3)*	pUDI103	This study
IMX1097	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pFBA1se-mRuby2-tENO2, URA3)*	pUDI186	This study
IMX1132	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pTPI1sc-mRuby2-tENO2, URA3)*	pUDI114	This study
IMX1133	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pTPI1sk-mRuby2-tENO2, URA3)*	pUDI115	This study
IMX1134	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pTPI1se-mRuby2-tENO2, URA3)*	pUDI116	This study
IMX1018	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pTDH3sc-mRuby2-tENO2, URA3)*	pUDI094	This study
IMX1128	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pTDH3sk-mRuby2-tENO2, URA3)*	pUDI110	This study
IMX1130	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pTDH3se-mRuby2-tENO2, URA3)*	pUDI112	This study
IMX1043	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPGK1sc-mRuby2-tENO2, URA3)*	pUDI100	This study

Strain	Genotype	Plasmid for integration	Source
IMX1019	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPGK1sk-mRuby2-tENO2, URA3)*	pUDI096	This study
IMX1069	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPGK1se-mRuby2-tENO2, URA3)*	pUDI102	This study
IMX1100	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pGPM1sc-mRuby2-tENO2, URA3)*	pUDI106	This study
IMX1071	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pGPM1sk-mRuby2-tENO2, URA3)*	pUDI104	This study
IMX1101	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pGPM1se-mRuby2-tENO2, URA3)*	pUDI107	This study
IMX1178	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pENO2sc-mRuby2-tENO2, URA3)*	pUDI122	This study
IMX1299	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pENO2sk-mRuby2-tENO2, URA3)*	pUDI123	This study
IMX1180	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pENO2se-mRuby2-tENO2, URA3)*	pUDI119	This study
IMX1181	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPYK1sc-Ruby2-tENO2, URA3)*	pUDI128	This study
IMX1182	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPYK1sk-mRuby2-tENO2, URA3)*	pUDI127	This study
IMX1183	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPYK1se-mRuby2-tENO2, URA3)*	pUDI129	This study
IMX1242	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPDC1sc-mRuby2-tENO2, URA3)*	pUDI161	This study
IMX1243	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPDC1sk-mRuby2-tENO2, URA3)*	pUDI162	This study
IMX1244	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pPDC1se-mRuby2-tENO2, URA3)*	pUDI163	This study
IMX1245	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pADH1sc-mRuby2-tENO2, URA3)*	pUDI158	This study
IMX1246	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pADH1sk-mRuby2-tENO2, URA3)*	pUDI159	This study
IMX1298	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pADH1se-mRuby2-tENO2, URA3)*	pUDI160	This study
IMX1099	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pACT1sc-mRuby2-tENO2, URA3)*	pUDI105	This study
IMX1168	*MATa ura3 HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3::(pTEF1sc-mRuby2-tENO2, URA3)*	pUDI124	This study
Other Saccharomyces species			
S. kudriavzevii CR85	*MATa/Matα*	-	(*39*)
S. eubayanus CBS12357	*MATa/Matα*	-	(*34*)

Molecular biology techniques

For high fidelity PCR amplification Phusion high fidelity polymerase (Thermo Scientific, Landsmeer, The Netherlands) was used according to manufacturer´s instructions. To improve efficiency of the PCR reactions, primer concentrations were decreased from 500 to 200 nM and the polymerase concentration was increased from 0.02 to 0.03 U μL^{-1}. PCR products were treated with 1µL DpnI FastDigest restriction enzyme (Thermo Fisher Scientific) for 1h at 37°C to remove residual circular templates. Afterwards, the mixture was purified using GenElute™ PCR Clean-Up Kit (Sigma-Aldrich, St. Louis, MO) according to manufacturer's protocol. PCR for diagnostic purposes was done using DreamTaq PCR mastermix (Thermo Fisher Scientific) according to manufacturer's recommendations. Primers used in this study are listed in Tables S1 and S2. PCR products were resolved on 1% agarose gel with Tris-acetate-EDTA (TAE) buffer. Genomic DNA used as template for PCR amplification of the promoter regions was isolated using YeaStar genomic DNA kit (Zymo Research, Orange, CA) according to manufacturer's protocol. Plasmids were extracted from *E. coli* using the GenElute plasmid miniprep kit (Sigma-Aldrich, St. Louis, MO) according to manufacturer's description and eluted with miliQ water. Restriction analysis of plasmids was done using FastDigest restriction enzymes with FastDigest Green Buffer (Thermo Fisher Scientific) incubating for 30 minutes at 37°C according to manufacturer's recommendations.

Promoters, plasmids and yeast strain construction

A schematic overview of the subsequent plasmid and strain construction steps is provided in Figure 1. Plasmids used in this study are reported in Table S3 (available at https://doi.org/10.3389/fgene.2018.00504). The *HXK2, PGI1, PFK1, PFK2, FBA1, TPI1, TDH3, PGK1, GPM1, ENO2, PYK1, PDC1* and *ADH1* and reference *TEF1* and *ACT1* promoter regions of approximately 800bp (see Table S4 for exact lengths) were PCR-amplified from *S. cerevisiae* CEN.PK113-7D, *S. kudriavzevii* CR85 and *S. eubayanus* CBS 12357 genomic DNA using primers listed in Table S1 (available at https://doi.org/10.3389/fgene.2018.00504). For compatibility with Golden Gate cloning, promoter sequences were flanked with BsaI and BsmBI restriction sites introduced as primer overhangs in the PCR amplification step.

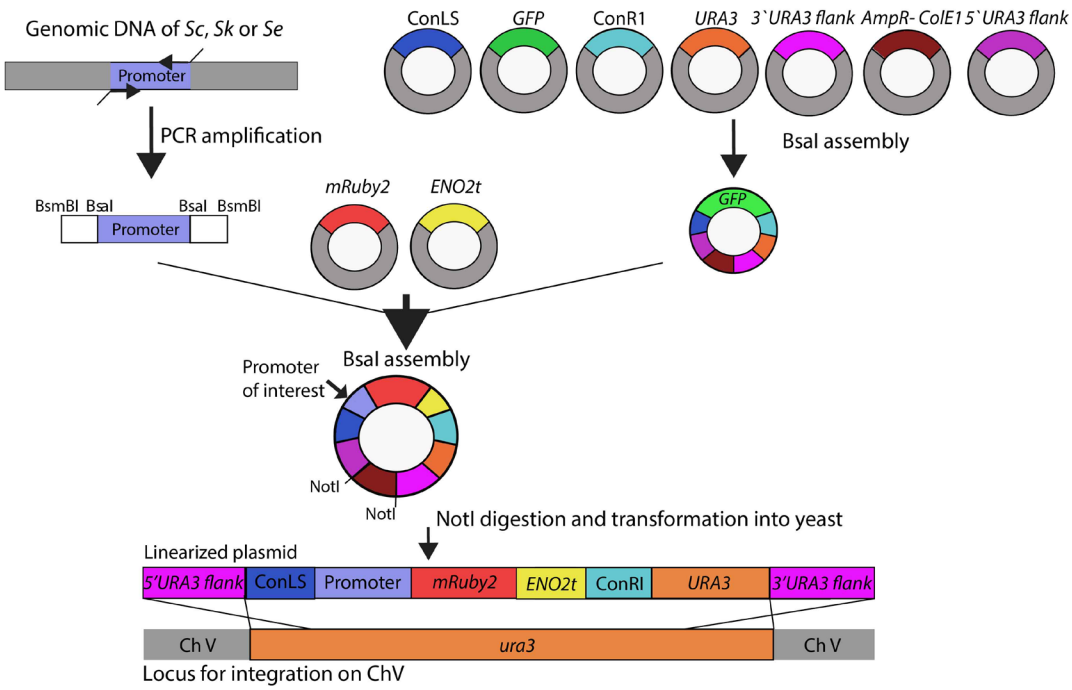


Figure 1 – Schematic representation of strain construction workflow. Glycolytic promoters of *S. cerevisiae (Sc), S. kudriavzevii (Sk)* and *S. eubayanus (Se)* were PCR-amplified using primers with specific BsaI flanks. First a 'GFP dropout' plasmid was assembled from the following parts containing all unique overhangs for assembly: two connectors ConLS and ConR, *URA3* marker, 5' and 3' *URA3* flanks and the Amp-ColE1 containing the marker and origin of replication for *E. coli*. This plasmid was used in a second round of BsaI golden gate assembly to replace the GFP fragment by the promoter of interest, *mRuby2* and *ENO2* terminator. The resulting plasmids were linearized by NotI restriction and integrated in the *ura3* locus of *S. cerevisiae* strain IMX1076.

The plasmid backbone was constructed by Golden Gate assembly using the collection of part plasmids provided in the Yeast Toolkit (*44*). To increase the efficiency of plasmid assembly, first a *GFP* dropout plasmid pUD428 was constructed containing a *URA3* marker, AmpR selection marker, bacterial origin of replication, two connector fragments and a *GFP* gene surrounded by *URA3* upstream and downstream homology flanks (Table S3). The correct assembly of plasmids was checked by restriction analysis. The *GFP* dropout cassette from pUD428 was subsequently replaced by the *mRuby2* gene flanked by a promoter of interest and by the *ENO2* terminator using Golden Gate cloning with BsaI. The reaction mixture was prepared with 1µL T4 DNA ligase buffer (Thermo Fisher Scientific), 0.5 µL T7 DNA ligase (NEB), 0.5µL FastDigest Eco31I (BsaI) (Thermo Fisher Scientific) and 10 ng of each DNA fragment. MiliQ H_2O was added to a final volume of 10 µL. The assembly was done in a thermocycler using 25 cycles of restriction and ligation: 42°C for 2 min, 16°C for 5 min, followed by a final digestion step (60°C for 10 min) and an inactivation step (80°C for 10 min). If one of the fragments contained an internal BsaI site, the final digestion and inactivation steps were omitted. 1 µL of the

assembly mix was transformed to *E. coli* (XL1-Blue) according to manufacturer's description and plated on selective LB medium. Correct ligation of the promoter-*mRuby2*-terminator construct in this plasmid resulted in the loss of the *GFP* gene, which could be easily screened based on colony colour. Additional plasmid confirmation was done by restriction analysis.

Prior to transformation to yeast, the constructed plasmids containing the promoter of interest, the *mRuby2* gene and the *ENO2* terminator were linearized by digestion with NotI (FastDigest, Thermo Fisher Scientific) according to manufacturer's protocol for 30 min at 37°C. 400 ng of each plasmid was digested and the mixture was directly transformed to the strain CEN.PK113-5D in which the linearized plasmid was integrated in the *ura3-52* locus. Yeast transformations were done according to Gietz and Woods (*45*). Colonies were screened by PCR (Table S2, available at https://doi.org/10.3389/fgene.2018.00504).

Batch cultivation in bioreactors

Samples for transcriptome analysis of *S. cerevisiae* (CEN.PK122), *S. kudriavzevii* (CR85) and *S. eubayanus* (CBS 12357) were obtained from aerobic batch cultures in bioreactors performed in independent duplicate. Batch cultures were performed in SMG supplemented with 0.2 g L^{-1} antifoam Emulsion C (Sigma-Aldrich, St. Louis, MO). The reactors were inoculated at a starting OD_{660} of 0.3 with cells resuspended in demineralized water, which were obtained from exponentially growing shake flask cultures incubated at the same temperature and with the same medium as was used in the bioreactors (SMG). Cultures were performed in 2 L bioreactors (Applikon, Schiedam, The Netherlands) containing a 1.4 L working volume. The cultures were constantly stirred at 800 rpm, sparged with 700 mL min^{-1} dried compressed air (Linde Gas Benelux, The Netherlands) and maintained at 30°C for *S. cerevisiae* and 25°C for *S. kudriavzevii* and *S. eubayanus.* The culture pH was kept at 5.0 during growth on glucose by automatic addition of 2M KOH.

Extracellular metabolites were determined by high-performance liquid chromatography (HPLC) analysis using a Aminex HPX-87H ion-exchange column operated at 60°C with 5 mM H_2SO_4 as the mobile phase at a flow rate of 0.6 mL min^{-1} (Agilent, Santa Clara, USA). Samples were centrifuged for 3 min at 20.000g and the supernatant was used for analysis.

Biomass dry weight was determined in analytical duplicate by filtration of 10 mL sample on filters (pore-size 0.45 µm, Whatman/GE Healthcare Life Sciences, United Kingdom) pre-dried in a microwave oven at 360W for 20 minutes, as previously described (*40*). Optical density at 660 nm (OD_{660}) was determined in a Libra S11

spectrophotometer (Biocrom, Cambridge, United Kingdom). The CO_2 and O_2 concentration in the gas outflow was analysed by a Rosemount NGA 2000 analyser (Baar, Switzerland), after cooling of the gas by a condenser (2°C) and drying using a PermaPure Dryer (model MD 110-8P-4; Inacom Instruments, Veenendaal, the Netherlands). Sampling for transcriptome analysis was done during mid-exponential growth on glucose at a biomass concentration of approximately 1 g L^{-1}. Sampling in liquid nitrogen and RNA extraction were performed as previously described (*46*).

Promoter activity assay

Promoter activity measurement of the *mRuby2* reporter strain library was performed in 96-well plates. Precultures were grown in 12-well plates in 1.5 mL volume in a thermoshaker (Grant-bio PHMP-4, United Kingdom) with constant shaking (800 rpm) and temperature. Precultures were grown at the temperature of the subsequent plate assay (30°C or 20°C). For the first preculture YPD medium was inoculated from glycerol stocks and grown overnight till saturation. From this culture 20 µL were transferred to new 12-well plates and the strains were grown under the conditions of interest till mid-exponential phase (corresponding to OD_{660} of 3 to 5). Afterwards the culture was centrifuged at 3000 g for 5 minutes, the supernatant was removed and cells were resuspended in fresh medium to an OD_{660} of 0.3 and transferred in volumes of 100 µL to a 96-well plate (Corning™ polystyrene white/transparent bottom, Greiner Bio-One) using six replicate wells per strain. To prevent evaporation, all plates, including preculture plates, were covered with sterile polyester acrylate sealing tape (Thermo Scientific). To supply sufficient levels of oxygen throughout the cultures, small openings were created in each well with a needle. The plate assays were performed in a plate reader (TECAN infinite M200 Pro) with constant temperature and shaking (orbital, 1mm). Every 20 minutes the optical density (OD_{660}) and the fluorescence using excitation and emission wavelengths 559 nm/600 nm were measured. Cultures were monitored till saturation. A non-fluorescent CEN.PK113-7D strain was taken along every run to determine the background fluorescence, as well as two reference reporter strains expressing *mRuby2* from the *TEF1* and *ACT1* promoters from *S. cerevisiae*. For every well, OD_{660} and fluorescence values from all time points during exponential growth were plotted against each other and the promoter activity was calculated as the slope of the linear regression between optical density and fluorescence.

Flow cytometry analysis

mRuby2 fluorescence intensity of individual cells from cultures grown in the TECAN plate reader was determined using flow cytometry. Mid-exponential cultures from the plate reader were diluted in Isoton II (Beckman Coulter) and the fluorescence intensity was determined for 10000 cells per sample on a BD FACSAriaII equipped with an 561

nm excitation laser and 582/15 nm emission filter. Data were analysed using FlowJo v10.2 (FlowJo LLC). As expected from strains in which the *mRuby2* expression system is integrated in the genome the fluorescence signal was homogeneously distributed among the yeast population (Figure S5).

Whole genome sequencing

To obtain genome sequences of high quality, the strain *S. kudriavzevii* CR85 was sequenced in-house both by Illumina Miseq sequencing (Illumina, San Diego, CA) and by Oxford Nanopore Technology MinION sequencing (Oxford Nanopore Technology, Oxford, United Kingdom). Genomic DNA was isolated using the Qiagen 100/G kit (Qiagen, Hilden, Germany) and the concentration was determined using Qubit® Fluorometer 2.0 (ThermoFisher Scientific). Illumina library preparation was done as described previously (*47*).

For Nanopore sequencing, 3 µg of genomic DNA were diluted in a total volume of 46 uL and then sheared with a g-TUBE (Covaris, Brighton, United Kingdom) to an average fragment size of 8-10 kb. The input DNA was then prepared for loading in a FLO-MIN106 flow cell with R9.4 chemistry and the 1D ligation sequencing kit (SQK-LSK108), following manufacturer's instructions with the exception of a size selection step with 0.4x (instead of 1x) AMPure beads after the End-Repair/dA tailing module and the use of 80% (instead of 70%) ethanol for washes. Raw files generated by MinKNOW were base called using Albacore (version 1.2.5; Oxford Nanopore Technology). Reads, in fastq format, with minimum length of 1000 bp were extracted, yielding 4.15 Gigabase sequence with an average read length of 4.3 kb.

De novo assembly was performed using Canu (v1.4, settings: genomesize=12m) (*48*) producing an 11.87 Megabase genome into 20 contigs of which 13 contigs in chromosome length plus 1 mitochondrial DNA, while 3 chromosomes consisted of 2 contigs each. The contig pairs were manually joined (with 1000 N's between the contigs) into 3 chromosomes (chromosomes VII, XII and XVI). Pilon (*49*) was then used to further correct assembly errors by aligning Illumina reads, using BWA (*50*) to the assembly using correction of only SNPs and short indels (–fix bases parameter). Gene annotations were performed using the MAKER2 annotation pipeline (version 2.31.9) (*51*) using SNAP (version 2013–11-29) (*52*) and Augustus (version 3.2.3) (*53*) as *ab initio* gene predictors. S288C EST and protein sequences were obtained from SGD (Saccharomyces Genome Database, http://www.yeastgenome.org) and were aligned using BLASTX (BLAST version 2.2.28+)(*54*). The translated protein sequence of the final gene model was aligned using BLASTP to S288C protein Swiss-Prot database (http://www.ebi.ac.uk/swissprot/). For CEN.PK113-7D and *S. eubayanus* CBS 12357

existing sequencing data was used (*43, 55*). The sequencing data are available at NCBI under bioproject accession number PRJNA480800.

RNA sequencing and data analysis

Library preparation and RNA sequencing were performed by Novogene Bioinformatics Technology Co., Ltd (Yuen Long, Hong Kong). Sequencing was done with Illumina paired end 150 bp sequencing read system (PE150) using a 250~300bp insert strand specific library which was prepared by Novogene. For the library preparation, as described by Novogene, mRNA enrichment was done using oligo(dT) beads. After random fragmentation of the mRNA, cDNA was synthetized from the mRNA using random hexamers primers. Afterwards, second strand synthesis was done by addition of a custom second strand synthesis buffer (Illumina), dNTPs, RNase H and DNA polymerase I. Finally, after terminal repair, A ligation and adaptor ligation, the double stranded cDNA library was finalized by size selection and PCR enrichment.

The sequencing data for the three strains, *S. cerevisiae* CEN.PK122, *S. kudriavzevii* CR85 and *S. eubayanus* CBS 12357 obtained by Novogene had an average read depth of 21, 24 and 24 million reads per sample respectively. For each sample, reads were aligned to the relevant reference genome using a two-pass STAR procedure (*56*). In the first pass, we assembled a splice junction database which was used to inform the second round of alignments. As paralogs in the glycolytic pathways were highly similar, we used stricter criteria for aligning and counting reads to facilitate delineation of paralogs. Introns were allowed to be between 15 and 4000 bp, and soft clipping was disabled to prevent low quality reads from being spuriously aligned. Ambiguously mapped reads were removed. Expression was quantified per transcript using htseq-count in strict intersection mode (*57*). As we wished to compare gene expression across genomes, where orthologs may have different gene lengths, data were normalized for gene length. Therefore the average FPKM expression counts for each gene in each species were calculated (*58*). The genomes from *S. cerevisiae* CEN.PK113-7D, *S. kudriavzevii* CR85 and *S. eubayanus* CBS 12357 were used as reference (NCBI BioProject accession numbers PRJNA52955, PRJNA480800 and PRJNA264003 respectively (https://www.ncbi.nlm.nih.gov/bioproject/)). Data are available at Gene Expression Omnibus with accession number GSE117404. CEN.PK113-7D transcriptome data is available on Gene Expression Omnibus database under accession number GSE63884.

Comparison of DNA sequences

Sequences from annotated glycolytic ORF and promoters of *S. cerevisiae* CEN.PK113-7D, *S. kudriavzevii* CR85 and *S. eubayanus* CBS 12357 were used for alignments with Clone Manager 9 Professional Edition, (NCBI BioProject accession numbers

PRJNA52955, PRJNA480800 and PRJNA264003 respectively (https://www.ncbi.nlm.nih.gov/bioproject/)). For the *TPI1* sequence alignment the sequences with the following accession numbers were used: CU928179 (*Z. rouxii*), HE605205 (*C. parapsilosis*), CP028453 (*Y. lipolytica*), AJ390491 (*C. albicans*), XM_002551264 (*C. tropicalis*), AJ012317 (*K. lactis*), FR839630 (*P. Pastoris*) AWRI1499 (D. bruxellensis), XM_018355487 (O. parapolymorpha), CR380954 (C. glabrata), CP002711 (A. gossypii), AP014602 (K. marxianus), XM_001642913 (K. polysporus), CP000501 (S. stipitis), XM_002616396 (C. lusitaniae) and CP028714 (E. coli). Alignment of these sequences was performed using multiple sequence alignment in Clustal Omega (*59, 60*) and the phylogenetic trees were obtained with JalView (version 2.10.4b1) using average distance and percentage identity (*61*).

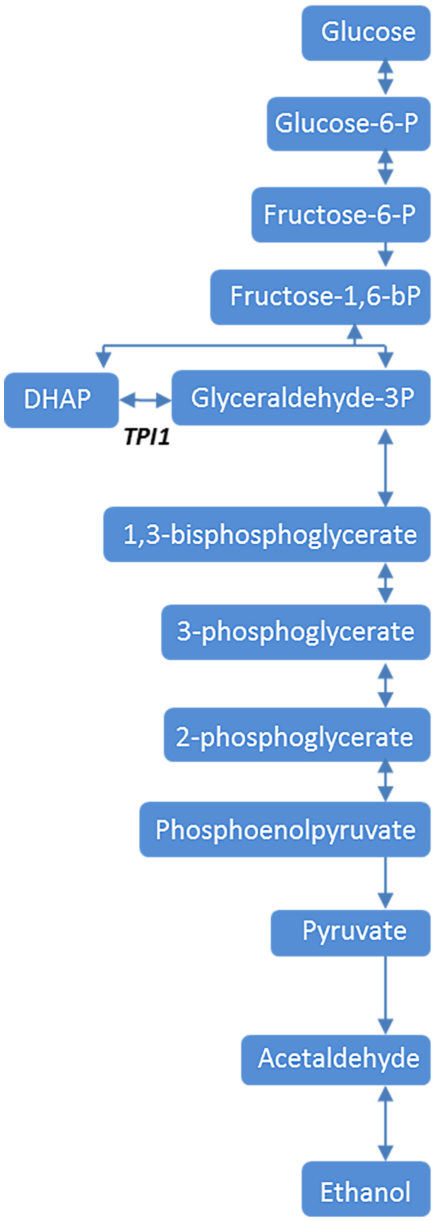
Statistics

Statistical analysis was performed using the software IBM SPSS statistics 23 (SPSS inc. Chicago, USA). For transcriptome data, fluorescence data and batch culture data analysis of variance (ANOVA) with Dunnett post-hoc test was performed to test if the results for *S. kudriavzevii* and *S. eubayanus* were statistically different from *S. cerevisiae.*

Results

Genetic makeup of the glycolytic and fermentative pathways in *S. cerevisiae* and its close relatives *S. kudriavzevii* and *S. eubayanus*

The genetic makeup of pathways involved in central carbon metabolism in *S. cerevisiae* has already been well characterized, and more particularly for glycolysis and alcoholic fermentation. The ten reactions of the glycolytic pathway and the two reactions of ethanolic fermentation in *S. cerevisiae* are catalyzed by a set of 26 enzymes encoded by 26 genes (Figure 2). High quality sequences are already available for *S. cerevisiae* and *S. eubayanus* (*43, 55*). To explore these pathways in *S. kudriavzevii* the strain *S. kudriavzevii* CR85 was sequenced using both Illumina and Oxford Nanopore technologies (see Materials and Methods section and Table S5). *S. cerevisiae*'s high genetic redundancy and the locations of the genes were fully mirrored in *S. kudriavzevii* and *S. eubayanus* genomes (Fig. 1). The only exception was the absence of *PDC6* in *S. kudriavzevii.* While a *ScPDC6* ortholog with 81% identity was identified in *S. eubayanus*, no ortholog could be found in *S. kudriavzevii*. For all other glycolytic genes from *S. cerevisiae*, genes with 80–97% homology of the coding regions were found in *S. kudriavzevii* and *S. eubayanus* (Figure 2). Overall, genes from *S. eubayanus* were slightly more distant from their *S. cerevisiae* orthologs than genes from *S. kudriavzevii,* which is in line with earlier reports (*62, 63*) (Figure 2).



Gene	Coding region % ID to *Sc*		Promoter % ID to *Sc*	
	Sk	*Se*	*Sk*	*Se*
HXK1	86	86	66	66
HXK2	90	89	64	64
GLK1	85	81	62	57
PGI1	91	90	54	51
PFK1	89	88	76	74
PFK2	90	90	77	72
FBA1	94	94	72	61
TPI1	95	96	74	71
TDH1	93	91	59	65
TDH2	96	96	66	63
TDH3	95	95	75	73
PGK1	96	94	76	68
GPM1	95	89	70	68
GPM2	83	81	49	53
GPM3	83	80	70	68
ENO1	93	93	70	65
ENO2	97	92	79	69
PYK1	95	96	69	69
PYK2	85	82	60	62
PDC1	95	95	82	74
PDC5	92	91	80	79
PDC6	absent	81	absent	56
ADH1	95	93	63	50
ADH2	88	85	70	68
ADH4	88	86	65	61
ADH5	85	83	51	48

40%
100%

Figure 2 – Genes and reactions involved in glycolysis and alcoholic fermentation in *S. cerevisiae* and sequence comparison between the promoters and coding regions of *S. cerevisiae (Sc) S. kudriavzevii (Sk)* and *S. eubayanus (Se)*. The major paralogs in *S. cerevisiae* are represented in bold. The coding regions and promoter regions (800 bp) of *S. kudriavzevii* and *S. eubayanus* were aligned to the corresponding *S. cerevisiae* sequences and the percentage identity is indicated. *PDC6* was absent in *S. kudriavzevii.*

In addition to the coding regions, the promoter regions were compared. Since the exact length of most promoter regions is not clearly defined, the 800 bp upstream of the coding regions were considered as promoters. Promoter sequences were substantially less conserved than the coding sequences, ranging from 43 to 78% identity when comparing *S. kudriavzevii* and *S. eubayanus* to *S. cerevisiae* promoters (Figure 2). Remarkably, some regions covering up to 45 bp were strictly conserved among the three species, whereas other parts of the promoter sequences hardly shared homology (see example of *PGK1p* on Figure S1). As promoter regions are poorly defined, promoters shorter than 800 bp might be fully functional. Alignment with shorter regions might therefore increase the degree of homology between promoters. Alignments using 500 bp upstream the coding region only slightly increased the alignment percentages (up to 7%), mostly as a consequence of the enrichment for conserved transcription factor binding sites located between 100 and 500 bp upstream of the ORF (*64*). Notably, orthologs with a relatively high or low degree of conservation between *S. cerevisiae* and *S. kudriavzevii* also displayed a similar pattern when comparing *S. eubayanus* to *S. cerevisiae.* For example, the *SkGPM2* and *SeGPM2* promoters both have a relatively low homology (49% and 53%) to the *ScGPM2* promoter, whereas the *SkPFK1* and *SePFK1* promoter have both a high degree (76% and 74%) of similarity to *ScPFK1.* Interestingly, the genes and promoters displaying a relatively low degree of homology between *S. cerevisiae* and its relatives, are homologs considered as minor in *S. cerevisiae* (for example *GPM2* and *PYK2*) (Figure 2). Blast searches did not identify additional glycolytic orthologs present in *S. eubayanus* or *S. kudriavzevii* but absent in *S. cerevisiae.*

The activity of a promoter strongly depends on the presence of regulatory sequences as the TATA box and other specific transcription factor binding sites. In *S. cerevisiae*, the most important glycolytic transcription factor is Gcr1, which has been experimentally shown to bind to most glycolytic promoters and to activate the expression of the corresponding genes as summarized before (*65*). Gcr1 binding sites are only active when located next to DNA consensus sequences bound by Rap1 (*66*), a more pleiotropic transcription factor involved in the transcriptional regulation of a wide variety of genes including many glycolytic genes (*65*). Another multifunctional transcription factor is Abf1 which binds to several glycolytic promoters (*65*). With a single exception, all binding sites for Rap1, Gcr1 and Abf1 which were experimentally proven to be active in *S. cerevisiae*, were conserved in *S. kudriavzevii* and *S. eubayanus* promoter regions (Figure 3). The exception was the *SeADH1* promoter in which the Rap1 and Gcr1 site which are conserved between *S. cerevisiae* and *S. kudriavzevii* could not be identified. Together with the presence and high protein similarity of the *Se*Rap1 (82%), *Sk*Rap1

(86%), *Se*Gcr1 (85%), and *Sk*Gcr1 (85%), proteins with *S. cerevisiae,* these results suggested that the regulation of the glycolytic genes might be similar in the three species.

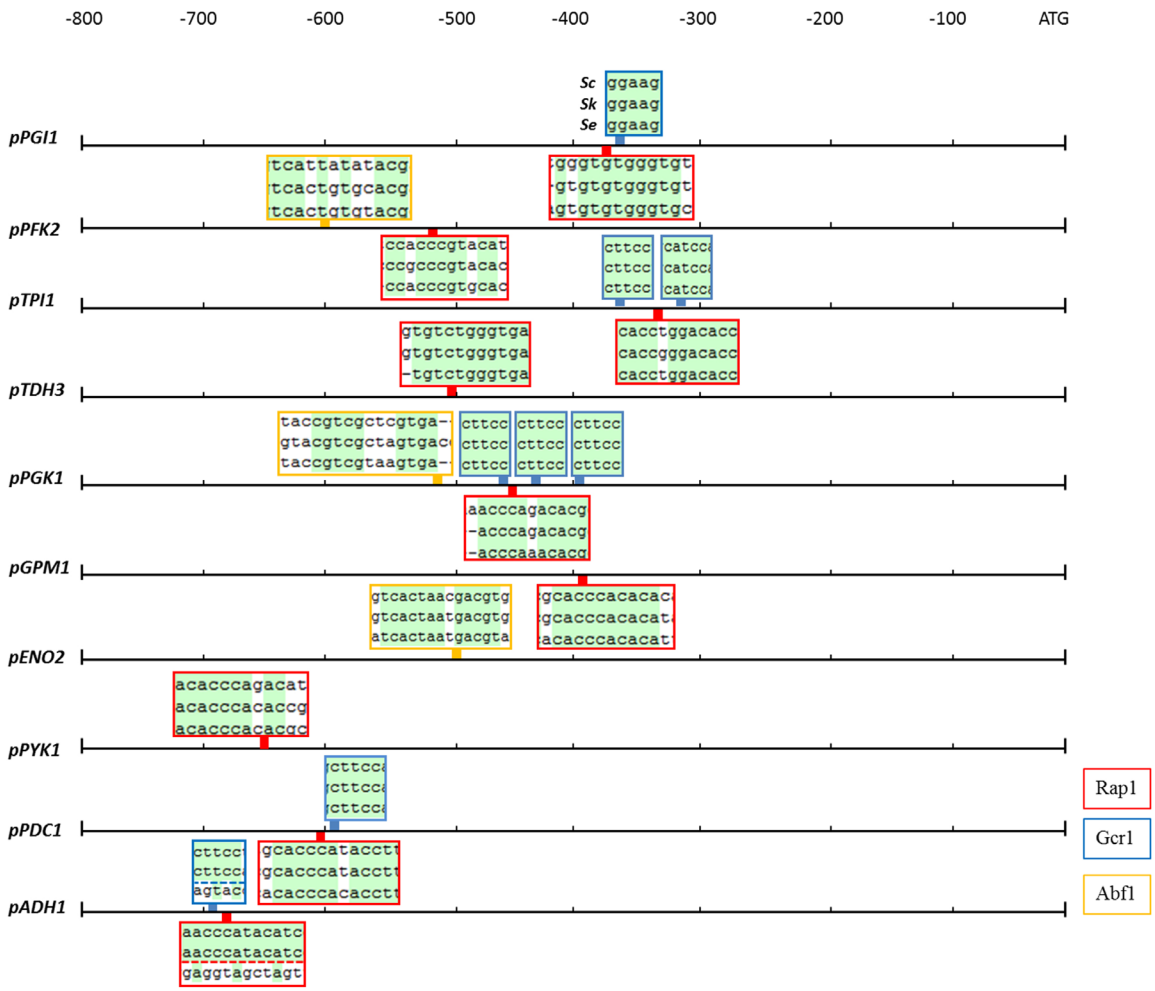


Figure 3 – Rap1, Gcr1 and Abf1 transcription factor binding sites which are conserved in the glycolytic and fermentation promoters of *S. cerevisiae, S. kudriavzevii* and *S. eubayanus.* The boxes indicate the location in the promoter of the binding sites for the Rap1 (red), Gcr1 (blue) and Abf1 (yellow) transcription factors which are experimentally shown to be functional in *S. cerevisiae.* The boxes contain the alignments of the three promoters at the transcription factor binding sites, conserved nucleotides are indicated in green. The Gcr1 and Rap1 sites in the *ADH1* promoter were not identified in *SeADH1p.*

Expression of the glycolytic genes during aerobic batch cultivation

To evaluate the similarity in glycolytic and fermentative gene expression, the transcriptome of *S. cerevisiae*, *S. kudriavzevii* and *S. eubayanus* was compared. *S. kudriavzevii* and *S. eubayanus* are both wild isolates and both diploid (*34, 39*). While many studies report the transcriptome of haploid *S. cerevisiae* strains, transcriptome data for diploid *S. cerevisiae* are scarce (*67, 68*). To obtain comparable transcriptome datasets for the three species, the diploid CEN.PK122 strain was used. The three diploid strains were grown in aerobic batch cultures in bioreactor using minimal chemically defined medium with glucose as sole carbon source. To ensure optimal growth conditions *S. cerevisiae* was cultivated at 30°C, while its cold-tolerant relatives that have lower temperature optima were cultivated at 25°C (*27, 69*). Under these conditions the maximum specific growth rate of *S. cerevisiae, S. kudriavzevii* and *S. eubayanus* was 0.38 h^{-1}, 0.25 h^{-1} and 0.33 h^{-1} respectively (Figure 4). Ethanol yields were similar for the three strains, but the biomass yield of *S. kudriavzevii* was significantly lower than that of its two relatives (Figure 4), which might reflect the higher relative cost of maintenance requirements at slow growth rates (*70*). For *S. eubayanus* we observed a lower glycerol yield as compared to its relatives, which was previously not observed under anaerobic conditions (*27*).

Transcriptome analysis of *S. cerevisiae*, *S. kudriavzevii* and *S. eubayanus* during mid-exponential growth phase revealed a remarkable similarity between the three species (Figure 5), despite differences in culture temperature. Furthermore, the major or minor classification of paralogous genes was fully conserved between the three species (Figure 5). From the genes considered as major paralogs the *SePFK1, SeFBA1, SkTDH3, SeTDH3, SkGPM1, SeENO2* and *SeADH1* genes displayed significantly lower expression levels as compared to *S. cerevisiae,* although only for *SeTDH3* and *SeADH1* the difference with *S. cerevisiae* was larger than 2-fold (8 and 3-fold respectively). For the minor paralogs slightly more variability was observed. Interestingly *SeHXK1* expression was 13-fold higher than its *S. cerevisiae* ortholog. All three *TDH* genes displayed a significantly lower expression in *S. kudriavzevii* and *S. eubayanus* as compared to *S. cerevisiae.* Likewise, for *ENO1* a lower expression was observed for *SeENO1* and even lower for *SkENO1* as compared to *ScENO1.* Finally, compared to *S. cerevisiae* a ca. 3-fold higher expression was observed for *SkPDC5* and *SeADH4.*

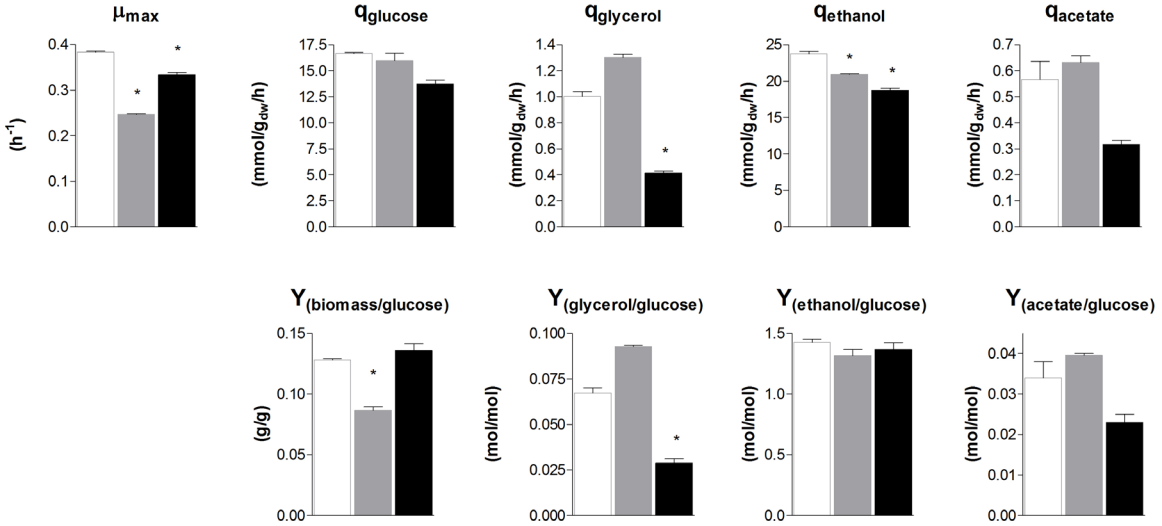


Figure 4 – Biomass specific rates and yields of *S. cerevisiae, S. kudriavzevii* and *S. eubayanus* batch cultivations in bioreactor. The strains were grown aerobically in synthetic medium supplemented with 20 g L^{-1} glucose. *S. cerevisiae* CEN.PK122 (white) was grown at 30°C, and *S. kudriavzevii* CR85 (grey) and *S. eubayanus* CBS 12357 (black) at 25°C. Asterisks indicate significant difference from *S. cerevisiae* (One-Way ANOVA, Dunnett post hoc test, P<0.01).

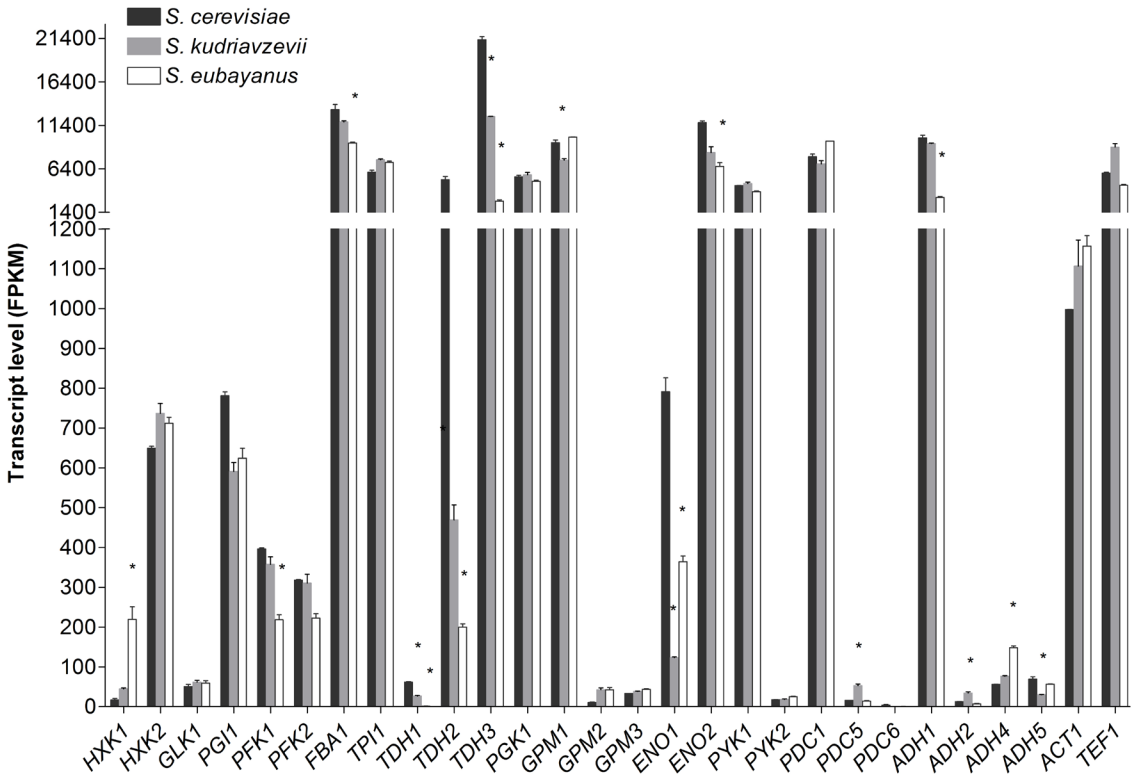


Figure 5 – Transcript levels of the diploid strains *S. cerevisiae* (black), *S. kudriavzevii* (grey) and *S. eubayanus* (white) from two biological replicates during mid-exponential growth in aerobic batch fermentations on glucose. Asterisks indicate significant difference from *S. cerevisiae* per gene (One-Way ANOVA, Dunnett post hoc test, P<0.01).

Optimization of microtiter plate assays to monitor promoter strength via fluorescent reporters

To explore the transferability of promoters within the *Saccharomyces* genus, the promoters of the major glycolytic and fermentative genes (indicated in bold in Figure 2) of *S. kudriavzevii* and *S. eubayanus* were functionally characterized in *S. cerevisiae*. A library of fluorescent reporter strains in which *mRuby2* expression was driven by heterologous promoters and, for comparison, by *S. cerevisiae* promoters, was constructed. To avoid bias due to gene copy number, the constructs were integrated in *S. cerevisiae* genome, at the *URA3* locus. The strains were cultured in 96-well plates, sealed with a transparent foil to prevent evaporation. Simultaneous monitoring of optical density and fluorescence revealed a premature saturation of the fluorescence signal as compared to biomass formation (Figure S2A and S2B). Fluorescent proteins have a strict requirement for molecular oxygen for the synthesis of their chromophores (*71*). The poor oxygenation of the cultures in sealed plates combined with the competition for oxygen between cellular respiration, anabolic reactions and mRuby2 maturation could explain the early saturation of the fluorescence signal. Unfortunately, this effect is rarely reported in literature and could be easily overlooked when fluorescence is measured at only one or few time points. Plate readers are widely used as method to characterize promoters with fluorescence reporters (*44, 72-74*) however, information provided in material and method sections are often scarce or incomplete, which makes reproduction of data by other groups difficult. To increase oxygen transfer while preventing evaporation, a small aperture was created in each well by puncturing the seal with a needle. The presence of an aperture had a strong impact on the fluorescence intensity of the cultures, enabling to monitor the cultures for a prolonged period of time (Figure S2B). Also during growth with ethanol as sole carbon source, for which oxygen requirement is substantially increased, no premature saturation of fluorescence was observed (Figure S2C and S2D). The location of the aperture in the well did not affect the fluorescence intensity (data not shown). To further evaluate the reliability of the fluorescence signal measured by the plate reader as well as the cell-to-cell heterogeneity of the fluorescence signal, measurements were also performed by flow cytometry. Comparing these data with the plate reader data revealed a very strong correlation of the fluorescence measured with these two techniques (R2 = 0,96, Figure S3)

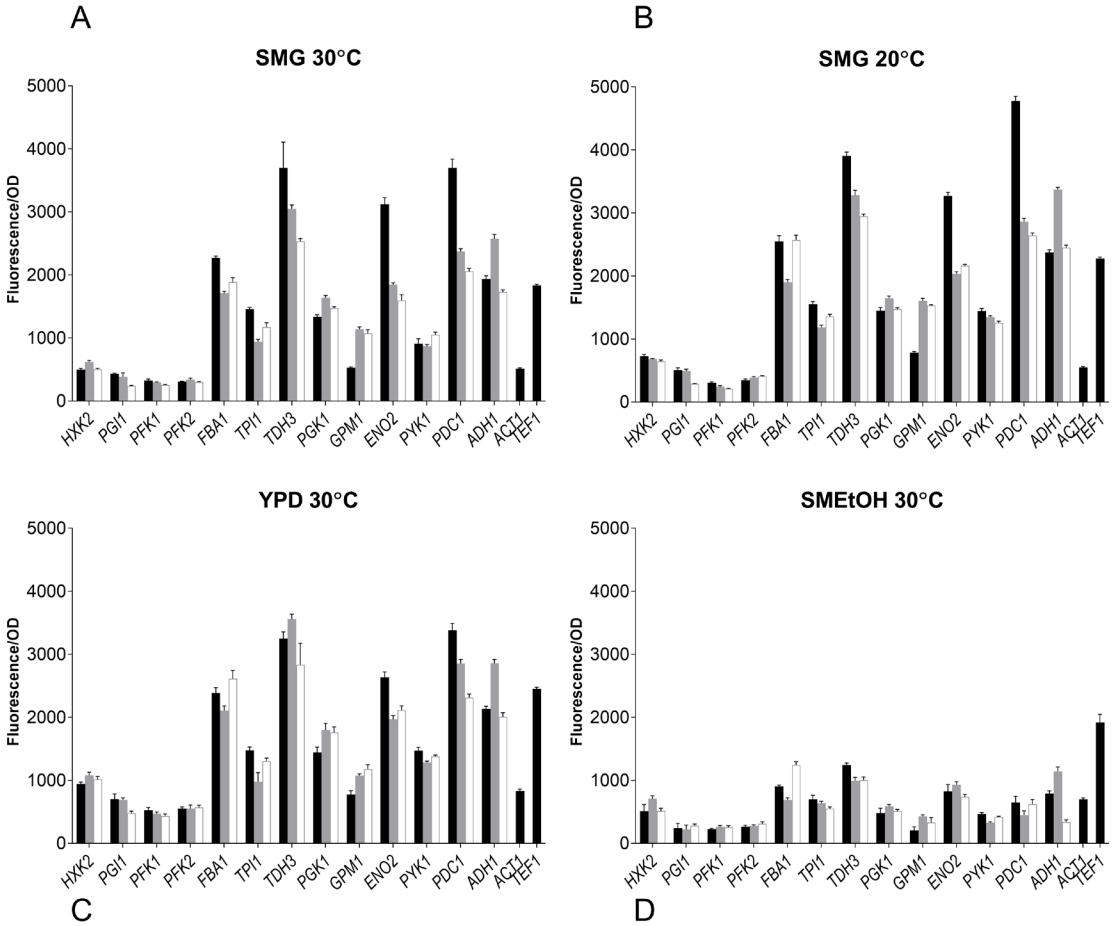


Figure 6 – Promoter activity of the major glycolytic promoters from *S. cerevisiae* (black)*, S. kudriavzevii* (grey) and *S. eubayanus* (white) expressing *mRuby2* in *S. cerevisiae.* During exponential growth in SMG **A),** SMG 20°C **B),** YPD **C)** and SMEtOH **D)** fluorescence and optical density were measured every 20 minutes and promoter activity was calculated as the slope of the linear regression between optical density and fluorescence. Two reference strains expressing *mRuby2* from the *ScTEF1* and *ScACT1* promoters were taken along in every plate. Error bars represent the standard deviation of the mean of six biological replicates.

Transferability and context-dependency of glycolytic and fermentative promoters within the *Saccharomyces* genus

The strain library grown in SMG at 30°C not only revealed that the *S. kudriavzevii* and *S. eubayanus* promoters could drive gene expression in *S. cerevisiae*, but also that their strength was remarkably similar to the strength of their *S. cerevisiae* orthologs (Figure 6). Additionally, two reporter strains expressing *mRuby2* from the constitutive *S. cerevisiae TEF1* and *ACT1* promoters were constructed and cultivated on all plates experiments. The activity of these two promoters was remarkably reproducible between independent culture replicates (Figure S4).

While, due to high data reproducibility, expression driven by *S. kudriavzevii* or *S. eubayanus* promoters was in most cases considered statistically different from the expression led by their *S. cerevisiae* orthologs (student t-test, $P<0.01$), differences in expression larger than 1.5-fold were rarely observed. Expression of *ENO2p* and *PDC1p* of *S. kudriavzevii* and *S. eubayanus* was lower than for their *S. cerevisiae* counterparts, while *SkGPM1p*, *SeGPM1p* and *SePYK1p* led to clearly higher expression levels than their *S. cerevisiae* homologs (Figure 6). These differences were not reflected in the transcript data (Figure 5). Conversely, the differential expression of *PFK1* and *TDH3* revealed by the RNAseq analysis was also found in the promoter transplantation study at SMG 30°C. Overall similarities and differences between the three species in transcript levels were mirrored by promoter activity.

To test the condition dependency of promoter activity, strains were tested under several culture conditions. YPD was used as rich medium, and ethanol was used as gluconeogenic carbon source (SMEtOH). Since *S. kudriavzevii* and *S. eubayanus* have a lower optimum growth temperature and hexokinase from *S. kudriavzevii* has been proposed to have a lower temperature optimum as compared to *S. cerevisiae* (*75*), the strains were also grown in SMG at 20°C. When grown in YPD and SMG at 20°C, all strains showed highly similar promoter activities as compared to cultures in SMG at 30°C even though the growth rates were different (SMG 30°C: 0.34 h^{-1}, SMG 20°C: 0.15 h^{-1}, YPD 30°C: 0.36 h^{-1}). However, during growth on ethanol (0.13 h^{-1}) promoter activity of the three species dropped tremendously as compared to glucose-grown cultures, in stark contrast with the fluorescence of the reference strains (*TEF1*p and *ACT1*p) that remained remarkably constant for all cultivation conditions. Nevertheless, also on SMEtOH *S. kudriavzevii* and *S. eubayanus* promoters showed expression levels very similar to their *S. cerevisiae* orthologs.

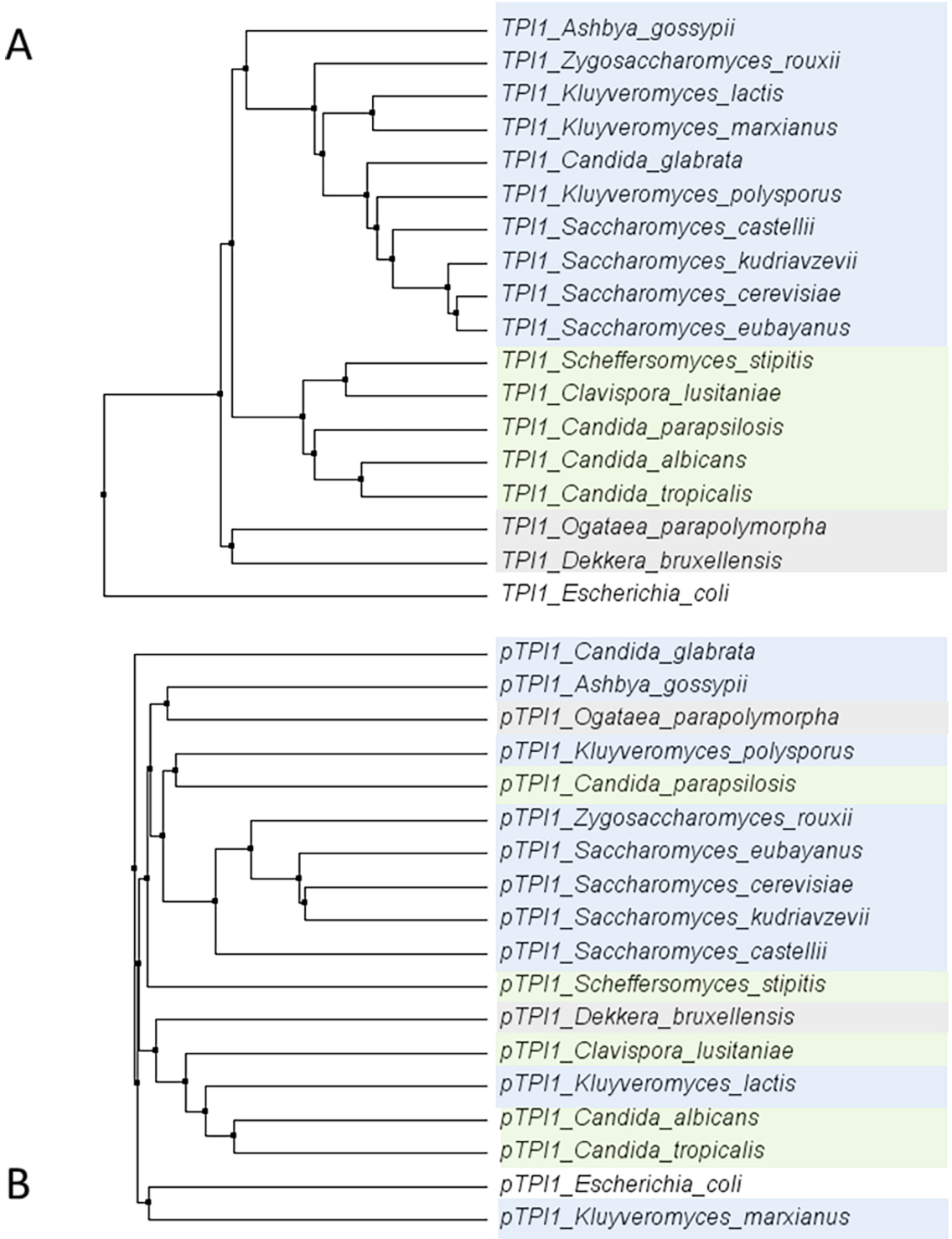


Figure 7 – Phylogenetic trees of the alignments of the *TPI1* **A)** genes (ORF) and **B)** promoters (800 bp) for a set of yeast species from the *Saccharomycotina* phylum. The coding regions are strongly conserved, whereas there is hardly any conservation among promoter regions. *E. coli* was used as an outgroup. Colour indicates groups as defined in Dujon (2010) (*62*).

Discussion

In this study we showed that the genetic makeup of the glycolytic and fermentative pathways is highly conserved among *S. cerevisiae, S. kudriavzevii* and *S. eubayanus*. For 11 out of 12 reactions, the exact same number of paralogs was found in the three species, reflecting that species divergence took place after whole genome and post-whole genome duplications. The only exception was the absence of the minor paralog *PDC6* in the *S. kudriavzevii* CR85 genome. In agreement with this observation, the presence of a pseudogene in *S. kudriavzevii* strains IFO1802 and ZP591 consisting of only about 15% of the full *PDC6* gene length has been reported (*76*). At the transcript level a strong conservation was also observed, suggesting that the classification between major and minor paralogs, confirmed in *S. cerevisiae* by mutant studies, could be extended to *S. kudriavzevii* and *S. eubayanus*. The slightly lower degree of conservation of minor paralogs (e.g. *GPM2, PYK2, ENO1, TDH1, TDH2, PDC6, ADH2, ADH4, ADH5*) is in line with the previously reported accelerated evolution of the *PYK2* and *ADH5* as compared to their *PYK1* and *ADH1* paralogs (*9*).

The glycolytic pathway is known to be highly conserved compared to most other pathways (*77, 78*). Recently it was shown that glycolytic coding regions from *E. coli* could replace the corresponding yeast genes (*79*). For promoter regions the conservation is in general lower as compared to coding regions, but a stronger conservation was found for the glycolytic promoters in the *Saccharomyces* genus than for other promoter regions (*80*). Combined with the remarkable conservation of binding sites for major transcriptional regulators (i.e. Rap1, Gcr1 and Abf1), these observations suggested a very similar transcriptional regulation of glycolytic and fermentative genes across the three species. Accordingly, transcriptome data showed a remarkable conservation in expression for the majority of glycolytic and fermentative genes in their native context. It is noteworthy that transcript levels of glycolytic and fermentative genes of these three diploid species were highly similar to the transcript levels of the haploid *S. cerevisiae* CEN.PK113-7D cultivated in the same condition as the *S. cerevisiae* diploid (*4*). The similarity in gene expression and the conservation of the main transcription factor binding sites in the three species suggested the possibility to introduce the promoters in *S. cerevisiae*, expecting similar regulation.

Until now a limited number of examples of heterologous glycolytic promoters driving gene expression in *S. cerevisiae* is available. Recently, it was shown that *S. kudriavzevii* glycolytic and fermentative promoters could drive gene expression in *S. cerevisiae* (*35*). More recently, it was shown that the *ADH2* promoter of several *Saccharomyces* species could drive gene expression in *S. cerevisiae* (*81*). Further, the glycolytic genes *PFK1,*

PFK2 and *PYK1* of the more distantly related yeast *Hanseniaspora uvarum*, expressed from their native promoters were shown to complement their *S. cerevisiae* orthologs (*82*).

To explore the conservation of glycolytic genes in a broader context, the sequence of the *TPI1* gene was compared across a set of 18 species within the *Saccharomycotina* subphylum (*62*). Within this subphylum, the coding region of *TPI1* was highly conserved (ranging from 64,7% – 96,3% identity to *S. cerevisiae*), while the promoters generally displayed a much weaker similarity (ranging from 28,5% - 71,1% identity to *S. cerevisiae*) (Figure 7). These observations are in line with studies reporting the loss of the gene encoding the Gcr1 transcription factor and the gain of new function by Rap1 in the CTG clade yeast *Candida albicans* (*83-85*). Indeed using the MEME suite motif discovery tool (*86*) gave only hits for Rap1 and Gcr1 motifs in the *Saccharomyces* family.

The present study shows the ability of all the major glycolytic promoters of *S. kudriavzevii* and *S. eubayanus* to drive gene expression in *S. cerevisiae* with similar strength and condition-dependency. Since many hybrids occur between *S. cerevisiae x S. eubayanus* and *S. cerevisiae x S. kudriavzevii,* it is not surprising that promoters are functional in *S. cerevisiae*. However, the similarity we found in promoter activities for most promoters transplanted to *S. cerevisiae* under different conditions is remarkable and indicates a strong conservation of the glycolytic regulatory mechanisms for *S. cerevisiae, S. kudriavzevii* and *S. eubayanus* (Figure S6). In general, these data do not correlate very well with the transcript data (Figure S7). This can most likely be explained by the relatively low dynamic measurement range of the plate reader compared to RNAseq, differences in cultivation conditions, length of promoters, choice of site for genomic integration (*87*) or differences in regulatory sequences in the promoters. During growth on ethanol a strong decrease in promoter activity was observed. This is in agreement with the previously reported drop in enzymatic activity of glycolysis during growth on ethanol (*88, 89*).

S. cerevisiae's proficiency in assembling and functionally expressing large (heterologous) pathways has propelled this yeast as preferred host for the production of complex molecules such as isoprenoids or opioids (*90, 91*). However, the successful expression of these pathways depends on the availability of suitable promoters. While *S. cerevisiae* has one of the most furbished molecular toolbox, the number of constitutive and well characterized promoters remains limited. Since *S. cerevisiae's* extremely efficient homologous recombination repeated usage of promoter sequences renders genetically unstable strains (*92*), this shortage of promoters presents a hurdle for extensive strain construction programs. While a lot of effort is invested in the design

of synthetic promoters and transcription amplifiers (*93-96*) using slightly distant but functional orthologous promoters presents an attractive alternative (*81, 97*). Usage of especially the *S. eubayanus* promoters, which are slightly more distant from *S. cerevisiae* than the *S. kudriavzevii* promoters, would reduce the length of the sequences being 100% identical to the native *S. cerevisiae* promoters. The minimum length which was found to be needed for efficient homologous recombination in *S. cerevisiae* was 30 bp with an optimal efficiency at a length of 60 bp or more (*92, 98*). In the *S. eubayanus* promoters, with one exception for the *PFK2* promoter, the longest sequence being identical to *S. cerevisiae* was found to be 34 bp. Usage of the *S. eubayanus* promoters would therefore substantially decrease the risk of instability and undesired recombination events during strain construction programs.

Conclusion

This study brings new insight in the genetic makeup and expression of glycolytic and fermentative genes in *S. eubayanus* and *S. kudriavzevii*. It also expands the molecular toolbox for *S. cerevisiae*, but also for its two relatives, with a set of strong, constitutive promoters. Furthermore, combining Illumina and Oxford Nanopore technologies, the present study offers a high quality sequence for *S. kudriavzevii* CR85, available from NCBI (PRJNA480800). Finally, the full set of transcript levels for the three diploid strains grown in tightly controlled conditions is available via GEO (See Material and Methods section) and can be mined to compare species-specific regulation of gene expression beyond the glycolytic and fermentative pathways.

Acknowledgements

We thank Rik Brouwer for his contribution to strain construction, Mark Bisschops for advice and help with fermentations and data analysis, Marijke Luttik for technical support and advice for the TECAN plate reader and flow cytometry analysis and Pilar de la Torre for performing the whole genome sequencing. Furthermore, we thank Jack Pronk for his advice on experimental design and Raúl Ortiz Merino for constructive comments on the manuscript. We thank Eladio Barrio Esparducer for kindly providing *S. kudriavzevii* CR85. This work was supported by a consolidator grant from the European Research Council (ERC).

Additional Material

Table S4 – Length in base pairs of promoters used in this study.

	S. cerevisiae	*S. kudriavzevii*	*S. eubayanus*
pHXK2	800	800	800
pPGI1	800	800	800
pPFK1	800	805	800
pPFK2	800	798	800
pFBA1	793	800	781
pTPI1	798	801	796
pTDH1	800	800	800
pTDH3	689	800	800
pPGK1	778	764	796
pGPM1	800	800	503
pENO2	800	800	800
pPYK1	811	800	807
pPDC1	800	800	800
pADH1	800	845	800

Table S5 – *S. kudriavzevii* sequencing.

The Nanopore assembly of the *S. kudriavzevii* CR85 sequencing data resulted in only 20 contigs, which represents a near 73-fold reduction in the number of contigs and captured an additional 200 kb as compared to IFO1802 assembly (*76*) which was Illumina sequenced. Except chromosomes 7,12 and 16 all chromosomes were assembled in single contigs with end-to-end coverage.

	***S. kudriavzevii* IFO1802**	***S. kudriavzevii* CR85**
Technology	Illumina	Nanopore
#Scaffolds/contigs	1455	20
N50 (Mbp)	0.151	0.86
Total assembly size (Mbp)	11.7	11.9

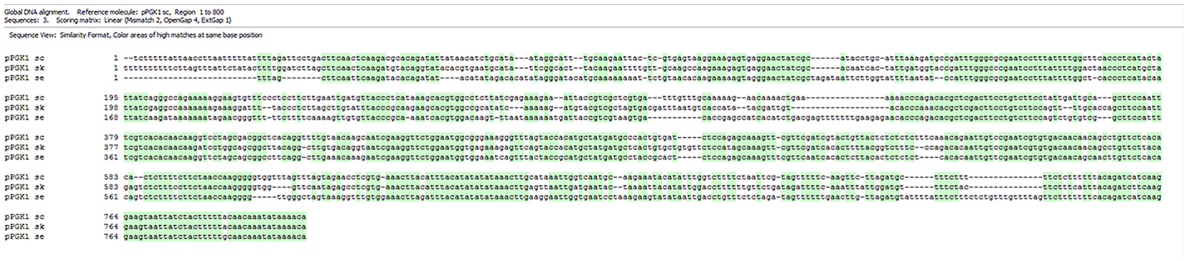


Figure S1 – Alignment of the *S. cerevisiae, S. kudriavzevii* and *S. eubayanus PGK1* promoters. 800 bp directly upstream of the ORF were used in the alignment. Green indicates nucleotides conserved in all three promoters.

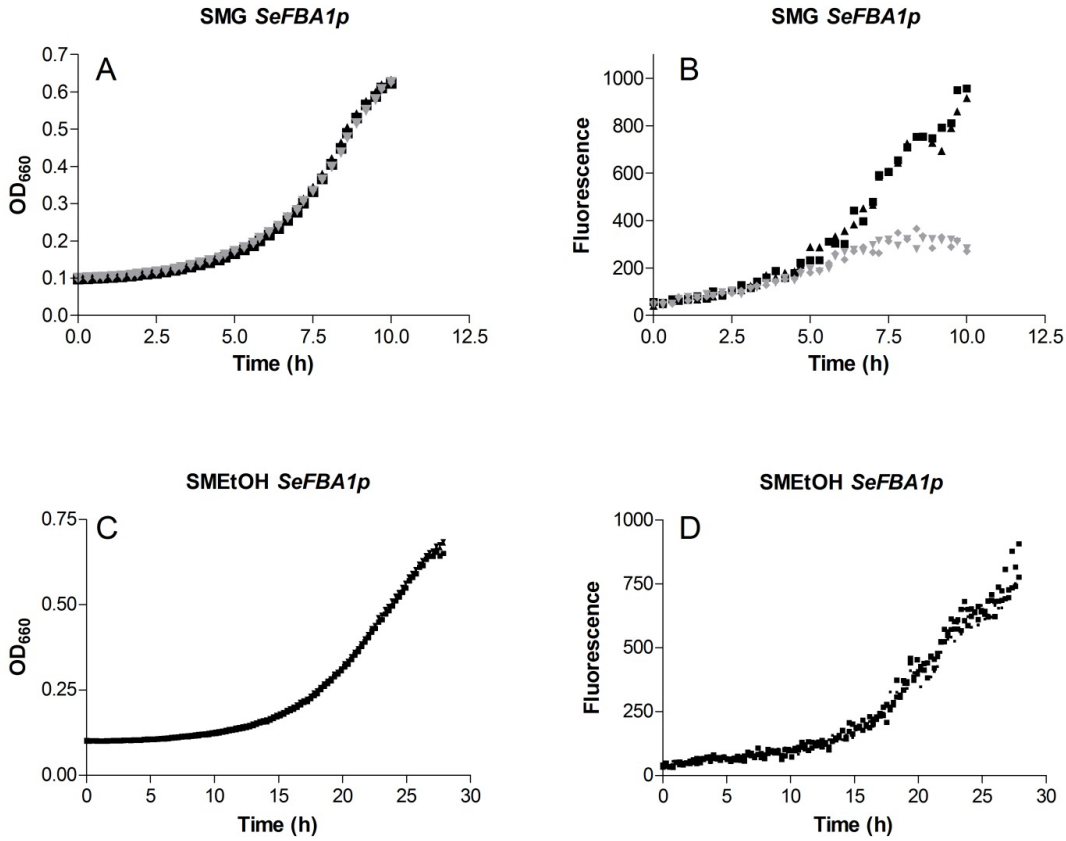


Figure S2 – Impact of oxygen availability and respiratory growth on mRuby2 fluorescence signal. Optical density and fluorescence intensity of strain IMX1097 expressing *mRuby2* from the *SeFBA1p* during growth in chemically defined medium with glucose (SMG) **(A,B)** and ethanol **(C,D)** as carbon source in 96-well plates. Cultures with intact seal are indicated in grey, and cultures with punctured seals are indicated in black. Two independent culture replicates are shown for each condition. During growth on ethanol there is a strict requirement for oxygen, but no limitation was observed in growth and fluorescence of the strains **(C,D).**

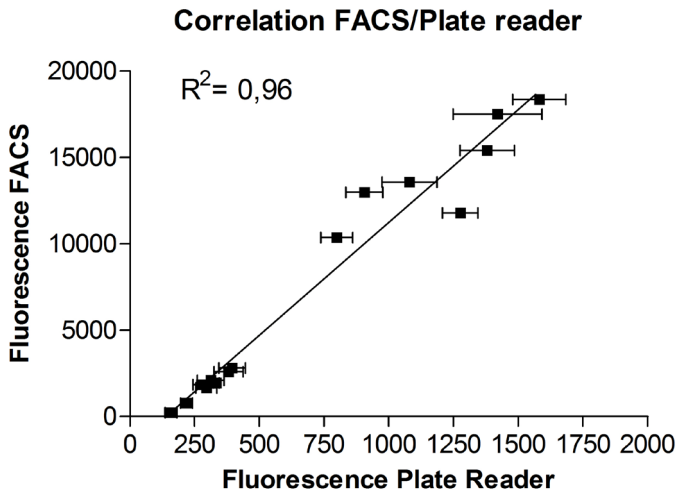


Figure S3 – Comparison of fluorescence intensity measurements by plate reader and by flow cytometry. Each data point represents the average fluorescence of five culture replicates from the same plate measured by the plate reader and the average fluorescence measurement of 10000 cells of the same cultures by flow cytometry, for a single strain. Results for 15 strains expressing mRuby2 with different promoter are shown. Error bars represent the SEM.

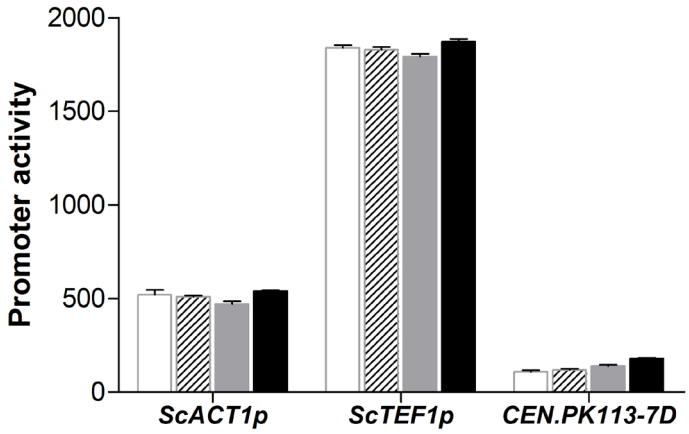


Figure S4 – Reproducibility of fluorescence measurements between 96-well plates. Fluorescence of strains expressing *mRuby2* driven by *S. cerevisiae ACT1* and *TEF1* promoters, and background fluorescence of the prototrophic non-fluorescent control strain CEN.PK113-7D. For each strain, the four bars represent the fluorescence measurements of four independent plate cultures. Bars and error bars represent the average and standard deviation of the mean of 6 biological replicates from the same plate, respectively.

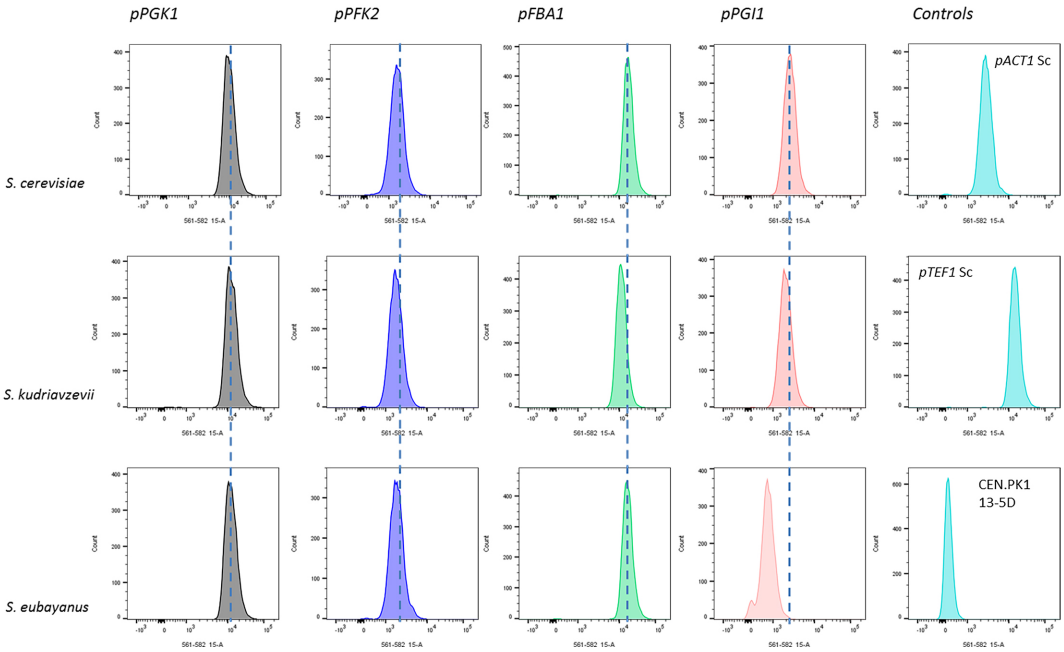


Figure S5 – Promoter strength characterization by flow cytometry.
The fluorescence of cells from exponentially growing cultures from a 96-well plate containing promoter-*mRuby2* reporter constructs was analysed by flow cytometry. For each strain the fluorescence profile for 10000 cells is shown. Strains expressing mRuby2 with the *ScACT1* and *ScTEF1* promoters and CEN.PK113-5D not expressing mRuby2 were taken as controls.

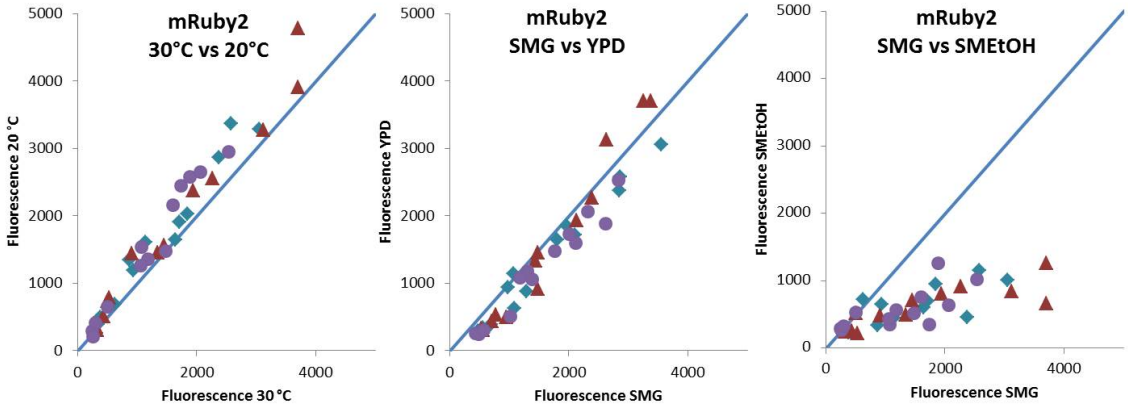


Figure S6 – Influence of growth conditions on the promoter strength of strains from the *S. cerevisiae* library expressing mRuby2 under the control of various glycolytic promoters. *S. cerevisiae* strains with glycolytic promoters from *S. cerevisiae* (triangle), *S. kudriavzevii* (diamond) and *S. eubayanus* (circle). Fluorescence was measured from whole cultures in plate reader.

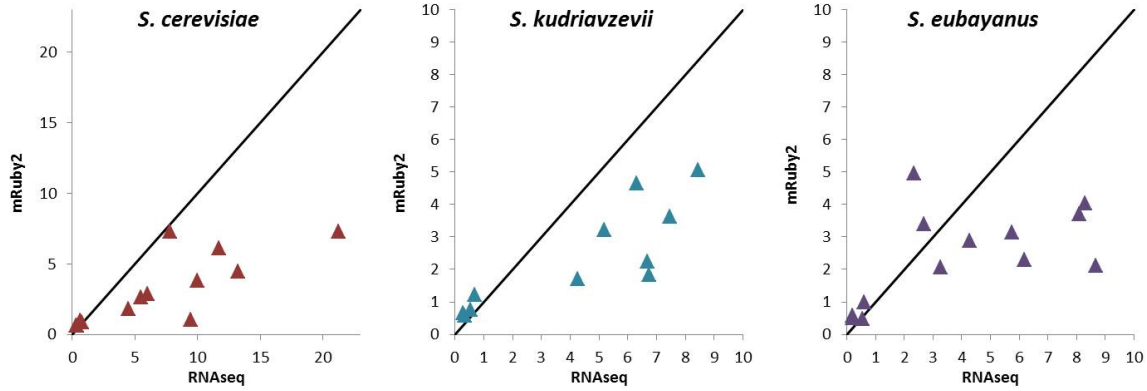


Figure S7 – Comparison of transcription measured by RNAseq and fluorescent reporter.
RNAseq data were normalized to *ACT1* transcript level and mRuby2 data to the fluorescence levels driven by the *ACT1* promoter.

References

1. D. Sicard, J.-L. Legras, Bread, beer and wine: yeast domestication in the *Saccharomyces* sensu stricto complex. *C. R. Biol.* **334**, 229-236 (2011).
2. J. A. Barnett, A history of research on yeasts 5: the fermentation pathway. *Yeast* **20**, 509-543 (2003).
3. J. A. Barnett, K. D. Entian, A history of research on yeasts 9: regulation of sugar metabolism. *Yeast* **22**, 835-894 (2005).
4. D. Solis-Escalante *et al.*, A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell* **14**, 804-816 (2015).
5. D. G. Fraenkel, The top genes: on the distance from transcript to function in yeast glycolysis. *Curr. Opin. Microbiol.* **6**, 198-201 (2003).
6. K. M. Carroll *et al.*, Absolute quantification of the glycolytic pathway in yeast: deployment of a complete QconCAT approach. *Mol. Cell. Proteomics* **10**, M111. 007633 (2011).
7. S. Ohno, Evolution by gene duplication. *Springer New York*, (1970).
8. K. H. Wolfe, D. C. Shields, Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708-713 (1997).
9. M. Kellis, B. W. Birren, E. S. Lander, Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617-624 (2004).
10. L. Kuepfer, U. Sauer, L. M. Blank, Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* **15**, 1421-1430 (2005).
11. G. C. Conant, K. H. Wolfe, Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol. Syst. Biol.* **3**, 129 (2007).
12. J. H. Van Heerden, F. J. Bruggeman, B. Teusink, Multi-tasking of biosynthetic and energetic functions of glycolysis explained by supply and demand logic. *Bioessays* **37**, 34-45 (2015).
13. A. Arvanitidis, J. J. Heinisch, Studies on the function of yeast phosphofructokinase subunits by *in vitro* mutagenesis. *J. Biol. Chem.* **269**, 8911-8918 (1994).
14. J. Heinisch, K. Vogelsang, C. P. Hollenberg, Transcriptional control of yeast phosphofructokinase gene expression. *FEBS Lett.* **289**, 77-82 (1991).
15. B. Peng, T. C. Williams, M. Henry, L. K. Nielsen, C. E. Vickers, Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microb. Cell Fact.* **14**, 91 (2015).
16. M. Ciriacy, Isolation and characterization of further cis-and trans-acting regulatory elements involved in the synthesis of glucose-repressible alcohol dehydrogenase (ADHII) in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **176**, 427-431 (1979).
17. V. M. Boer, J. H. De Winde, J. T. Pronk, M. D. Piper, The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J. Biol. Chem.* **278**, 3265-3274 (2003).
18. T. A. Knijnenburg *et al.*, 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 (2009).
19. G. C. Conant, K. H. Wolfe, Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* **9**, 938-950 (2008).
20. M. Fauchon *et al.*, Sulfur sparing in the yeast proteome in response to sulfur demand. *Mol. Cell* **9**, 713-723 (2002).
21. T. Replansky, V. Koufopanou, D. Greig, G. Bell, *Saccharomyces* sensu stricto as a model system for evolution and ecology. *Trends Ecol. Evol.* **23**, 494-501 (2008).
22. C. T. Hittinger, *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* **29**, 309-317 (2013).
23. S. Naseeb *et al.*, *Saccharomyces jurei* sp. nov., Isolation and genetic identification of a novel yeast species from *Quercus robur*. *Int. J. Syst. Evol. Microbiol.* **67**, 2046-2052 (2017).
24. F. N. Arroyo-López, R. Pérez-Torrado, A. Querol, E. Barrio, Modulation of the glycerol and ethanol syntheses in the yeast *Saccharomyces kudriavzevii* differs from that exhibited by *Saccharomyces cerevisiae* and their hybrid. *Food Microbiol.* **27**, 628-637 (2010).
25. I. Masneuf-Pomarède, M. Bely, P. Marullo, A. Lonvaud-Funel, D. Dubourdieu, Reassessment of phenotypic traits for *Saccharomyces bayanus* var. *uvarum* wine yeast strains. *Int. J. Food Microbiol.* **139**, 79-86 (2010).

26. Z. Salvadó *et al.*, Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Appl. Environ. Microbiol.* **77**, 2292-2302 (2011).
27. M. Hebly *et al.*, *S. cerevisiae× S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res.* **15**, fov005 (2015).
28. U. Bond, The genomes of lager yeasts. *Adv. Appl. Microbiol.* **69**, 159-182 (2009).
29. S. S. González, E. Barrio, J. Gafner, A. Querol, Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res.* **6**, 1221-1234 (2006).
30. S. S. González, E. Barrio, A. Querol, Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. *Appl. Environ. Microbiol.* **74**, 2314-2320 (2008).
31. D. Peris, C. A. Lopes, C. Belloch, A. Querol, E. Barrio, Comparative genomics among *Saccharomyces cerevisiae× Saccharomyces kudriavzevii* natural hybrid strains isolated from wine and beer reveals different origins. *BMC Genomics* **13**, 407 (2012).
32. H.-V. Nguyen, T. Boekhout, Characterization of Saccharomyces uvarum (Beijerinck, 1898) and related hybrids: assessment of molecular markers that predict the parent and hybrid genomes and a proposal to name yeast hybrids. *FEMS Yeast Res.* **17**, (2017).
33. C. Belloch, S. Orlic, E. Barrio, A. Querol, Fermentative stress adaptation of hybrids within the *Saccharomyces* sensu stricto complex. *Int. J. Food Microbiol.* **122**, 188-195 (2008).
34. D. Libkind *et al.*, Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. USA* **108**, 14539-14544 (2011).
35. N. G. Kuijpers *et al.*, Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. USA* **113**, 15060-15065 (2016).
36. K.-D. Entian, P. Kötter, 25 Yeast genetic strain and plasmid collections. *Methods in Microbiology* **36**, 629-666 (2007).
37. J. Van Dijken *et al.*, An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* **26**, 706-714 (2000).
38. J. F. Nijkamp *et al.*, De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* **11**, 36 (2012).
39. C. A. Lopes, E. Barrio, A. Querol, Natural hybrids of *S. cerevisiae x S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*. *FEMS Yeast Res.* **10**, 412-421 (2010).
40. C. Verduyn, E. Postma, W. A. Scheffers, J. P. Van Dijken, 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 (1992).
41. G. Bertani, Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* **186**, 595-600 (2004).
42. G. Bertani, Studies on lysogenis I.: The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293-300 (1951).
43. A. N. Salazar *et al.*, Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.* **17**, 10.1093/femsyr/fox1074 (2017).
44. M. E. Lee, W. C. DeLoache, B. Cervantes, J. E. Dueber, A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4**, 975-986 (2015).
45. R. D. Gietz, R. A. Woods, Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87-96 (2002).
46. M. D. Piper *et al.*, Reproducibility of Oligonucleotide Microarray Transcriptome Analyses an interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **277**, 37001-37008 (2002).
47. M. A. Swiat *et al.*, FnCpf1: a novel and efficient genome editing tool for Saccharomyces cerevisiae. *Nucleic Acids Res.* **45**, 12585-12598 (2017).
48. S. Koren *et al.*, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722-736 (2017).
49. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* **9**, e112963 (2014).
50. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).

51. C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
52. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
53. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215-ii225 (2003).
54. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
55. E. Baker *et al.*, The genome sequence of Saccharomyces eubayanus and the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* **32**, 2818-2831 (2015).
56. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
57. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
58. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511-515 (2010).
59. M. Goujon *et al.*, A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* **38**, W695-W699 (2010).
60. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
61. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).
62. B. Dujon, Yeast evolutionary genomics. *Nat. Rev. Genet.* **11**, 512-524 (2010).
63. X.-X. Shen *et al.*, Reconstructing the backbone of the *Saccharomycotina* yeast phylogeny using genome-scale data. *G3: Genes, Genomes, Genetics* **6**, 3927-3939 (2016).
64. C. T. Harbison *et al.*, Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104 (2004).
65. A. Chambers, E. A. Packham, I. R. Graham, Control of glycolytic gene expression in the budding yeast (*Saccharomyces cerevisiae*). *Curr. Genet.* **29**, 1-9 (1995).
66. C. M. Drazinic, J. B. Smerage, M. C. López, H. V. Baker, Activation mechanism of the multifunctional transcription factor repressor-activator protein 1 (Rap1p). *Mol. Cell. Biol.* **16**, 3187-3196 (1996).
67. T. Galitski, A. J. Saldanha, C. A. Styles, E. S. Lander, G. R. Fink, Ploidy regulation of gene expression. *Science* **285**, 251-254 (1999).
68. B.-Z. Li, J.-S. Cheng, M.-Z. Ding, Y.-J. Yuan, Transcriptome analysis of differential responses of diploid and haploid yeast to ethanol stress. *J. Biotechnol.* **148**, 194-203 (2010).
69. F. N. Arroyo-López, S. Orlić, A. Querol, E. Barrio, Effects of temperature, pH and sugar concentration on the growth parameters of *Saccharomyces cerevisiae, S. kudriavzevii* and their interspecific hybrid. *Int. J. Food Microbiol.* **131**, 120-127 (2009).
70. S. Pirt, Maintenance energy: a general model for energy-limited and energy-sufficient growth. *Arch. Microbiol.* **133**, 300-302 (1982).
71. R. Y. Tsien, The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509-544 (1998).
72. J. H. Davis, A. J. Rubin, R. T. Sauer, Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* **39**, 1131-1141 (2010).
73. D. Zeevi *et al.*, Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res.* **21**, 2114-2128 (2011).
74. L. Keren *et al.*, Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.* **9**, 701 (2013).
75. P. Gonçalves, E. Valério, C. Correia, J. M. de Almeida, J. P. Sampaio, Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One* **6**, e20739 (2011).
76. D. R. Scannell *et al.*, The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces* sensu stricto genus. *G3: Genes, Genomes, Genetics* **1**, 11-25 (2011).
77. L. A. Fothergill-Gilmore, P. A. Michels, Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* **59**, 105-235 (1993).
78. K. A. Webster, Evolution of the coordinate regulation of glycolytic enzyme genes by hypoxia. *J. Exp. Biol.* **206**, 2911-2922 (2003).
79. A. H. Kachroo *et al.*, Systematic bacterialization of yeast genes identifies a near-universally swappable pathway. *eLife* **6**, e25093 (2017).

80. Z. Kuang, S. Pinglay, H. Ji, J. D. Boeke, Msn2/4 regulate expression of glycolytic enzymes and control transition from quiescence to growth. *Elife* **6**, (2017).
81. C. J. Harvey *et al.*, HEx: a heterologous expression platform for the discovery of fungal natural products. *bioRxiv*, 247940 (2018).
82. A.-K. Langenberg *et al.*, 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.*, e01580-01517 (2017).
83. C. Askew *et al.*, Transcriptional regulation of carbohydrate metabolism in the human pathogen *Candida albicans*. *PLoS Path.* **5**, e1000612 (2009).
84. H. Lavoie *et al.*, Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol.* **8**, e1000329 (2010).
85. M. T. Weirauch, T. R. Hughes, Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet.* **26**, 66-74 (2010).
86. T. L. Bailey *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202-W208 (2009).
87. D. B. Flagfeldt, V. Siewers, L. Huang, J. Nielsen, Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* **26**, 545-551 (2009).
88. H. Peter Smits *et al.*, Simultaneous overexpression of enzymes of the lower part of glycolysis can enhance the fermentative capacity of *Saccharomyces cerevisiae*. *Yeast* **16**, 1325-1334 (2000).
89. P. Van Hoek, J. P. Van Dijken, J. T. Pronk, Effect of specific growth rate on fermentative capacity of baker's yeast. *Appl. Environ. Microbiol.* **64**, 4226-4233 (1998).
90. C. J. Paddon *et al.*, High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528-532 (2013).
91. S. Galanie, K. Thodey, I. J. Trenchard, M. F. Interrante, C. D. Smolke, Complete biosynthesis of opioids in yeast. *Science* **349**, 1095-1100 (2015).
92. P. Manivasakam, S. C. Weber, J. McElver, R. H. Schiestl, Micro-homology mediated PCR targeting in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **23**, 2799-2800 (1995).
93. H. Redden, H. S. Alper, The development and characterization of synthetic minimal yeast promoters. *Nat. Common.* **6**, 7810 (2015).
94. A. Rantasalo *et al.*, Synthetic transcription amplifier system for orthogonal control of gene expression in *Saccharomyces cerevisiae*. *PloS One* **11**, e0148320 (2016).
95. F. Machens, S. Balazadeh, B. Mueller-Roeber, K. Messerschmidt, Synthetic promoters and transcription factors for heterologous protein expression in *Saccharomyces cerevisiae*. *Front. Bioeng. Biotechnol.* **5**, 63 (2017).
96. G. Naseri *et al.*, Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **6**, 1742-1756 (2017).
97. M. Naesby *et al.*, Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* **8**, 45 (2009).
98. S.-b. Hua, M. Qiu, E. Chan, L. Zhu, Y. Luo, Minimum length of sequence homology required for *in vivo* cloning by homologous recombination in yeast. *Plasmid* **38**, 91-96 (1997).

Chapter 4

Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences

Francine J. Boonekamp#, Sofia Dashko#, Donna Duiker, Thies Gehrmann, Marcel van den Broek, Maxime den Ridder, Martin Pabst, Vincent Robert, Thomas Abeel, Eline D. Postma, Jean-Marc Daran, Pascale Daran-Lapujade

# Authors contributed equally to this work

Published in ACS Synthetic Biology *9,* 1361–1375 (2020)

Abstract

The construction of powerful cell factories requires intensive and extensive remodelling of microbial genomes. Considering the rapidly increasing number of these synthetic biology endeavours, there is an increasing need for DNA watermarking strategies that enable the discrimination between synthetic and native gene copies. While it is well documented that codon usage can affect translation, and most likely mRNA stability in eukaryotes, remarkably few quantitative studies explore the impact of watermarking on transcription, protein expression and physiology in the popular model and industrial yeast *Saccharomyces cerevisiae*. The present study, using *S. cerevisiae* as eukaryotic paradigm, designed, implemented and experimentally validated a systematic strategy to watermark DNA with minimal alteration of yeast physiology. The thirteen genes encoding proteins involved in the major pathway for sugar utilization (i.e glycolysis and alcoholic fermentation) were simultaneously watermarked in a yeast strain using the previously published pathway swapping strategy. Carefully swapping codons of these naturally codon optimized, highly expressed genes, did not affect yeast physiology and did not alter transcript abundance, protein abundance and protein activity besides a mild effect on Gpm1. The markerQuant bioinformatics method could reliably discriminate native from watermarked genes and transcripts. Furthermore, presence of watermarks enabled selective CRISPR/Cas genome editing, specifically targeting the native gene copy while leaving the synthetic, watermarked variant intact. This study offers a validated strategy to simply watermark genes in *S. cerevisiae.*

Introduction

A DNA watermark is a unique synthetic nucleotide sequence that enables the identification and traceability of its carrier when applying PCR amplification and sequencing techniques. Application of the watermarks in living organisms started recently with as purpose to protect R&D investments, to create an information storage source or to enable traceability of pathogenic or endangered species (*1-3*). Literature reports successful embedding and subsequent detection of the watermarks in DNA strands *in vitro* (*4*), as well as *in vivo* using several model microorganisms (i.e. *Bacillus subtilis*, *Escherichia coli*, *Saccharomyces cerevisiae*, *Mycoplasma mycoides* and *Mycoplasma capricolum*), plants and viruses (*1, 2, 5-11*). All these studies focused on a single locus for the watermark introduction, with a few notable exceptions. First, the *Mycoplasma* genome *de novo* synthesis in which four large watermarks (ca. 1kb) were introduced to enable the differentiation between the natural and synthetic copies of the *Mycoplasma* genome (*5*). Second, the Synthetic Yeast 2.0 (Sc2.0) project, where several approximately 28 bp regions of each open reading frame were recoded to distinguish synthetic from native genes by PCR (*12*). Lastly, the recoding of the *E.coli* genome, such that it uses 61 instead of 64 codons (*10*). The successes of these projects reveal the potential of the watermarks for future development in synthetic biology, particularly during large-scale genome remodeling projects, where tagging the synthetic gene copies can enable the discrimination between synthetic and native homologs. For instance, Kuijpers and coworkers recently reported the pathway swapping strategy that enables to redesign large, native essential pathways (*13*). Pathway swapping was demonstrated on the glycolytic and fermentation pathways of *S. cerevisiae,* involving 12 catalytic steps encoded by 26 genes. After a first genetic reduction leading to a minimal glycolysis set of 13 genes (*14*), a second, synthetic set of these 13 genes was integrated in a single locus on chromosome IX. Subsequently, the native copies of these 13 glycolytic genes were removed from their original chromosomal loci, leading to SwYG, a yeast strain with a single locus, minimal glycolytic pathway. However, the presence of two identical gene copies for all glycolytic genes during the strain construction process led to complications. Firstly, in this intermediate strain carrying both native, scattered and synthetic, co-localized glycolytic genes, removal of the native gene copies without harming the synthetic, identical copies integrated on chromosome IX, was challenging. Secondly, expression of the native and synthetic genes could not be measured and compared. Both problems can easily be addressed by embedding watermarks in the synthetic genes. When judiciously placed in Protospacer Adjacent Motifs (PAM), watermarks can disable CRISPR/Cas editing in the synthetic genes (*15*). When designed

in coding regions (CDS), watermarks can be used to identify native from watermarked mRNA molecules.

Whether inserted in coding or non-coding regions, the major downside of watermarks is the risk of unintended changes in the host physiology. Watermarking in coding regions is potentially less challenging as watermarks can be embedded in the CDS as silent mutation, taking advantage of the redundancy of the genetic code encompassing 61 codons for only 20 amino acids. However, while 'silent' or synonymous mutations in CDS do not affect the amino acid sequence of the corresponding protein, they can alter cells at different levels. Codons can be classified as optimal and non-optimal based on their frequency in the genome and the abundance of tRNAs with complementary anticodons (*16-18*). It is now well established that cells use codon optimality to tune protein expression. Highly expressed genes, such as genes encoding the highly abundant glycolytic proteins, are enriched for optimal codons (*19, 20*). Furthermore, by tuning the translation rate, codon optimality regulates the co-translational folding of polypeptides and plays a role in shaping proteins conformational states (*21-24*). More recently, it has been shown that codon optimality also modifies mRNA structure, splicing and stability (*25-28*). Codon optimality preservation is therefore an important criterion to consider when introducing watermarks without causing undesirable changes in gene function. There is however little known about the impact of watermarking on cell physiology, and remarkably few studies are dedicated to *Saccharomyces cerevisiae*, a microbe intensively used in synthetic biology developments (*29, 30*). Heider and Barnekow demonstrated that watermarking of *VAM7* in *S. cerevisiae* did not affect the vacuolar function of the corresponding protein (*6*). Liss and co-workers expressed a watermarked GFP in *S. cerevisiae* and showed minimal impact on GFP protein by Western blotting (*7*). In the Sc2.0 project in every ORF larger than 500 bp at least two 19-28 bp PCRtags were introduced which were recoded approximately 33%-60%. Every strain with a native chromosome replaced by a synthetic version showed no or minor fitness defects and transcript profiling showed only few genes changed in expression (*9, 12, 31-36*). Whether these transcript changes originated from the PCRtags was not always investigated and it is unclear whether these PCRtags allow discrimination between native and synthetic mRNAs when both are present in the cell. Therefore, there remains a strong need for studies proposing a watermarking strategy with the ability to distinguish between native and synthetic DNA and mRNA, validated by a systematic, quantitative exploration of the impact of watermarking on transcription, translation and general physiology (*37*).

To fill this knowledge gap, using *S. cerevisiae* as eukaryotic paradigm, this study designed, implemented and experimentally validated a systematic approach to

watermark DNA with minimal alteration of yeast physiology. The impact of simultaneously watermarking 13 genes encoding abundant proteins involved in the major pathway for sugar utilization (i.e glycolysis and alcoholic fermentation) on metabolism, transcriptome and enzyme activity was explored using batch cultures in tightly controlled bioreactors. Watermarked transcripts were segregated from native ones using the karyollelle specific expression detection method (*38*). Finally, the ability of watermarks to protect synthetic genes from CRISPR/Cas9 DNA editing was evaluated.

Material and Methods

Strains and cultivation conditions

The *S. cerevisiae* strains used in the study belong to CEN.PK family (*39-41*) and are listed in Table 1. Liquid cultures were grown in 500 ml shake flasks filled with 100 ml of medium at 30° C with 200 rpm agitation. Complex media (further referred to as YPD) contained 10 $g.L^{-1}$ yeast extract, 20 $g.L^{-1}$ peptone and 20 $g.L^{-1}$ glucose. Synthetic minimal medium (further referred as SMG) consisted of 3 g $g.L^{-1}$ KH_2PO_4, 0.5 $g.L^{-1}$ $MgSO_4 \cdot 7H_2O$, 5 $g.L^{-1}$ $(NH_4)_2SO_4$, 1 $mL.L^{-1}$ of a trace element solution, and 1 $mL.L^{-1}$ of a vitamin solution as previously described[13] and supplemented with 20 $g.L^{-1}$ glucose. For solid medium, 20 $g.L^{-1}$ agar was added prior autoclaving. When selection in SMG was required, $(NH_4)_2SO_4$ was replaced with 3 $g.L^{-1}$ K_2SO_4 and 2.3 $g.L^{-1}$ filter-sterilized urea, and the medium was supplemented with 200 mg.L-1 of G418, hygromycin B or 10 mM acetamide (*42, 43*). For the counterselection purpose, 1 $mg.mL^{-1}$ 5-FOA (Zymo Research, Irvine, US) was added to SMG supplemented with uracil (150 $mg.L^{-1}$ (*42*)). For plasmid propagation, *E. coli* XL1-Blue cells (Agilent Technologies, Santa Clara, CA, USA) were grown in Lysogeny broth (LB) medium supplied with 100 mg L ampicillin or 25 $mg.L^{-1}$ chloramphenicol at 37°C with 180 rpm agitation. Yeast and bacterial frozen stocks were prepared by addition of 30% (v/v) glycerol to exponentially growing cultures. Strain aliquots were stored at –80°C.

Table 1 – Strains used in this study.

Strains characterized in this study

Strain name	Genotype	Source and description
IMX1770	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 glk1::Sphis5Δ::(pGAL1-I-SceI-tCYC1) can1::(ARS418$_{CAN1\ AH}$ FBA1_*$_{AH\ H}$ TPI1_*$_{HP}$ PGK1_*$_{PQ}$ ADH1_*$_{QN}$ PYK1_*$_{NO}$ TDH3_*$_{OA}$ ENO2_*$_{AB}$ HXK2_*$_{BC}$ PGI1_*$_{CD}$ PFK1_*$_{DJ}$ PFK2_*$_{J\ BP}$ HIS3$_{BP\ L}$ GPM1_*$_{LM}$ PDC1_*$_{M\ AR}$ ARS1211$_{AR\ CAN1}$) sga1::KlURA3*	This study. Prototrophic strain with watermarked single locus glycolysis (WMG strain). Derived from IMX1717
IMX1771	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 glk1::Sphis5Δ::(pGAL1-I-SceI-tCYC1) can1::(ARS418 $_{CAN1\ AH}$ FBA1 $_{AH\ H}$ TPI1 $_{HP}$ PGK1 $_{PQ}$ ADH1 $_{QN}$ PYK1 $_{NO}$ TDH3 $_{OA}$ ENO2 $_{AB}$ HXK2 $_{BC}$ PGI1 $_{CD}$ PFK1 $_{DJ}$ PFK2 $_{J\ BP}$ HIS3 $_{BP\ L}$ GPM1 $_{LM}$ PDC1 $_{M\ AR}$ ARS1211 $_{AR\ CAN1}$) sga1::KlURA3*	This study. Prototrophic strain with native single locus glycolysis (NG strain). Derived from IMX1747
IMX2028	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 glk1::Sphis5Δ::(pGAL1-I SceI-tCYC1) can1::(ARS418 $_{CAN1\ AH}$ FBA1 $_{AH\ H}$ TPI1 $_{HP}$ PGK1 $_{PQ}$ ADH1 $_{QN}$ PYK1 $_{NO}$ TDH3 $_{OA}$ ENO2 $_{AB}$ HXK2 $_{BC}$ PGI1 $_{CD}$ PFK1 $_{DJ}$ PFK2 $_{J\ BP}$ HIS3 $_{BP\ L}$ GPM1 $_{LM}$ PDC1 $_{M\ AR}$ ARS1211 $_{AR\ CAN1}$) sga1::(ARS418$_{sga1\ AH}$ FBA1_*$_{AH\ H}$ TPI1_*$_{HP}$ PGK1_*$_{PQ}$ ADH1_*$_{QN}$ PYK1_*$_{NO}$ TDH3_*$_{OA}$ ENO2_*$_{AB}$ HXK2_*$_{BC}$ PGI1_*$_{CD}$ PFK1_*$_{DJ}$ PFK2_*$_{J\ BP}$ HIS3$_{BP\ L}$ GPM1_*$_{LM}$ PDC1_*$_{M\ AR}$ ARS1211$_{AR\ sga1}$)*	This study. Prototrophic strain with native and watermarked single locus glycolysis (DG strain). Derived from IMX1748

Strains used as starting point or intermediate in the construction of the above strains

CEN.PK113-7D	*MATa MAL2–8c SUC2*	Control strain[46, 47]
IMX589	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1GH TPI1HP PGK1PQ ADH1QN PYK1NO TDH3OA ENO2AB HXK2BC PGI1CD PFK1DJ PFK2JK AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-SYN$_{MF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*	Starting strain for all construction work[13]. Contains a SinLoG in Chr. IX, with variable promoter and terminator length. Uracil auxotroph.

IMX1338	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-$_{SYNMF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 glk1::Sphis5Δ::(pGAL1-I SceI-tCYC1)*	This study. Derived from IMX589. Contains a SinLoG in Chr. IX. Uracil and histidine auxotroph.
IMX1717	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-$_{SYNMF}$) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 glk1::Sphis5Δ::(pGAL1-I SceI-tCYC1) can1::(ARS418$_{AH}$ FBA1_*$_{AH.H}$ TPI1_*$_{HP}$ PGK1_*$_{PQ}$ ADH1_*$_{QN}$ PYK1_*$_{NO}$ TDH3_*$_{OA}$ ENO2_*$_{AB}$ HXK2_*$_{BC}$ PGI1_*$_{CD}$ PFK1_*$_{DJ}$ PFK2_*$_{J.BP}$ HIS3$_{BP.L}$ GPM1_*$_{LM}$ PDC1_*$_{M.AR}$ ARS1211$_{AR}$)*	This study. Derived from IMX1338. Strain with native (variable prom & term length) and watermarked SinLoG in Chr. IX and V respectively. Histidine auxotroph.
IMX1747	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{GH}$ TPI1$_{HP}$ PGK1$_{PQ}$ ADH1$_{QN}$ PYK1$_{NO}$ TDH3$_{OA}$ ENO2$_{AB}$ HXK2$_{BC}$ PGI1$_{CD}$ PFK1$_{DJ}$ PFK2$_{JK}$ AmdSYM$_{KL}$ GPM1$_{LM}$ PDC1-$_{SYNMF}$) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 glk1::Sphis5Δ::(pGAL1-I SceI-tCYC1) can1::(pGAL1-I-SceI-tCYC1) can1::(ARS418 $_{CAN1\ AH}$ FBA1 $_{AH\ H}$ TPI1 $_{HP}$ PGK1 $_{PQ}$ ADH1 $_{QN}$ PYK1 $_{NO}$ TDH3 $_{OA}$ ENO2 $_{AB}$ HXK2 $_{BC}$ PGI1 $_{CD}$ PFK1 $_{DJ}$ PFK2 $_{J\ BP}$ HIS3 $_{BP\ L}$ GPM1 $_{LM}$ PDC1 $_{M\ AR}$ ARS1211 $_{AR\ CAN1}$)*	This study. Derived from IMX1338. Strain with native SinLoG with variable prom & term length in Chr. IX and native SinLoG with standardized prom and term length in Chr. V. Histidine auxotroph.
IMX1748	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 glk1::Sphis5Δ::(pGAL1-I-SceI-tCYC1) can1::(ARS418 $_{CAN1\ AH}$ FBA1 $_{AH\ H}$ TPI1 $_{HP}$ PGK1 $_{PQ}$ ADH1 $_{QN}$ PYK1 $_{NO}$ TDH3 $_{OA}$ ENO2 $_{AB}$ HXK2 $_{BC}$ PGI1 $_{CD}$ PFK1 $_{DJ}$ PFK2 $_{J\ BP}$ HIS3 $_{BP\ L}$ GPM1 $_{LM}$ PDC1 $_{M\ AR}$ ARS1211 $_{AR\ CAN1}$)*	This study. Derived from IMX1771. Strain with watermarked SinLoG in Chr. IX and native SinLoG with standardized prom and term length in Chr. V. Uracil auxotroph.

Molecular biology techniques

PCR reactions for diagnostic purposes were performed using DreamTaq DNA polymerase Master Mix (Thermo Fisher Scientific, Walthman, MA, USA) according to the manufacturer's instructions. For high fidelity PCR reactions, Phusion® High-Fidelity DNA polymerase (Thermo Fisher Scientific) was used following the supplier's manual. Oligonucleotides of desalted or PAGE quality, depending on the purpose, were purchased from Sigma Aldrich (St Louis, MO, USA). DNA fragments were resolved in agarose gels and purified using PCR clean-up kit from the reaction mixture (Sigma Aldrich, St Louis, MO, USA) or excised from the agarose gel and purified using Zymoclean gel purification kit (Zymo Research, Irvine, CA, USA) when required. Circular templates were removed by applying DpnI enzyme restriction according to the producer's manual (Thermo Fisher Scientific).

Plasmids were isolated from *E. coli* cultures using Sigma GenElute Plasmid kit (Sigma-Aldrich, St Louis, MO, USA). *E. coli* transformations were performed using chemical competent XL-1 Blue cells (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer instructions. Golden Gate Assembly was performed as previously described (*44*) using equimolar concentrations of 20 fmol for each fragment. For a 10 µL reaction mixture 1 µL T4 DNA ligase buffer (Thermo Fisher Scientific), 0.5 µL T7 DNA ligase (NEB New England Biolabs, Ipswich, MA), and 0.5 µL of either FastDigest Eco31I (BsaI) or BsmBI (NEB) were added.

Gibson Assembly was performed using Gibson Assembly Master Mix (New England Biolabs, Ipswich, MA) according to the manufacturer's protocol.

All plasmids are reported in Table S4 and primers in Table S7.

***In silico* design of the watermarks**

Watermarks were introduced in the genes of interest according to the guidelines described in Box 1 using the Clone Manager software.

Watermarked CDS were ordered as a synthetic gene from GenArt (Thermo Fisher, Regensburg). The list of synthesized plasmids encoding watermarked CDS (pGGKp137 to pGGKp150) can be found in Table S4.

The change in codon usage in a gene caused by watermarking was calculated as:

$$\sum_{i=1}^{n}(|fraction\ of\ native\ codon_i - fraction\ of\ watermarked\ codon_i|)$$

Where *i* represents each codon substitution in a gene.

Box 1: watermarking strategy

- Mutating the coding region only
- Leaving first 101 bp of the ORF intact
- Mutating 2 regions of 100 bp, one in the middle and one towards the end of the ORF
- Introducing at least 5 SNPs in each 100 bp stretch, evenly spread along the sequence.
- Only targeting codons with 4 to 6 alternative triplets
- Replacing native codon with the triplet with the most similar abundance
- Setting maximum variation in abundance between native and watermarked triplet to 20%

Construction of libraries encoding transcriptional units of watermarked glycolysis

The sequences of the watermarked genes, promoters (800bp) and terminators (300bp) were ordered from GeneArt (Thermo Fisher, Regensburg, Germany). For compatibility with Golden Gate Cloning, the sequences were ordered flanked with BsaI and BsmBI restriction sites. The promoters and terminators were delivered by GeneArt subcloned in the entry vector pUD565 and for the watermarked genes the subcloning into pUD565 was done in house using BsmBI Golden Gate cloning. An exception was made for *pTDH3*, *pPGK1*, *tPGK1, tENO2* and *tADH1* which were amplified from genomic DNA of CEN.PK113-7D using primers with flanks containing BsaI restriction sites listed in Table S7. Subsequently, the assembly of the promoter, gene and terminator was done in the preassembled vector pGGKd012 using Golden Gate cloning as described in the previous section. pGGKd012 was assembled from the Yeast toolkit (*44*) plasmids pYTK-002, 047, 072, 078, 081 and 083 (Table S4). Correct plasmid assembly was verified by enzyme digestion with either BsaI, BsmBI (New England Biolabs) or FastDigest enzymes (Thermo Fisher Scientific) following the manufacturer's instructions.

Construction of gRNA plasmids used in the study

The guide RNA (gRNA) plasmids pUDR413 and pUDR529 for the yeast strain construction were designed and constructed according to Mans et al. (2015) (*45*). gRNA targets were selected using the Yeastriction tool (*45*) in case of pUDR413, or designed

manually for the *K. lactis URA3* target in plasmid pUDR529. For pUDR413, the 2 µm fragment was constructed in two parts using the primer 6131 and 5975 and primer 6296 together with 5941 using pROS12 as a template. For pUDR529, the 2 µm fragment was obtained by PCR using primer 14549 and pROS12 as a template. The backbone for pUDR413 was amplified with primers 6005 and 6006 using pROS13 as a template, while for pUDR529 same primer pair was used to amplify the backbone from pROS12. For both plasmids, 100 ng of each purified fragment was used in the Gibson Assembly and correct plasmid assembly was verified with the primers 3841, 5941 and 6070 in case of pUDR413 and 4034 and 5941 for pUDR529.

The guide RNA plasmids for selective native copy gene removal of *TPI1* and *PYK1*, named pUDR531 and pUDR532 respectively (Table S4), were constructed as described in Mans et al. (2015) (*45*) with the modifications regarding the design of the gRNA. gRNAs were designed manually to target the native CDS containing a PAM which was removed in the watermarked copy of the CDS. Each gRNA was ordered as a primer (Table S7, primers 14515, 14517, 14519, 14521). The 2 µm fragment for four gRNA plasmids was obtained by PCR using corresponding gRNA primer (Table S7, primers 14515, 14517, 14519, 14521) and pROS13 as a template. The backbone for the four plasmids was obtained by amplification with primers 6005 and 6006 using pROS12 as a template. For the assembly, 100 ng of purified backbone and gRNA fragments were used in the Gibson Assembly and correct plasmid assemblies were verified with the primers 3841 and 5941 in combination with gRNA specific primers listed in Table S7.

Construction of SwYG strains with native and watermarked glycolysis and Double glycolysis strain

A schematic overview of the strain construction approach is shown in Figure 3. All yeast transformations were performed according to Gietz and Woods (2002) (*46*) . For highly efficient targeted integration CRISPR/Cas9 mediated editing was applied. To this end, 350 ng of a plasmid carrying a corresponding guide RNA (further gRNA) was transformed into the yeast strain together with a purified PCR fragment (150 fmol) containing 60 bp homology to the integration site and acting as donor DNA (Primers list in Table S7). gRNA plasmids and the donor DNA were specific for each strain construction step and will be specified below. When donor DNA was consisting of multiple fragments, 60 bp sequences for homologous recombination (SHR) were flanking each of the fragments to enable *in vivo* assembly by homologous recombination. PCR fragments for the native SinLoG genes and for *ARS418, ARS1211* and *HIS3* were obtained using CEN.PK113-7D genomic DNA as a template, while fragments encoding the watermarked SinLoG were amplified from plasmids encoding the corresponding transcriptional units (Table S4, Table S7).

To obtain a double auxotrophic host strain named IMX1338, the *Schizosaccharomyces pombe HIS5* gene previously inserted in the *glk1* locus of IMX589 containing the SinLoG in chromosome IX (*13*) was replaced by the *I-SceI* expression cassette (p*GAL1* – *I-SceI* – t*ADH1*), which was amplified from the plasmid pUDC073 (primers 10708 and 10709). The replacement was mediated by a Cas9 gRNA plasmid assembled *in vivo* from two PCR fragments amplified from the pMEL10 plasmid using primers 6005 and 6006 in combination with 10904 (gRNA primer). Transformants were selected on SMG, and the gRNA plasmid with *URA3* marker was removed by two sequential restreaks on SMG with 5-FOA. The correct genotype was confirmed by diagnostic PCR using primers 6190+1525 and 1553+6189 and later by whole genome sequencing. To construct IMX1717 and IMX1747, IMX1338 was transformed with the p426-SNR52p-gRNA.CAN1.Y-SUP4t plasmid targeting the *CAN1* locus (*47*), and PCR fragments of the 13 native or watermarked SinLoG glycolytic genes together with *ARS418*, *ARS1211* and the *HIS3* marker gene (Figure 2). Transformants were selected on SMG media and after strain confirmation by PCR (Table S7) the gRNA plasmid encoding the *KlURA3* marker was removed. As the next step, the SinLoG with variable length of promoters and terminators was removed from the *SGA1* locus in the strains IMX1717 and IMX1747. To this end, both strains were transformed with plasmid pUDR413 and 1 µg of *KlURA3* repair fragment amplified with primers 13273 and 13274 introducing homology flanks to the *SGA1* site (Figure 2). Transformants were selected on SMG supplemented with G418 and after strain confirmation by PCR using primers 11898 + 7479, 11898 + 2363 and 170+7479, the plasmid was removed. For the construction of the strain IMX2028 containing the native SinLoG in Ch V and watermarked SinLoG in Ch IX, first, the intermediate strain IMX1748 was constructed by removing the *KlURA3* gene from IMX1771. This was done by transformation with plasmid pUDR529 encoding a gRNA for the *KlURA3* gene and a repair fragment amplified with the primers 4223 and 4224 and containing homology to the *SGA1* locus. Colonies were selected on YPD media supplemented with Hygromycin B and correct strain construction was confirmed by PCR using primers 4223 and 4224. After *KlURA3* marker removal, IMX1748 was transformed with plasmid pUDR314 and the mixture of fragments for the watermarked SinLoG, *ARS418*, *ARS1211,* and *KlURA3* marker gene resulting in strain IMX2028. Correct assembly of the fragments was confirmed by PCR.

Selective CRISPR/Cas9 genome editing

For selective CRISR/Cas9 genome editing, IMX1717 (Table 1) was transformed with 1 µg of a 120 bp repair fragment with homology to the beginning and end of the gene and with 1µg of plasmids pUDR531 or pUDR532 containing a gRNA for *TPI1* and *PYK1* respectively as described in the section Construction of gRNA plasmids and Table S4.

Cells were plated on YPD with Hygromycin B. Repair fragments (120 nt-long) and diagnostic primers are listed in Table S7-G.

Whole genome sequencing and data analysis

Yeast genomic DNA was isolated using the Qiagen Genomic DNA Buffer Set and Genomic-tip 100/G tips (Qiagen, Hilden, Germany) according to the manufacturer's manual. The incubation step with zymolyase was performed for 11 hours and the incubation step for digestion with proteinase K was performed overnight. The concentration of the genomic DNA mixture was measured with the BR ds DNA kit (Invitrogen, Carlsbad, CA, USA) using a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific) and the purity was verified with a Nanodrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific).

IMX1770, IMX1771 and IMX2028 genomes were sequenced on an Illumina MiSeq Sequencer (Illumina, San Diego,CA, USA) using the MiSeq® Reagent Kit v3 with 2 × 300 bp read length. Extracted DNA was mechanically sheared to an aimed average size of 550 bp with the M220 ultrasonicator (Covaris, Wolburn, MA, USA). DNA libraries were prepared using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina) according to the manufacturer's manual. Quantification of the libraries was done by qPCR using the KAPA Library Quantification Kit for Illumina platforms (Kapa Biosystems, Wilmington, MA, USA) on a Rotor-Gene Q PCR cycler (Qiagen). The genome of CEN.PK113-7D, the *in silico* constructed watermarked and reference (native) SinLoG sequences and the *KlURA3* repair fragment were used as a reference to map sequence reads of genomic DNA onto using the Burrows-Wheeler Alignment tool (BWA) (*48*). The sequence alignment was further processed using SAMtools (*49*). Coverage of the sequence reads was also calculated using the Magnolya algorithm (*50*).

All Illumina sequences are available at NCBI (http://www.ncbi.nlm.nih.gov/) under the bioproject accession number PRJNA554743.

Batch cultivations of IMX1770, IMX1771 and IMX2028

Batch cultivations were performed in biologically independent triplicates in 2-Liter fermenters (Applikon, Delft, the Netherlands) with a working volume of 1.4 L. Cells from exponentially growing SMG shake flask cultures were inoculated into the fermenters containing SMG supplied with 0,2 $g.L^{-1}$ antifoam Emulsion C (Sigma Aldrich, St Louis, MO) at an OD660 of 0.4. The fermenters were sparged with dried compressed air at a rate of 700 mL/min (Linde, Gas Benelux, The Netherlands). The broth was stirred constantly at 800 rpm, kept at a constant temperature of 30°C and at a pH of 5 by automatic addition of 2M KOH by an Applikon ADI 1030 Bio Controller.

Optical density was measured every hour at 660 nm with a Jenway 7200 spectrophotometer (Staffordshire, United Kingdom). For extracellular metabolite

analysis 1 mL of the broth was centrifuged for at least 10 min at 13,000 rpm and the supernatant was analyzed with high-performance liquid chromatography (HPLC) using an Agilent 1100 (Agilent Technologies, Santa Clara, CA, USA) with an Aminex HPX-87H ion-exchange column (BioRad, Veenendaal, the Netherlands) operated with 5 mM sulfuric acid as mobile phase at a flow rate of 0.6 mL/min. The carbon dioxide and oxygen concentration in the gas outflow were analyzed by a Rosemount NGA 2000 analyzer (Baar, Switzerland), after cooling of the gas by a condenser (2°C) and drying using a PermaPure Dryer (model MD 110-8P-4; Inacom Instruments, Veenendaal, the Netherlands). Biomass dry weight was measured 5 – 6 times by filtering (pore size 0.45 µm, Gelman Laboratory, Ann Arbor, MI, USA), as described previously (*51*). Sampling for RNA was done directly from the reactor in liquid nitrogen as described by Piper *et al*. (*52*). The cells were stored at -80˚C for maximally two weeks until further processing and RNA was extracted as previously described [23]. An equivalent of 48 mg dry weight per sample was used. At the same time points as the samples that were taken for RNA isolation, approximately 62.5 mg dry weight was sampled for the enzyme assays, stored at -20˚C in 4 mL aliquots and further process as previously described (*51*). Optical densities of the cultures at the moment of sampling for RNA analysis and enzyme assays can be found in Table S9.

Determination of *in vitro* enzyme activities

On the day of the enzyme assays, frozen samples were thawed and prepared for assays as described by Postma *et al.* (*51*). Assays were performed using a U-3010 spectrophotometer (Hitachi, Tokio, Japan) at 30 °C and 340 nm as described by Jansen *et al.* (2005)(*53*), with the exception of Pfk, which was performed according to Cruz *et al.* (2012) (*54*). Reported activities are based on at least two technical replicates, measured with different cell extract concentrations. When necessary, cell extracts were diluted in 100 mM monopotassium phosphate buffer and with 2 mM magnesium chloride (pH 7.5), or in demineralized water when triose phosphate isomerase activity was measured. The protein concentration of the cell extracts was determined as described by Lowry *et al.* (*55*) using bovine serum albumin as a standard.

RNA Sequencing simulation

To evaluate the watermarking methods, and to compare the markerQuant tool with traditional alignment, we generated artificial RNA-Seq reads from the native and watermarked sequences. Using the polyester R package (*56*), we simulated two conditions, in which the second condition has a four-fold expression of each transcript compared to the first condition. We used the standard *error_rate* parameter of 0.005. Generated reads were paired-end, each end 100 bp in length.

RNA sequencing and data analysis

RNA libraries and sequencing were performed by Novogene Bioinformatics Technology Co., Ltd (Yueng Long, Hong Kong). Sequencing was performed using HiSeq 150 bp paired-end reads system using 250 – 300 bp insert strand-specific library. As described by Novogene, library preparation involved mRNA enrichment using oligo (dT) beads, followed by random fragmentation of the mRNA. cDNA was synthetized from mRNA using random hexamer primers and a second strand synthesis was done applying a custom second strand synthesis buffer (Illumina), dNTPs, RNase H, and DNA polymerase I. After adaptor ligation, double-stranded cDNA library was finalized by size selection and PCR enrichment and samples were sequenced. Obtained data had an average of 23.08M reads per sample (Table S8). To quantify the abundance of glycolytic genes with and without watermarks, a similar scheme as the k-mer algorithm of Gehrmann and co-workers (*38*) was applied. Briefly, for each transcript, we identify sequence markers of 21bp that are unique in the transcript relative to the entire transcriptome and genome. With an exact matching algorithm, these markers can uniquely identify the transcript of origin of a read in RNA-Seq data. In contrast to the previous work (*38*), we did not remove overlapping markers (we did not remove redundant markers) but merged them into larger sequences in which any 21bp k-mer would uniquely identify the transcript of origin. This allowed us to recover a higher percentage of reads per transcript. Gaps in these merged sequences that are not unique relative to the genome and transcriptome were ignored in the marker quantification step. As in previous work, we used an Aho-Corasick exact string-matching algorithm to quantify transcripts. Differential expression was performed using DE-Seq2 (*57*).

RNA sequencing data analysis implementation and code availability

The marker discovery and quantification tools were developed in scala, and the entire pipeline is implemented in python using Snakemake (*58*). In addition to the k-mer method, a traditional alignment pipeline is also implemented in the markerQuant utility. All code, including an example dataset, is available at https://github.com/thiesgehrmann/markerQuant.

Label free quantification (LFQ) by shot-gun proteomics

Cultivation and sampling

For proteomics analysis, the yeast strains pre-grown to exponential phase in SMG in shake flask were used to inoculate fresh SMG flasks. 5 ml of these cultures in mid-exponential phase were centrifuged for 10 min at 5000g at 4 °C and the cell pellet was directly stored at -80°C. Cultures were performed in biological triplicates for strains with watermarked and native glycolysis. To verify if the difference in Pgm1 activity

observed in bioreactor between the strains with watermarked and native glycolysis was also present in shake flask culture, Gpm1 activity was assayed in cell samples from the shake flasks. This additional analysis confirmed the lower specific activity of Gpm1 in the watermarked strain (Fig S9).

Protein extraction and trypsin proteolytic digestion

Cell pellets were resuspended in lysis buffer composed of 100 mM TEAB containing 1% SDS and phosphatase/protease inhibitors. Yeast cells were lysed by glass bead milling and thus shaken 10 times for 1 minute with a bead beater alternated with 1 min rest on ice. Proteins were reduced by addition of 5 mM DTT and incubated for 1 hour at 37°C. Subsequently, the proteins were alkylated for 60 min at room temperature in the dark by addition of 50 mM acrylamide. Protein precipitation was performed by addition of four volumes of ice-cold acetone (-20°C) and proceeded for 1 hour at -20°C. The proteins were solubilized using 100 mM ammonium bicarbonate. Proteolytic digestion was performed by Trypsin (Promega, Madison, WI), 1:100 enzyme to protein ratio, and incubated at 37°C overnight. Solid phase extraction was performed with an Oasis HLB 96-well µElution plate (Waters, Milford, USA) to desalt the mixture. Eluates were dried using a SpeedVac vacuum concentrator at 45°C. Dried peptides were resuspended in 3% ACN/0.01% TFA prior to MS-analysis to give an approximate concentration of 250 ng per µl.

Large-scale shot-gun proteomics

An aliquot corresponding to approx. 250 ng protein digest was analysed using an one dimensional shot-gun proteomics approach (*59*). Briefly, the samples were analysed using a nano-liquid-chromatography system consisting of an ESAY nano LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 µm x 150 mm, 2µm), and an QE plus Orbitrap mass spectrometer (Thermo). The flow rate was maintained at 350 nL/min over a linear gradient from 6% to 26% solvent B over 45 minutes, followed by back equilibration to starting conditions. Data were acquired from 5 to 60 min. Solvent A was H_2O containing 0.1% formic acid, and solvent B consisted of 80% acetonitrile in H_2O and 0.1% formic acid. The Orbitrap was operated in data depended acquisition mode acquiring peptide signals from 385-1250 m/z at 70K resolution. The top 10 signals were isolated at a window of 2.0 m/z and fragmented using a NCE of 28. Fragments were acquired at 17K resolution.

Database search, label free quantification and visualisation

Data were analysed against the proteome database from *Saccharomyces cerevisiae* (Uniprot, strain ATCC 204508 / S288C, Tax ID: 559292, July 2018) using PEAKS Studio

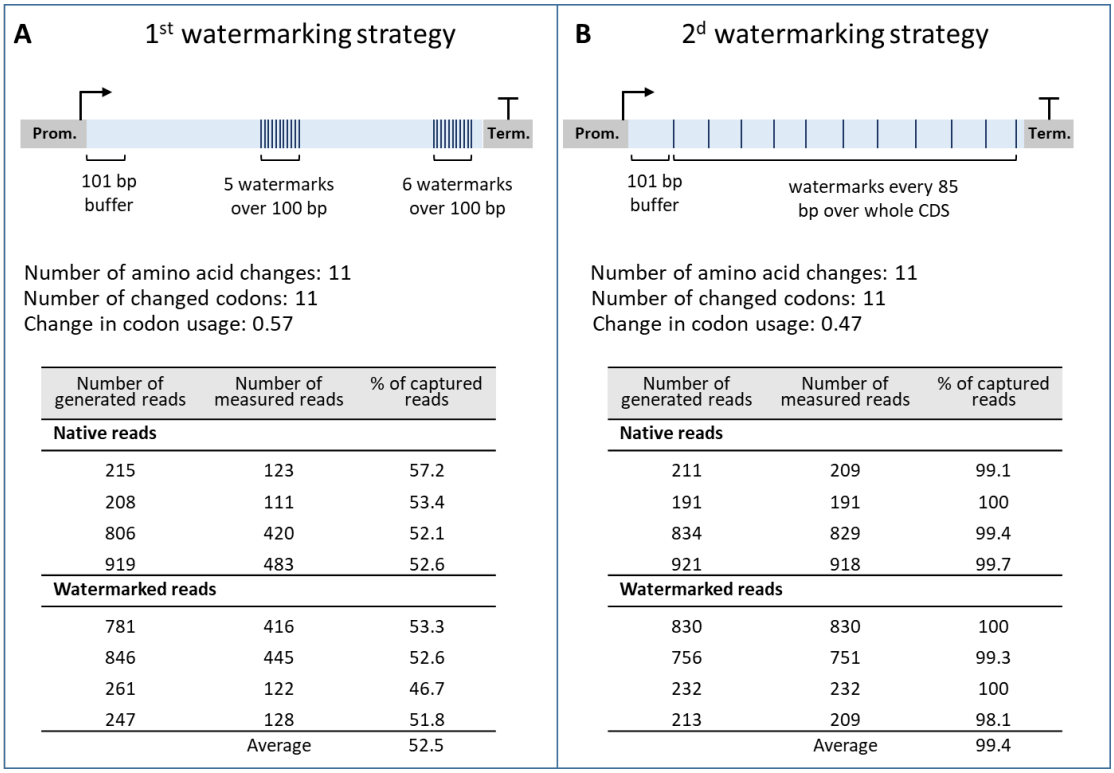
X (Bioinformatics Solutions Inc)[2] allowing for 20 ppm parent ion and 0.02 m/z fragment ion mass error, 2 missed cleavages, acrylamide as fixed and methionine oxidation and N/Q deamidation as variable modifications. Peptide spectrum matches were filtered against 1% false discovery rates (FDR) and identifications with ≥ 2 unique peptides. Changes in protein abundances between both strains IMX1770 and IMX1771 using the label free quantification (LFQ) option provided by the PEAKS Q software tool (Bioinformatics Solutions Inc) (*60*). Protein areas were normalised to the total ion count (TIC) of the respective analysis run before performing pairwise comparison between the above mentioned strains. LFQ was performed using protein identifications containing at least 2 unique peptides, which peptide identifications were filtered against 1% FDR. The significance method for evaluating the observed abundance changes was set to ANOVA. The abundances of the glycolytic enzymes were further visualised as bar graphs using Matlab2018b. The area of the biological triplicates were averaged and standard deviations were represented as error bars.

Results and discussion

Design and *in silico* validation of the watermarking strategy

The presence of watermarks in the CDS of glycolytic genes shall enable discrimination of the watermarked versus native DNA and mRNA sequences with a minimal effect on transcript and protein levels, activity of enzymes in the glycolytic pathway and ultimately, yeast physiology. Finding the optimal tradeoff between robust watermark detection by sequencing and minimal physiological impact was therefore the main design principle of the watermarking strategy. Based on current RNA sequencing resolution (Illumina platform with an error rate of <1%), at least five nucleotide substitutions were required to distinguish watermarked from native sequences using random Single Nucleotide Polymorphisms (SNPs). Codon replacement was performed on the amino acids encoded by four to six alternative triplets (A, G, P, T, V, L, R, S), favouring triplets for which only the third base pair of the triplet was different from the original codon. The codon with the most similar percentage of abundance when referring to the codon usage table of *S. cerevisiae* (Table S1) was chosen, avoiding triplets leading to more than 20% variation in abundance when possible. The structure of the 5' region of the mRNA is important for translation efficiency. Not only does the folding energy at the 5' end affect translation initiation, but the presence of non-optimal codons close to the initiation site can stall ribosomes, thereby hampering translation initiation (*61, 62*). Furthermore, as translation initiation is considered as translation limiting step (*63*), and following the example of Annaluru and colleagues (*9*), the first 101 nt of the CDS were preserved. The optimal distribution of watermarks over the remaining CDS stretch was tested with two *in silico* approaches using *ADH1*. In the first approach, watermarks were colocalised in two 100 nt regions, one in the middle of the CDS, and the other located 10 nt upstream of the stop codon (Figure 1A). In the second approach, base pair substitutions were equally spread over the CDS sequence, every 85 nt (Figure 1B). In both approaches 11 watermarks were introduced, which resulted in an overall change in codon usage of 0.57 for the first and 0.47 for the second method (Figure 1, Figure S1. See Materials and Methods section for calculation of the change in codon usage). In order to evaluate the discriminatory potential of RNA sequencing with these two strategies, 100bp paired-end sequence reads were simulated for both watermarked and native *ADH1* copies (See Materials and Methods section). These data were processed using the k-mer method developed by Gehrmann et. al. (*38*), only considering reads containing watermarks (see Material and Methods section), to selectively quantify watermarked and native reads. On average 52.5% of the reads were captured when using the first approach with clustered watermarks, while 99.4% were detected using the second approach, with watermarks spread over the CDS (Figure 1).

The Pearson correlation coefficient between generated and measured reads was above 0.99 for both methods, indicating that both methods are able to retrieve the variation in abundance across the samples, required for differential expression. The second approach resulted in a better sequence coverage and slightly lower codon usage change. However,the first approach is less labour intensive when manual design is performed, and is less likely to affect co-translational folding (*21*), as a shorter part of the CDS undergoes codon usage change. The first method was therefore selected as watermarking strategy (detailed in Box 1) and used to edit *in silico* the CDS of 13 genes of glycolysis and alcoholic fermentation (*HXK2*, *PGI1*, *PFK1*, *PFK2*, *FBA1*, *TPI1*, *TDH3*, *PGK1*, *GPM1*, *ENO2*, *PYK1*, *PDC1* and *ADH1*, see example for watermarking for *FBA1* and *ENO2* in Figure S3). This resulted in a reasonably low change in codon usage of the watermarked genes of 0.5 on average (Table S2). Using simulated data, we compared the performance of the k-mer method with traditional alignment and found that the k-mer method was able to achieve a higher read retrieval rate than alignment indicating a more accurate transcription estimate (Table S3). The watermarked CDS were synthetized with flanks compatible with Golden Gate assembly (plasmids pGGKp137 to pGGKp150, Table S4).



Number of generated reads	Number of measured reads	% of captured reads
Native reads		
215	123	57.2
208	111	53.4
806	420	52.1
919	483	52.6
Watermarked reads		
781	416	53.3
846	445	52.6
261	122	46.7
247	128	51.8
	Average	52.5

Number of generated reads	Number of measured reads	% of captured reads
Native reads		
211	209	99.1
191	191	100
834	829	99.4
921	918	99.7
Watermarked reads		
830	830	100
756	751	99.3
232	232	100
213	209	98.1
	Average	99.4

Figure 1 – Comparison of two watermarking strategies. **A)** First strategy with clustered watermarks. **B)** Second strategy with watermarks distributed over the whole coding region. The tables in panel A and B represent the % of sequencing reads that can be captured by the two watermarking strategies, calculated from *in silico* simulated 100bp paired-end sequencing reads.

Strain construction strategy and confirmation

In the SwYG strain (*13*), the set of genes involved in glycolysis and fermentation was reduced from 26 to 13 and relocalized to a single locus (Single Locus Glycolysis, SinLoG) on chromosome IX. The SwYG strain is a perfect platform to rapidly remodel glycolysis and alcoholic fermentation and test multiple (heterologous) variants. SwYG was therefore used as starting strain to express the watermarked genes. Using simultaneous Cas9-mediated genome editing and *in vivo* assembly, the entire glycolytic and fermentation pathways composed of 13 watermarked genes were integrated in one step in the *CAN1* locus on chromosome V. The watermarked genes were framed by the native, standardized corresponding promoters and terminators (800 bp and 300 bp respectively). Three helper elements, two Autonomously Replicating Sequences (ARS) and a selection marker were included in the SinLoGs design (Figure 2). Two active ARSs (*ARS418* and *ARS1211*) were added on both ends of the ca. 35Kb long SinLoGs to minimize the risk of perturbing DNA replication of this long DNA stretch. A selection marker was used to facilitate screening for correct integration and removal of the SinLoGs. The native SinLoG, present in the *SGA1* locus on chromosome IX, was then removed using the Cas9 endonuclease, resulting in strain IMX1770 (Figure 2 and Figure 3). To obtain an isogenic control strain, the same procedure was followed to construct a strain with native SinLoG, framed by the same promoters and terminators as the watermarked genes, and integrated in the same *CAN1* locus on chromosome V (Figure 2 and 3). This control strain was named IMX1771. The genome of both strains was sequenced, confirming the presence of a single, correctly assembled glycolytic pathway at the targeted chromosomal location. Sequencing revealed the absence of mutations in the coding regions of the glycolytic and marker genes but identified a few mutations in the promoter and terminator regions of the glycolytic and selection marker expression cassettes (Table S5). In IMX1770, a single Single Nucleotide Variation (SNV) was found in the promoter of *PFK1* and *HIS3* and in the terminator of *PGK1* and *ENO2*. In IMX1771, a single SNV was identified in the *HIS3* terminator, and a short TA stretch was missing in the promoter of *GPM1*. A third strain, IMX2028 was constructed. IMX2028 harbored a double SinLoG, one located on chromosome IX carrying the watermarked genes and another on one carrying the native yeast genes on chromosome V (IMX2028, Figure 1 and Table 1). Unfortunately, genome sequencing revealed the deletion of a large region of the mitochondrial DNA (Figure S4). The strain IMX2028 was constructed to evaluate the performance of the k-mer method to discriminate between watermarked and native genes when present in the same strain. Despite IMX2028 respiration deficiency, the watermarked and native SinLoG that this strain carried were essentially faithful to the

in silico design (Table S5) which made this strain still valuable for differential quantification of watermarked and native genes.

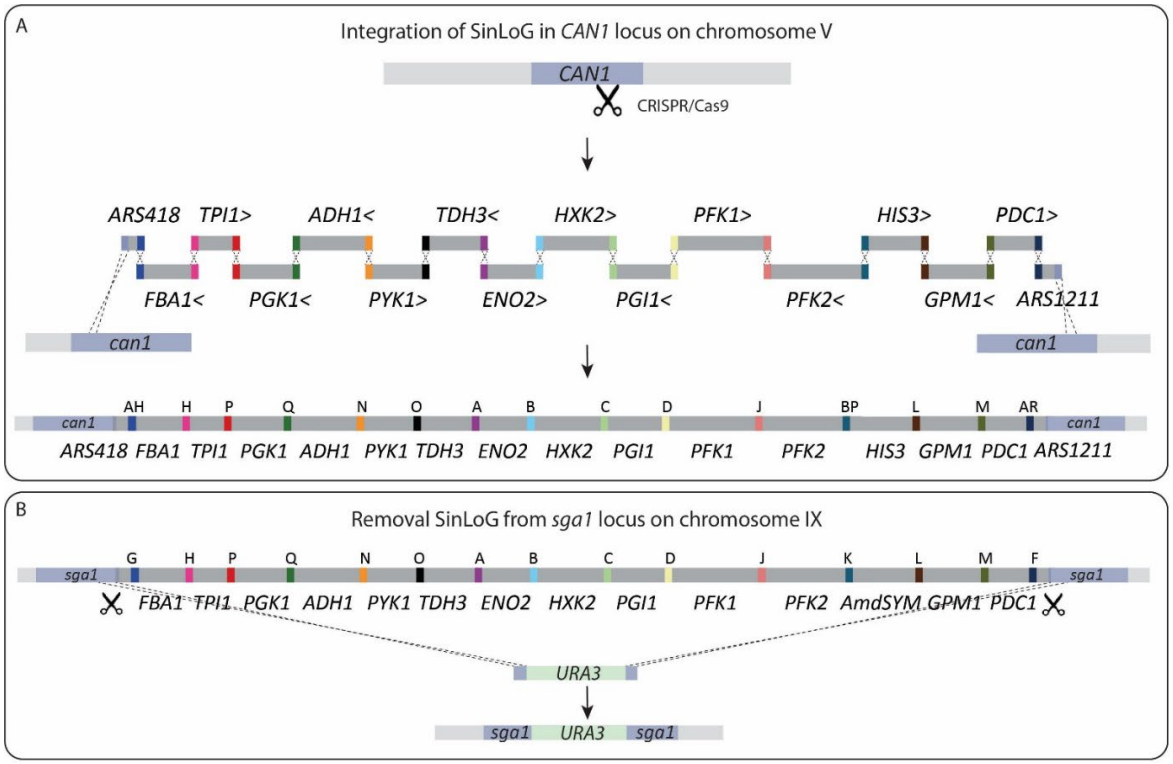


Figure 2 – Construction of SinLoG (Single Locus Glycolysis) strains IMX1770 and IMX1771 using the glycolysis swapping strategy (*13*). **A)** A newly designed glycolysis is integrated in the *CAN1* locus by simultaneous CRISPR/Cas9-aided editing of *CAN1* and *in vivo* assembly of glycolytic expression cassettes and helper fragments (ARS418, ARS1211 and the selection marker HIS3). The > and < signs next to the gene names indicate the directionality of transcription and letters indicate the synthetic homologous recombination (SHR) sequence which was used for assembly. **B)** Subsequently, the Single Locus Glycolysis present in the *SGA1* locus was excised by double editing using CRISPR/Cas9 and replaced by the *URA3* selection marker. The set of genes integrated in *CAN1* is then the sole set of glycolytic genes present in the newly constructed strain and is essential for growth on glucose.

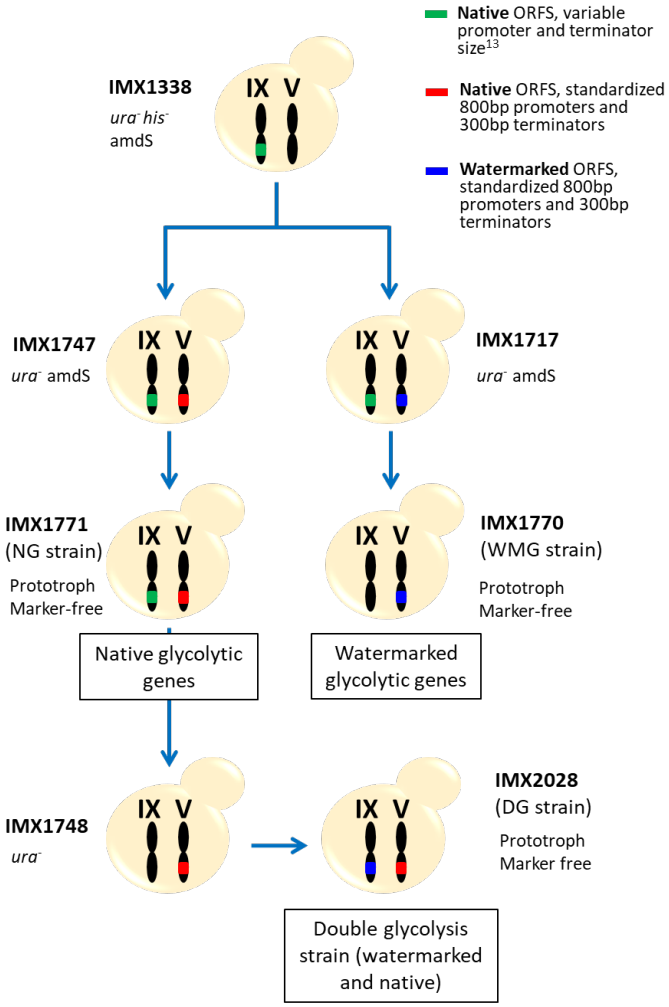


Figure 3 – Strain construction workflow. The Switchable Yeast Glycolysis (SwYG) strain, IMX1338, served as parental strain to introduce in chromosome V a SinLoG (Single Locus Glycolysis) with native ORFs and standardized promoters/terminators (IMX1747) as well as with watermarked ORFs and standardized promoters/terminators (IMX1717). From both strains the native SinLoG in chromosome V with variable promoters and terminators was removed (resulting in strain IMX1771 and IMX1770, respectively). After removal of *URA3* from strain IMX1771 (native ORFs) the SinLoG with watermarked ORFs was introduced in chromosome IX, resulting in a strain with double glycolysis (IMX2028).

Watermarks do not affect yeast physiology

To evaluate the impact of DNA watermarking on yeast physiology, the watermarked strain IMX1770 and its isogenic control IMX1771 were grown in aerobic batch bioreactors and their growth kinetics were compared. Both strains were prototrophic, meaning that they fully relied on glucose, the sole carbon and energy source catabolized via glycolysis, to produce the required cellular building blocks and therefore to grow. The two strains displayed identical growth rates (0.33 ± 0.004 h^{-1} and 0.32 h^{-1} ± 0.002 h^{-1} for IMX1770 and IMX1771 respectively) as well as glucose and O_2 uptake rates, ethanol and CO_2 production rates and yields (Figure 4 and Table 2). Both strains passed the diauxic shift and grew equally well using the ethanol which was produced during

growth on glucose, as carbon and energy source (Figure S5). Watermarking of glycolytic and alcoholic fermentation genes therefore did not alter metabolic fluxes and the overall physiological responses during fast respiro-fermentative on glucose and full respiratory growth on ethanol.

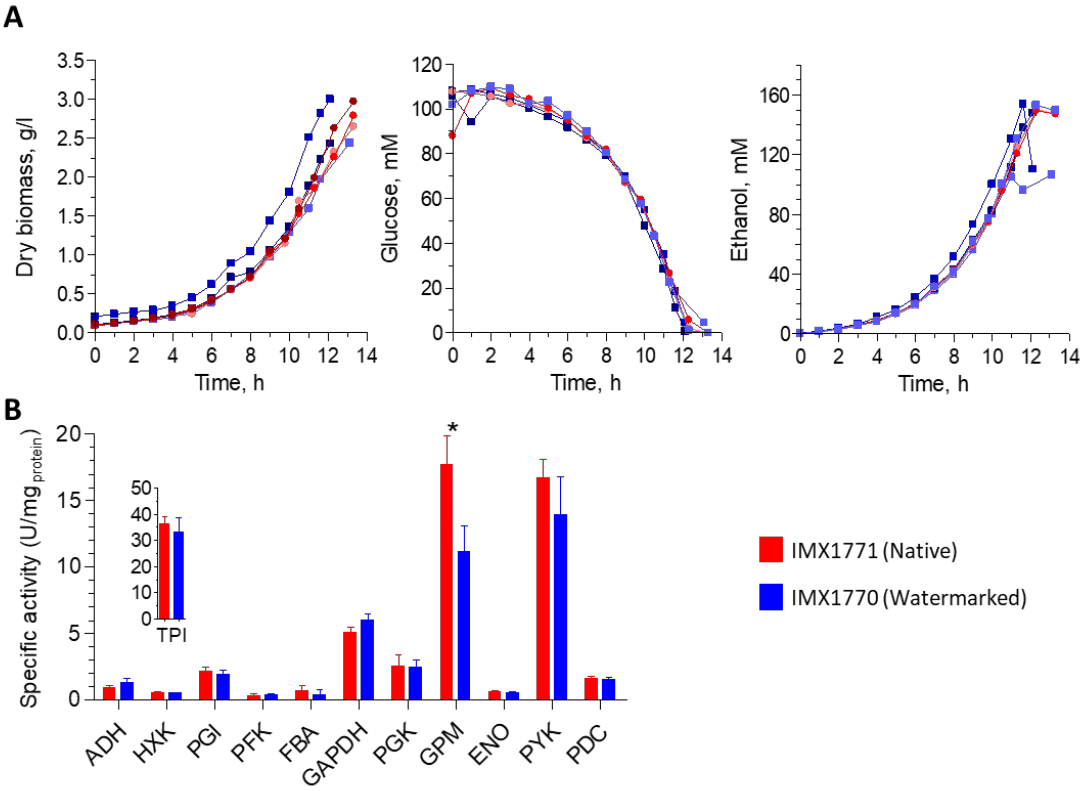


Figure 4 –Physiological characterization of strains with native (IMX1771) and watermarked (IMX1770) glycolytic genes during aerobic batch cultures in bioreactors. **A)** left panel, biomass concentration (gram dry biomass per liter), central panel, glucose concentration (mM), right panel, ethanol concentration (mM). Three independent culture replicates are represented for each strain. Shades of blue with square symbols, IMX1770, shades of red with round symbols, IMX1771. **B)** specific enzyme activities of the 12 reactions encoded by the 13 glycolytic enzymes (Pfk1 and Pfk2 form an enzyme complex) of the strains with native (IMX1771, red bars) and watermarked glycolysis (IMX1770, blue bars). Samples were taken in mid-exponential phase. Bars represent the average and standard deviation of measurements from three independent batch cultures for each strain. Stars indicate enzyme activities that are significantly different between the two strains (Student t-test, p-value threshold 0.05, two-tailed test, homoscedastic).

Watermarking might affect protein folding and consequently function. However, as yeast glycolysis is characterized by an overcapacity of its enzymes, mild variations of glycolytic enzymes activities might not be detectable by growth kinetics. The 12 specific activity assays of the 13 enzymes (Pfk1 and Pfk2 are subunits of a hetero-octameric phosphofructokinase (*64*)) encoded by the watermarked genes were therefore assayed *in vitro*. The specific activity of these 13 enzymes was, with the exception of Gpm1,

remarkably similar between watermarked and native strains (p-values above 0.05; Student t-test, two-tailed, homoscedastic). For all enzymes, specific activities were remarkably similar to protein abundance, including a 1.6-fold decrease in specific activity and protein abundance for Gpm1 (Figure 4, Figure S7 and S8). Watermarking therefore did not affect or marginally affect protein expression and functionality (Figure 4).

Table 2 – Physiological characterization of strains with native (IMX1771) and watermarked (IMX1770) glycolytic genes during aerobic batch cultures in bioreactors. Data represent the average and standard deviation of measurements from three independent batch cultures for each strain. Statistical analysis (Student t-test, p-value threshold 0.05, two-tailed test, homoscedastic) revealed no significant differences between the two strains.

		IMX1771	IMX1770
Yields	Y_{sx} ($g_{dry\ weight}/g_{glucose}$)	0.12 ± 0.01	0.13 ± 0.01
	$Y_{sglycerol}$ (mol/mol)	0.07 ± 0.01	0.05 ± 0.01
	$Y_{sethanol}$ (mol/mol)	1.47 ± 0.05	1.46 ± 0.02
	$Y_{sacetate}$ (mol/mol)	0.05 ± 0.00	0.06 ± 0.00
Specific rates	μ_{max} (h^{-1})	0.32 ± 0.00	0.33 ± 0.01
	$q_{glucose}$ ($mmol/g^{-1}.h^{-1}$)	-14.6 ± 0.7	-14.7 ± 0.9
	$q_{glycerol}$ ($mmol/g^{-1}.h^{-1}$)	0.96 ± 0.03	0.76 ± 0.17
	$q_{ethanol}$ ($mmol/g^{-1}.h^{-1}$)	21.4 ± 0.3	21.4 ± 1.7
	$q_{acetate}$ ($mmol/g^{-1}.h^{-1}$)	0.79 ± 0.05	0.88 ± 0.06
Carbon balances (%)		105 ± 2	103 ± 3

To further explore the potential impact of watermarking on yeast physiology, the transcriptome of IMX1771 and IMX1770 grown in aerobic batch reactors was compared. The transcriptional response of these two strains was remarkably similar (Figure 5). The native and watermarked glycolytic genes were the only differentially expressed genes between the two strains. This differential expression reflects the absence of the native genes and therefore their lack of expression in IMX1770, and the absence and lack of expression of the watermarked genes in IMX1771. However, expression levels of the native and watermarked genes in IMX1771 and IMX1770, respectively, were highly similar (Figure 6A).

Physiological characterization of IMX2028 confirmed the respiration deficiency suggested by the absence of mitochondrial DNA. The k-mer method was able to selectively quantify expression of the native and watermarked genes. While expression of glycolytic and respiration genes might differ in IMX2028 as compared to IMX1770 and IMX1771 due to the mutations in mitochondrial DNA and associated respiration deficiency, the relative expression of glycolytic and fermentation genes, expressed from the same promoters in the native and watermarked SinLoG, was not expected to differ between the native and watermarked genes in this strain. Accordingly, and in agreement with the similarity of the expression levels between IMX1770 and IMX1771, transcript levels of native and watermarked genes in IMX2028 were identical (Figure 6B). Watermarking of 13 highly expressed genes of central carbon metabolism, essential for glucose utilization, had therefore no impact on yeast transcriptome and physiology.

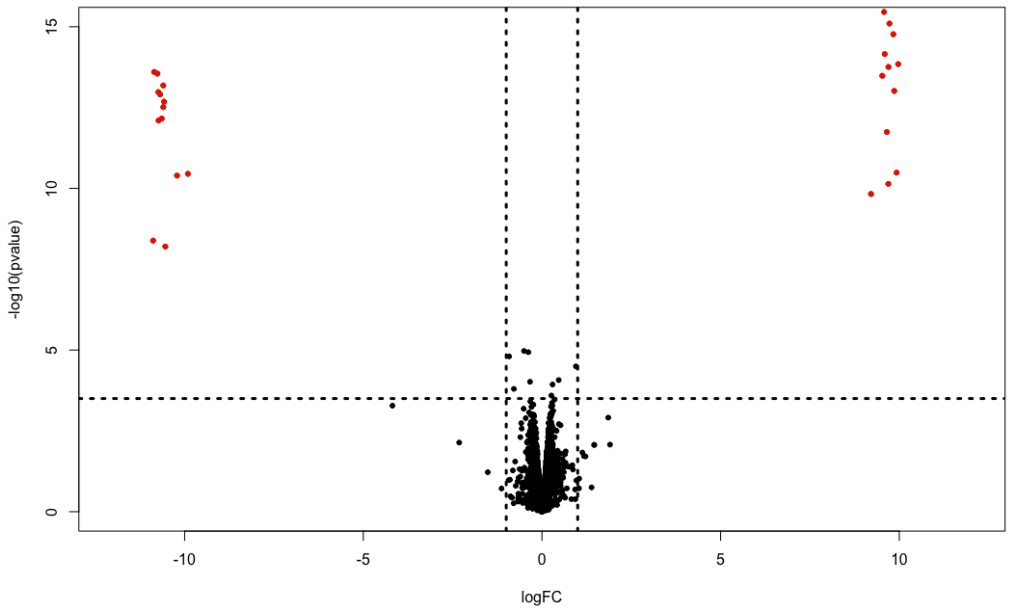


Figure 5 – Genome-wide transcriptome analysis of IMX1770 and IMX1771. The x-axis represents the log fold change in expression, and the y-axis represents the -log p-value. Each point represents a transcript. A negative log fold change reflects higher expression in the native strain than in the watermarked strain, and vice versa. The horizontal, dashed line represents the FDR corrected p-value threshold of 0.05, and the vertical dashed lines represent a log fold change threshold of 1. Red points indicate significantly differentially expressed transcripts (FDR-corrected p-value above 0.05 and Log fold change higher than 1).

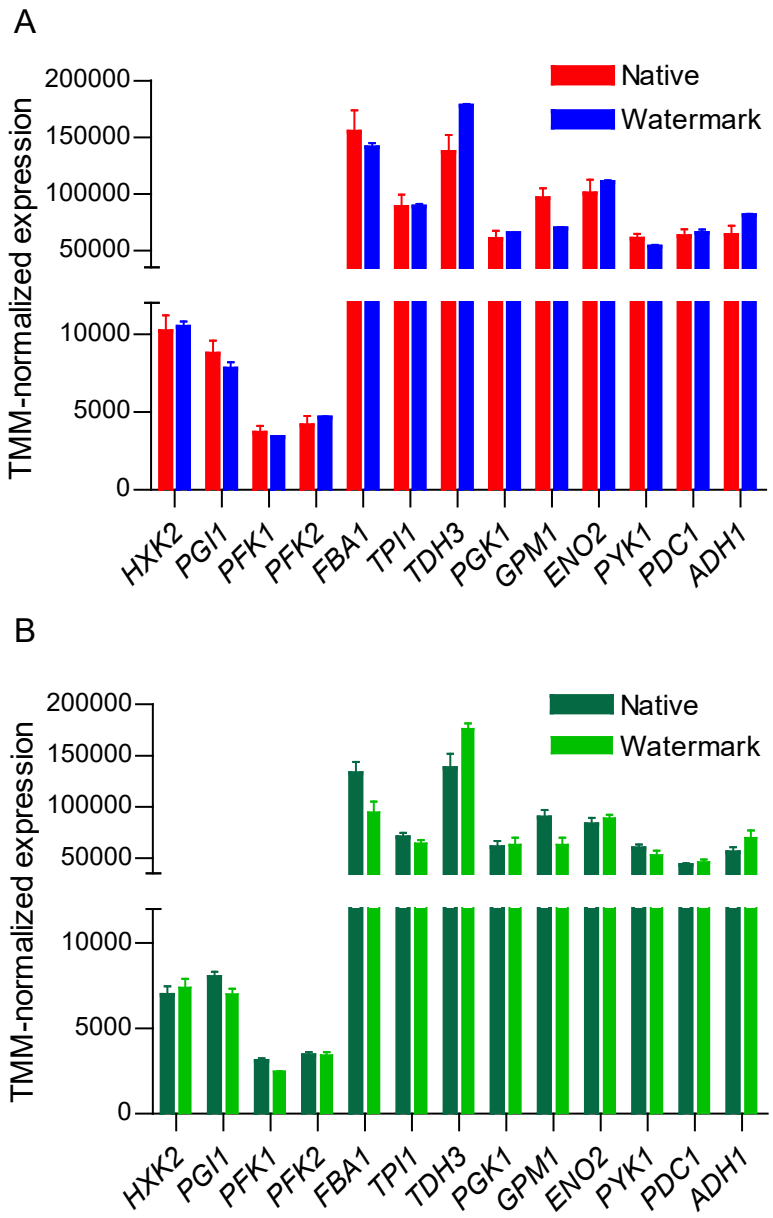


Figure 6 – Glycolysis and fermentation transcript levels of *S. cerevisiae* grown in aerobic batch cultures in bioreactors. **A)** Watermarked transcript levels of IMX1770 (blue) and native transcript levels of IMX1771 (red). **B)** Watermarked and native transcript levels of IMX2028. Bars represent the average and standard deviation of three independent cultures replicates. Samples were taken in mid-exponential phase (Table S9). No significant change in expression was found between watermarked and native genes (Student t-test, p-value threshold 0.05, two-tailed test, homoscedastic) between IMX1770 and IMX1771 and within IMX2028.

Watermarking enables selective *CRISPR/Cas9* genome editing

DNA binding and editing by CRISPR/Cas9 requires the presence at the targeted site of a specific PAM recognition sequence (*15*). A single nucleotide variation in this sequence can abolish Cas9 ability to introduce a double strand DNA break (*15*). This feature is particularly interesting when considering selective editing of identical or highly similar sequences. If strategically designed, watermarks can enable targeted editing of a watermarked gene, leaving the native copy intact or conversely, prevent editing of the watermarked gene while cutting the native copy. Guide RNAs (gRNAs) selectively targeting the native copies of *PYK1* and *TPI1* for CRISPR/Cas9 editing were designed (Table S6), inserted into expression vectors and transformed to IMX1717, a double SinLoG strain and direct ancestor of IMX1770 (Figure 3). Double-stranded DNA fragments of 120 nt were supplied during transformation to repair via homologous recombination the break induced by CRISPR/Cas9. As the sequence of the native and watermarked genes is identical with the exception of watermarks, a single primer set designed just outside the open reading frame can be used to amplify both copies of *PYK1* or *TPI1* in a single PCR reaction using IMX1717 genomic DNA as template. Ran on a gel, the PCR products of this reaction would lead to a single band corresponding to both the native and watermarked copies of *PYK1* or *TPI1*. Selective editing would lead to the appearance of a second, smaller band on gel, corresponding to the edited copy of *PYK1* or *TPI1*. Out of 15 colonies of IMX1717 transformed with the gRNA targeting *PYK1*, three displayed two bands demonstrating editing of a single *PYK1* copy (Figure 7). Five out of 15 colonies of IMX1717 transformed with the gRNA targeting *TPI1* showed selective editing (Figure 7). For two transformants per gene (*TPI1* and *PYK1*) showing two bands on the gel, sequencing the largest band confirmed the presence of the watermarked sequence only, confirming selective editing of the native *TPI1* and *PYK1* (Figure S6). It has recently been shown that cells can use chromosomal DNA with high homology to repair a CRISPR/Cas mediated DNA break, leading to loss of heterozygosity (*65*). In the present case, it means that cells could repair the induced DNA break in the targeted, native gene copy with its watermarked homolog, resulting in two copies of the watermarked gene, but a single PCR product and therefore a single band on gel. Sequencing of the unique PCR product of four colonies in which editing of *PYK1* and *TPI1* was considered unsuccessful revealed that, for all tested colonies, the *PYK1* and *TPI1* genes were cut by CRISPR/Cas but repaired by (part of) the watermarked allele. Editing of the targeted, native genes by CRISPR/Cas9 was therefore highly efficient (100% of the tested colonies), however the DNA break was repaired either by the supplied repair DNA fragment or by the watermarked homolog.

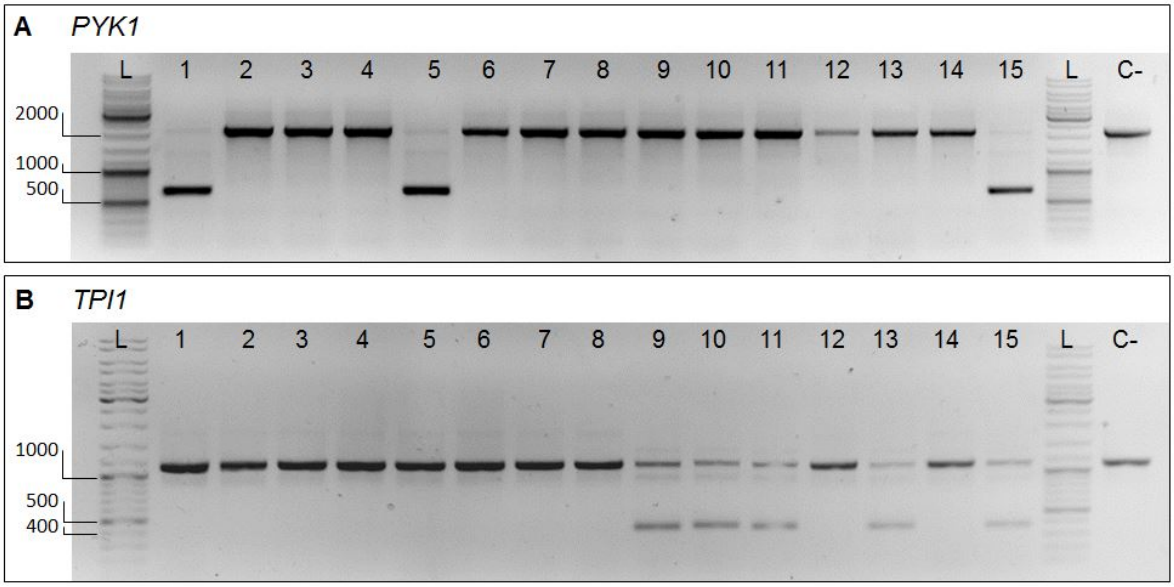


Figure 7 – Diagnostic PCR for selective editing of native glycolytic genes. Separation of PCR products resulting from outside-outside amplification to identify edited (non-watermarked) and non-edited (watermarked) loci for *PYK1* **(A)** and *TPI1* **(B)** from transformants of IMX1717 (double SinLoG)**. A)** Lanes 1-15 show the PCR results of amplification of the *PYK1* locus of randomly picked colonies. Successful editing of the locus results in a DNA fragment with a length of 670 bp. No editing of the locus results in a DNA fragment with a length of 2177 bp. Primers 11915 and 4667 were used. Lanes 1, 5 and 15 display bands of both sizes revealing selective editing. **B)** Lane 1-15 show the PCR results of amplification of the *TPI1* locus of randomly picked colonies. Successful editing of the locus results in a DNA fragment with a length of 378 bp. No editing of the locus results in a DNA fragment with a length of 1125 bp. Primers 3514 and 6406 were used. Lanes 9-11, 13 and 15 display bands of both sizes revealing selective editing. A negative control is indicated with "C-" (IMX1338, SinLog). In the lanes indicated with "L", GeneRuler™ DNA ladder mix was loaded. 1% (w/v) agarose in TAE.

Conclusion

The present study offers an innocuous watermarking strategy for coding regions that enables the discrimination of DNA and mRNA by sequencing through a k-mer approach and facilitates selective editing of watermarked and non-watermarked sequences. While the design of watermarked genes was performed manually in the present study, it can easily be automated when a larger number of genes is concerned, with software similar to for example GeneDesign (*66*).

The set of genes chosen to test the watermarking strategy encodes highly abundant proteins that are generally considered to operate at overcapacity, which means that the capacity of the enzymes is considerably larger than the actual *in vivo* flux. While this overcapacity might obscure physiological responses, a closer inspection of transcript levels and enzyme activities confirmed the watermarks harmlessness for *S. cerevisiae*. Only one of the 13 tested genes showed an activity of the enzyme encoded by the

watermarked allele significantly decreased (ca. 1.6-fold) as compared to the activity of the enzyme encoded by the native allele (phosphoglucomutase encoded by *GPM1*), which could be explained by a similar decrease in Gpm1 protein abundance. As native and watermarked transcript levels were identical for *GPM1*, the lower enzyme abundance in the watermarked strain might result from a slightly lower translation efficiency. Neither the watermarking specifics (type of codon substitution, change in codon usage, etc.) of this particular protein, nor information from literature hinted towards the mechanism underlying this decreased protein abundance. A recent study, combining measurements of protein synthesis rate with ribosome footprinting data confirmed that Gpm1, like most glycolytic proteins, has a fast synthesis rate (*67*). Applying the same approach to watermarked strains could help characterizing the impact of nucleotide substitution on translational efficiency.

Another particularity of glycolytic genes is their high codon optimality (on average ca. 90% of optimal codons according to Hanson and Coller (*24*)). While one could argue that this set of genes might not be representative of the yeast genome, to the best of our knowledge there is no evidence that such genes are more or less robust towards changes in codon frequency than genes with lower levels of codon optimality. For future studies it would be interesting to explore if codon optimality affects genes sensitivity to watermarking.

Acknowledgements

We thank Marijke Luttik for assaying the glycolytic enzymes activity, Jordi Geelhoed for the confirmation of selective CRISPR/Cas editing by PCR and sequencing and Pilar de la Torre for RNA samples processing. This work was supported by a consolidator grant from the European Research Council (ERC).

Additional Material

Table S1 – Codon usage in *S. cerevisiae.*

Codons used for codon optimization and their abundance (data from https://www.yeastgenome.org/)

Amino acid	codon	Total nb[a]	/1000[b]	Fraction[c]
Ala	GCG	17988	6.16	0.11
	GCA	47538	16.27	0.3
	GCT	59300	20.29	0.37
	GCC	35410	12.12	0.22
Arg	AGG	27561	9.43	0.21
	AGA	61537	21.06	0.48
	CGG	5299	1.81	0.04
	CGA	9050	3.1	0.07
	CGT	18272	6.25	0.14
	CGC	7644	2.62	0.06
Gly	GGG	17673	6.05	0.12
	GGA	32723	11.2	0.23
	GGT	66198	22.66	0.46
	GGC	28522	9.76	0.2
Leu	TTG	77261	26.44	0.28
	TTA	77615	26.56	0.28
	CTG	31054	10.63	0.11
	CTA	39440	13.5	0.14
	CTT	35753	12.24	0.13
	CTC	16086	5.51	0.06
Pro	CCG	15778	5.4	0.12
	CCA	51993	17.79	0.41
	CCT	39685	13.58	0.31
	CCC	20139	6.89	0.16
Ser	AGT	42657	14.6	0.16
	AGC	29003	9.93	0.11
	TCG	25381	8.69	0.1
	TCA	55725	19.07	0.21
	TCT	68207	23.34	0.26
	TCC	40972	14.02	0.16
Thr	ACG	23766	8.13	0.14
	ACA	53147	18.19	0.31
	ACT	59096	20.23	0.34
	ACC	36395	12.46	0.21
Val	GTG	31266	10.7	0.19
	GTA	35397	12.11	0.22
	GTT	62735	21.47	0.39
	GTC	32738	11.2	0.2

[a] Total number of occurrences of this codon in the yeast genome
[b] Number of occurrences of this codon per 1000 codons in the yeast genome
[c] Fraction of occurrence of this codon usage from the set of codons representing the same amino acid

Table S2 – Watermarked glycolytic genes.

List of watermarks introduced in the glycolytic and fermentative genes and of the resulting change in codon usage

Gene	Position (bp)	Native codon	WM codon	Fraction of native codon[a]	Fraction of WM codon[b]	Change in fraction (a-b)
TPI1	375	TTG	TTA	0.28	0.28	0
	408	GCC	GCA	0.22	0.3	0.08
	420	TTG	TTA	0.28	0.28	0
	441	TTG	TTA	0.28	0.28	0
	453	TTG	TTA	0.28	0.28	0
	645	AGC	AGT	0.11	0.16	0.05
	651	GCC	GCA	0.22	0.3	0.08
	657	ACC	ACA	0.21	0.31	0.1
	678	GTC	GTA	0.2	0.22	0.02
	693	GTC	GTA	0.2	0.22	0.02
	708	TTG	TTA	0.28	0.28	0
	Sum of all substitutions[c]					**0.35**
FBA1	558	GTT	GTC	0.39	0.2	-0.19
	573	GCT	GCC	0.37	0.22	-0.15
	588	TTG	TTA	0.28	0.28	0
	618	GTC	GTT	0.2	0.39	0.19
	630	TTG	TTA	0.28	0.28	0
	1005	GTT	GTC	0.39	0.2	-0.19
	999	GTC	GTT	0.2	0.39	0.19
	1023	ACC	ACA	0.21	0.31	0.1
	1041	ACC	ACA	0.21	0.31	0.1
	1065	ACC	ACA	0.21	0.31	0.1
	Sum of all substitutions					**1.21**
GPM1	381	TCT	TCA	0.26	0.21	-0.05
	393	TCT	TCA	0.26	0.21	-0.05
	435	GTC	GTA	0.2	0.22	0.02
	447	ACT	ACA	0.34	0.31	-0.03
	456	TTG	TTA	0.28	0.28	0
	462	TTG	TTA	0.28	0.28	0
	639	GTC	GTA	0.2	0.22	0.02
	660	TTG	TTA	0.28	0.28	0
	699	GCT	GCC	0.37	0.22	-0.15
	702	GCC	GCT	0.22	0.37	0.15
	726	GCC	GCA	0.22	0.3	0.08
	Sum of all substitutions					**0.55**
HXK2	735	TCC	TCA	0.16	0.21	0.05
	750	CTA	CTT	0.14	0.13	-0.01
	762	CTA	CTT	0.14	0.13	-0.01
	795	GCC	GCA	0.22	0.3	0.08
	816	TCC	TCA	0.16	0.21	0.05
	1365	CCT	CCC	0.31	0.16	-0.15
	1380	TCC	TCA	0.16	0.21	0.05
	1392	GCC	GCA	0.22	0.3	0.08
	1413	GCC	GCA	0.22	0.3	0.08
	1440	TCC	TCA	0.16	0.21	0.05
	Sum of all substitutions					**0.61**
PDC1	885	GTC	GTA	0.2	0.22	0.02
	867	TTG	TTA	0.28	0.28	0

	882	ACC	ACA	0.21	0.31	0.1
	909	ACC	ACA	0.21	0.31	0.1
	933	TCC	TCA	0.16	0.21	0.05
	1593	TCT	TCA	0.26	0.21	-0.05
	1626	GTC	GTA	0.2	0.22	0.02
	1647	TTG	TTA	0.28	0.28	0
	1665	TTG	TTA	0.28	0.28	0
	1677	ACC	ACA	0.21	0.31	0.1
	Sum of all substitutions					**0.44**
PFK1	1506	TTA	TTG	0.28	0.28	0
	1512	ACT	ACA	0.34	0.31	-0.03
	1515	CTA	CTT	0.14	0.13	-0.01
	1545	CTG	CTT	0.11	0.13	0.02
	1554	ACC	ACA	0.21	0.31	0.1
	1563	ACT	ACA	0.34	0.31	-0.03
	1575	TTA	TTG	0.28	0.28	0
	2880	GCT	GCC	0.37	0.22	-0.15
	2907	CTG	CTT	0.11	0.13	0.02
	2910	TCC	TCA	0.16	0.21	0.05
	2937	GTA	GTG	0.22	0.19	-0.03
	2940	GCC	GCA	0.22	0.3	0.08
	2946	TTA	TTG	0.28	0.28	0
	2949	GCC	GCA	0.22	0.3	0.08
	Sum of all substitutions					**0.6**
PFK2	1461	ACT	ACA	0.34	0.31	-0.03
	1479	CGT	CGA	0.14	0.07	-0.07
	1494	TTA	TTG	0.28	0.28	0
	1503	CTT	CTA	0.13	0.14	0.01
	1530	TCC	TCA	0.16	0.21	0.05
	2775	TTG	TTA	0.28	0.28	0
	2799	TCC	TCA	0.16	0.21	0.05
	2817	GTC	GTA	0.2	0.22	0.02
	2841	CTC	CTG	0.06	0.11	0.05
	2862	GGA	GGC	0.23	0.2	-0.03
	Sum of all substitutions					**0.31**
PGI1	837	GTC	GTA	0.2	0.22	0.02
	855	GTC	GTA	0.2	0.22	0.02
	861	TCG	TCC	0.1	0.16	0.06
	882	GCC	GCA	0.22	0.3	0.08
	894	GGC	GGA	0.2	0.23	0.03
	1563	GTC	GTA	0.2	0.22	0.02
	1581	GGC	GGA	0.2	0.23	0.03
	1599	TCC	TCA	0.16	0.21	0.05
	1611	TCT	TCA	0.26	0.21	-0.05
	1629	ACC	ACA	0.21	0.31	0.1
	Sum of all substitutions					**0.46**
PGK1	630	TTA	TTG	0.28	0.28	0
	639	GCC	GCA	0.22	0.3	0.08
	663	TTG	TTA	0.28	0.28	0
	687	GTC	GTA	0.2	0.22	0.02
	693	TCT	TCA	0.26	0.21	-0.05
	1164	TCC	TCA	0.16	0.21	0.05
	1176	ACT	ACA	0.34	0.31	-0.03
	1200	TTA	TTG	0.28	0.28	0
	1218	TTG	TTA	0.28	0.28	0
	1239	TCC	TCA	0.16	0.21	0.05
	Sum of all substitutions					**0.28**

PYK1	762	TTG	TTA	0.28	0.28	0
	789	GCC	GCA	0.22	0.3	0.08
	819	GCC	GCA	0.22	0.3	0.08
	837	GTC	GTA	0.2	0.22	0.02
	849	TTG	TTA	0.28	0.28	0
	1419	TTG	TTA	0.28	0.28	0
	1443	TCC	TCA	0.16	0.21	0.05
	1461	GCC	GCA	0.22	0.3	0.08
	1476	TCC	TCA	0.16	0.21	0.05
	1491	GTC	GTA	0.2	0.22	0.02
	Sum of all substitutions					**0.38**
TDH3	516	TTG	TTA	0.28	0.28	0
	522	ACC	ACA	0.21	0.31	0.1
	555	ACT	ACA	0.34	0.31	-0.03
	570	TCC	TCA	0.16	0.21	0.05
	597	ACC	ACA	0.21	0.31	0.1
	903	TTG	TTA	0.28	0.28	0
	918	GTC	GTA	0.2	0.22	0.02
	930	TCC	TCA	0.16	0.21	0.05
	960	ACC	ACA	0.21	0.31	0.1
	975	TTG	TTA	0.28	0.28	0
	Sum of all substitutions					**0.45**
ADH1	531	TCC	TCA	0.16	0.21	0.05
	549	CTA	CTT	0.14	0.13	-0.01
	558	TTG	TTA	0.28	0.28	0
	573	GCC	GCA	0.22	0.3	0.08
	597	TTG	TTA	0.28	0.28	0
	939	TTG	TTA	0.28	0.28	0
	942	GTC	GTA	0.2	0.22	0.02
	966	GGC	GGA	0.2	0.23	0.03
	978	TTG	TTA	0.28	0.28	0
	1029	GTT	GTC	0.39	0.2	-0.19
	1032	GTT	GTC	0.39	0.2	-0.19
	Sum of all substitutions					**0.57**
ENO2	663	ACC	ACA	0.21	0.31	0.1
	678	TTG	TTA	0.28	0.28	0
	684	TTG	TTA	0.28	0.28	0
	726	GTC	GTA	0.2	0.22	0.02
	738	TTG	TTA	0.28	0.28	0
	750	TCC	TCA	0.16	0.21	0.05
	1212	TCC	TCA	0.16	0.21	0.05
	1242	TTG	TTA	0.28	0.28	0
	1260	TTG	TTA	0.28	0.28	0
	1272	GCT	GCC	0.37	0.22	-0.15
	1281	GCC	GCA	0.22	0.3	0.08
	Sum of all substitutions					**0.45**

[a] and [b], fraction of occurrence of this codon is used from the set of codons representing the same amino acid according to Table S1

[c], sum of all substitutions for each watermarked gene, calculated as $\sum_{i=1}^{n}(|a_i - b_i|)$

Table S3 – Comparing Alignment and markerQuant for differential quantification of watermarked and native glycolytic transcripts. Here, we show, for simulated data, the ability of markerQuant and STAR (*68*) to retrieve the real count of reads originating from either the native sequence, or the watermarked sequence. As for evaluating the watermarking method, we simulated RNA-Sequencing reads using the polyester R package (*56*). We compared the % retrieval between STAR and markerQuant. For STAR, we removed reads that aligned to more than one transcript using samtools (*69*). This results in a lower retrieval fraction of the generated reads using STAR than when using markerQuant. It shows that while STAR may also be able to discern between the native and watermarked transcripts, markerQuant is able to retrieve a higher fraction of the reads.

	Generated counts				Star Quant				MarkerQuant				Star % Retrieved					MarkerQuant % Retrieved				
	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	Mean	S1	S2	S3	S4	Mean
ADH1_coding	158	195	751	838	74	79	317	392	85	95	377	452	46,8	40,5	42,2	46,8	44,1	53,8	48,7	50,2	53,9	51,7
ADH1_w	725	828	167	191	315	368	70	89	375	433	84	108	43,4	44,4	41,9	46,6	44,1	51,7	52,3	50,3	56,5	52,7
ADH3_coding	246	238	957	994	102	101	383	385	121	121	456	475	41,5	42,4	40,0	38,7	40,7	49,2	50,8	47,6	47,8	48,9
ADH3_w	977	917	288	248	401	352	120	99	458	430	141	116	41,0	38,4	41,7	39,9	40,3	46,9	46,9	49,0	46,8	47,4
ENO2_coding	237	238	935	971	90	70	318	339	106	93	381	416	38,0	29,4	34,0	34,9	34,1	44,7	39,1	40,7	42,8	41,8
ENO2_w	1055	1018	273	211	327	361	98	64	408	414	116	80	31,0	35,5	35,9	30,3	33,2	38,7	40,7	42,5	37,9	39,9
FBA1_coding	195	223	854	868	77	98	339	344	103	115	395	402	39,5	43,9	39,7	39,6	40,7	52,8	51,6	46,3	46,3	49,2
FBA1_w	793	822	271	219	321	336	99	87	369	402	118	102	40,5	40,9	36,5	39,7	39,4	46,5	48,9	43,5	46,6	46,4
GPM1_coding	127	130	633	635	74	81	370	384	91	93	446	450	58,3	62,3	58,5	60,5	59,9	71,7	71,5	70,5	70,9	71,1
GPM1_w	599	596	134	139	344	369	86	78	428	435	100	104	57,4	61,9	64,2	56,1	59,9	71,5	73,0	74,6	74,8	73,5
HXK2_coding	240	288	1084	1200	66	75	339	351	77	96	409	412	27,5	26,0	31,3	29,3	28,5	32,1	33,3	37,7	34,3	34,4
HXK2_w	1170	1162	277	218	342	337	82	61	400	408	95	71	29,2	29,0	29,6	28,0	29,0	34,2	35,1	34,3	32,6	34,0
PDC1_coding	360	368	1331	1451	80	84	318	346	93	99	392	411	22,2	22,8	23,9	23,8	23,2	25,8	26,9	29,5	28,3	27,6
PDC1_w	1340	1343	307	323	336	323	88	86	397	390	104	101	25,1	24,1	28,7	26,6	26,1	29,6	29,0	33,9	31,3	31,0
PFK1_coding	633	663	2523	2413	92	60	315	332	100	86	370	391	14,5	9,0	12,5	13,8	12,5	15,8	13,0	14,7	16,2	14,9
PFK1_w	2087	2388	609	598	286	286	88	61	341	337	101	75	13,7	12,0	14,4	10,2	12,6	16,3	14,1	16,6	12,5	14,9
PFK2_coding	514	561	2367	2342	58	84	305	298	77	103	377	365	11,3	15,0	12,9	12,7	13,0	15,0	18,4	15,9	15,6	16,2
PFK2_w	2174	2316	608	542	273	307	78	92	316	358	87	104	12,6	13,3	12,8	17,0	13,9	14,5	15,5	14,3	19,2	15,9
PGI1_coding	381	275	1302	1202	94	59	300	287	123	72	365	332	24,7	21,5	23,0	23,9	23,3	32,3	26,2	28,0	27,6	28,5
PGI1_w	1228	1371	311	288	276	332	68	77	330	394	77	89	22,5	24,2	21,9	26,7	23,8	26,9	28,7	24,8	30,9	27,8
PGK1_coding	264	246	945	1015	100	84	327	352	112	105	381	410	37,9	34,1	34,6	34,7	35,3	42,4	42,7	40,3	40,4	41,5
PGK1_w	941	1019	245	173	329	329	78	51	382	379	95	64	35,0	32,3	31,8	29,5	32,1	40,6	37,2	38,8	37,0	38,4
PYK1_coding	284	317	1408	1314	81	94	389	370	100	107	473	431	28,5	29,7	27,6	28,2	28,5	35,2	33,8	33,6	32,8	33,8
PYK1_w	1283	1115	285	291	379	318	77	74	443	381	88	94	29,5	28,5	27,0	25,4	27,6	34,5	34,2	30,9	32,3	33,0
TDH3_coding	249	164	797	745	108	78	379	335	134	91	444	413	43,4	47,6	47,6	45,0	45,9	53,8	55,5	55,7	55,4	55,1
TDH3_w	817	861	214	219	374	401	85	108	444	473	108	126	45,8	46,6	39,7	49,3	45,3	54,3	54,9	50,5	57,5	54,3
TPI1_coding	144	138	643	523	94	82	400	339	112	99	477	399	65,3	59,4	62,2	64,8	62,9	77,8	71,7	74,2	76,3	75,0
TPI1_w	605	605	150	159	377	367	94	93	446	443	117	105	62,3	60,7	62,7	58,5	61,0	73,7	73,2	78,0	66,0	72,7
													Mean				**35,0**	**Mean**				**41,8**

Table S4 – Plasmids used in this study available at doi: 10.1021/acssynbio.0c00045

Table S5 – Whole genome sequencing of the constructed strains.

List of mutations in glycolytic expression cassettes and helper elements in IMX1770, IMX1771 and IMX2028 as compared to the parental strain IMX589.

	Fragment	Location	Mutation
IMX1770	*PFK1* promoter	-581	A > AT
	PGK1 terminator	211	G > GA
	ENO2 terminator	280	C > A
	HIS3 promoter	-132	C > CA
IMX1771	*GPM1* promoter	-459	CTATATATATA > C
	HIS3 terminator	114	C > CA
IMX2028	*GPM1* promoter	-459	CTATATATATA > C
	HIS3 terminator	114	C > CA
	FBA1 terminator*	294	CA>C
	PFK2 promoter*	-22	A>G

* In watermarked glycolysis cassette.

Table S6 – gRNA sequence for selective gene editing. For each native glycolytic gene targeted for DNA editing, the PAM is underlined. In the corresponding watermarked gene, the bases modified to prevent editing by Cas9 are shown in bold font. Only gRNA's for *PYK1* and *TPI1* were used in this study.

Gene	Native gRNA sequence	Watermarked gRNA sequence
PGK1	TTGAATCTTGTCAGCAACCTT**G**G	TTGAATCTTGTCAGCAACCTT**T**G
PYK1	ACTTAGCAAT**C**AATTTCTTTTG**G**	ACTTAGCAAT**T**AATTTCTTTTG**T**
PDC1	GTGGAATTCGACAATGTTCTT**G**G	GTGGAATTCGACAATGTTCT**T**G
TPI1	AACAACATC**C**AAAGTCTTACC**G**G	AACAACATC**T**AAAGTCTTACC**T**G

Table S7 – Primers is available at doi: 10.1021/acssynbio.0c00045.

Table S8 – RNA-Sequencing Depth.

Strain	Replicate	Depth
Native	1	22089278
Native	2	22216419
Native	3	24257740
Watermarked	1	24070180
Watermarked	2	24780729
Watermarked	3	23334043
Double	1	23480289
Double	2	25448312
Double	3	23022264

Table S9 – Sampling OD for RNA seq.

Optical density of IMX1770, IMX1771 and IMX2028 batch cultures analysed by RNA sequencing

Strain	Reactor	OD_{660}
IMX1770	27	7,8
IMX1770	28	9,95
IMX1770	29	8,25
IMX1771	27	7,75
IMX1771	28	8,05
IMX1771	29	7,95
IMX2028	27	4,5
IMX2028	29	4,51
IMX2028	30	5,79

A

```
ADH1 native      1 atgtctatcccagaaactcaaaaaggtgttatcttctacgaatcccacggtaagttggaatacaaagatattccagttccaaagccaaaggccaacgaattgttgatcaacgttaaatactctggtgtctgtcacactgacttgcacgcttggcacggtgactggccattgccagttaagctaccattagtcggtggtcacgaaggtgcc
ADH1_wm1         1 ........................................................................................................................................................................

ADH1 native    211 ggtgtcgttgtcggcatgggtgaaaacgttaagggctggaagatcggtgactacgccggtatcaaatggttgaacggttcttgtatggcctgtgaatactgtgaattgggtaacgaatccaactgtcctcacgctgacttgtctggttacacccatgacggttctttccaacaatacgctactgctgacgcggtgcaagccgctcgtatt
ADH1_wm1       211 ........................................................................................................................................................................

ADH1 native    421 cccgaagggaccgacttggcccaagtcgcccccatcttgtgtgctggtatcaccgtctacaaggctttgaagtctgctaacttgatggccggtcactgggttgctatctccggtgctgctggtggtctaggttctttggctgttcaatacgccaaggctatgggttacagagtcttgggtattgacggtggtgaaggtaaggaagaatta
ADH1_wm1       421 ...................................................................................a...................t........a...............a...........................a...............................

ADH1 native    631 ttcagatccatcggtggtgaagtcttcattgacttcactaaggaaaaggacattgtcggtgctgttctaaaggccactgacggtggtgctcacggtgtcatcaacgtttccgtttccgaagccgctattgaagcttctaccagatacgttagagctaacggtaccaccgttttggtcggtatgccagctggtgccaagtgttgttctgat
ADH1_wm1       631 ........................................................................................................................................................................

ADH1 native    841 gtcttcaaccaagtcgtcaagtccatctctattgttggttcttacgtcggtaacagagctgacaccagagaagctttggacttcttcgccagaggtttggtcaagtctccaatcaaggttgtcggcttgtctaccttgccagaaatttacgaaaagatggaaaagggtcaaatcgttggtagatacgttgttgacacttctaaataa
ADH1_wm1       841 ...............................................................................a..a.......................a............a.....................................c..c...............


ADH1 native      1 atgtctatcccagaaactcaaaaaggtgttatcttctacgaatcccacggtaagttggaatacaaagatattccagttccaaagccaaaggccaacgaattgttgatcaacgttaaatactctggtgtctgtcacactgacttgcacgcttggcacggtgactggccattgccagttaagctaccattagtcggtggtcacgaaggtgcc
ADH1_wm2         1 ...................................................................................a.....................................................................a..................

ADH1 native    211 ggtgtcgttgtcggcatgggtgaaaacgttaagggctggaagatcggtgactacgccggtatcaaatggttgaacggttcttgtatggcctgtgaatactgtgaattgggtaacgaatccaactgtcctcacgctgacttgtctggttacaccCcatgacggttctttccaacaatacgctactgctgacgcggtgcaagccgctcgtatt
ADH1_wm2       211 ......................................................................a.....................................................g.......................................................

ADH1 native    421 cccgaagggaccgacttggcccaagtcgcccccatcttgtgtgctggtatcaccgtctacaaggctttgaagtctgctaacttgatggccggtcactgggttgctatctccggtgctgctggtggtctaggttctttggctgttcaatacgccaaggctatgggttacagagtcttgggtattgacggtggtgaaggtaaggaagaatta
ADH1_wm2       421 ..........................a.......................................................................a.......................................................................g

ADH1 native    631 ttcagatccatcggtggtgaagtcttcattgacttcactaaggaaaaggacattgtcggtgctgttctaaaggccactgacggtggtgctcacggtgtcatcaacgtttccgtttccgaagccgctattgaagcttctaccagatacgttagagctaacggtaccaccgttttggtcggtatgccagctggtgccaagtgttgttctgat
ADH1_wm2       631 ......................................................................a.......................................................................a..................

ADH1 native    841 gtcttcaaccaagtcgtcaagtccatctctattgttggttcttacgtcggtaacagagctgacaccagagaagctttggacttcttcgccagaggtttggtcaagtctccaatcaaggttgtcggcttgtctaccttgccagaaatttacgaaaagatggaaaagggtcaaatcgttggtagatacgttgttgacacttctaataa
ADH1_wm2       841 ..............................................a.......................................................a.......................................................
```

B

Figure S1 - *In silico* comparison of two watermarking approaches for *ADH1.*

Alignment of the native and watermarked *ADH1* sequence using watermarking strategy 1 (two watermarking areas, **A**) or 2 (watermarks evenly spread over the coding region, **B**). Mismatches corresponding to the watermarks are highlighted in orange.

A

```
FBA1_native          1 atgggtgttgaacaaatcttaaagagaaagaccggtgtcatcgttggtgaagatgtccacaacttattcacttacgctaaggaacacaagttcgctattccagctattaacgtcacctcttcttctactgccgtcgctgctttagaagctgctagagacagcaagtccccaatcattttgcaaacctctaacggtggtgctgcttacttc
FBA1_watermark       1 ....................................................................................................................................................................................

FBA1_native        211 gctggtaagggtatctctaacgaaggtcaaaatgcttccatcaagggtgctattgccgctgcccactacatcagatccattgctccagcttacggtatcccagttgtcttacactctgaccactgtgccaagaagttgttgccatggttcgatggtatgttggaagctgatgaagcttacttcaaggaacacggtgaaccattattctcc
FBA1_watermark     211 ....................................................................................................................................................................................

FBA1_native        421 tcccacatgttggatttgtctgaagaaaccgatgaagaaaacatctctacttgtgtcaagtacttcaagagaatggccgctatggaccaatggttagaaatggaaatcggtattaccggtggtgaagaagatggtgttaacaacgaaaacgctgacaaggaagacttgtacaccaagccagaacaagtttacaacgtctacaaggctttg
FBA1_watermark     421 ....................................................................................................c...............a.....a.......................a...........a

FBA1_native        631 cacccaatctctccaaacttctccattgctgctgctttcggtaactgtcacggtttgtacgctggtgacatcgctttgagaccagaaatcttggctgaacaccaaaagtacacccagagaacaagttggttgcaaggaagaaaagccattgttcttggtcttccacggtggttccggttctactgtccaagaattccacactggtattgac
FBA1_watermark     631 ....................................................................................................................................................................................

FBA1_native        841 aacggtgttgtcaaggtcaacttggacactgactgtcaatacgcttacttgactggtatcagagactacgtcttgaacaagaaggactacataatgtccccagtcggtaacccagaaggtccagaaaagccaaacaagaagttcttcgacccaagagtctgggttagagaaggtgaaaagaccatgggtgctaagatcaccaagtctttg
FBA1_watermark     841 ....................................................................................................................a.......................a.................a.........

FBA1_native       1051 gaaactttccgtaccactaacactttataa
FBA1_watermark    1051 .....a........a..............
```

B

```
ENO2_native          1 atggctgtctctaaagtttacgctagatccgtctacgactcccgtggtaacccaaccgtcgaagtcgaattaaccaccgaaaagggtgttttcagatccattgttccatctggtgcctccacccggtgtccacgaagctttggaaatgagagatgaagacaaatccaagtggatgggtaaggggtgttatgaacgctgtcaacaacgtcaac
ENO2_watermark       1 ....................................................................................................................................................................................

ENO2_native        211 aacgtcattgctgctgctttcgtcaaggccaacctagatgttaaggaccaaaaggccgtcgatgacttcttgttgtctttggatggtaccgccaacaagtccaagttgggtgctaacgctatcttgggtgtctccatggccgctgctagagccgctgctgctgaaaagaacgtcccattgtaccaacatttggctgacttgtctaagtcc
ENO2_watermark     211 ....................................................................................................................................................................................

ENO2_native        421 aagacctctccatacgttttgccagttccattcttgaacgttttgaacggtggttcccacgctggtggtgctttggctttgcaagaattcatgattgctccaactggtgctaagaccttcgctgaagccatgagaattggttccgaagtttaccacaacttgaagtctttgaccaagaagagatacggtgcttctgccggtaacgtcggt
ENO2_watermark     421 ....................................................................................a...................a............a.........................a.........a...

ENO2_native        631 gacgaaggtggtgttgctcccaaacattcaaaccgctgaagaagctttggacttgattgttgacgctatcaaggctgctggtcacgacggtaaggtcaagatcggtttggactgtgcttcctctgaattcttcaaggacggtaagtacgacttggacttcaagaacccagaatctgacaaatccaagtggttgactggtgtcgaattagct
ENO2_watermark     631 ....................................................................................................................................................................................

ENO2_native        841 gacatgtaccactccttgatgaagagatacccaattgtctccatcgaagatccatttgctgaagatgactgggaagcttggtctcacttcttcaagaccgctggtatccaaattgttgctgatgacttgactgtcaccaacccagctagaattgctaccgccatcgaaaagaaggctgctgacgctttgttgttgaaggttaaccaaatc
ENO2_watermark     841 ....................................................................................................................................................................................

ENO2_native       1051 ggtacottgtctgaatccatcaaggctgctcaagactctttcgctgccaactgggggtgttatggtttcccacagatctggtgaaactgaagacactttcattgctgacttggttgtcggtttgagaactggtcaaatcaagactggtgctccagctagatccgaaagattggctaagttgaaccaattgttgagaatcgaagaagaattg
ENO2_watermark    1051 ....................................................................................................a........a...................a.....................a

ENO2_native       1261 ggtgacaaggctgtctacgccggtgaaaacttccaccacggtgacaagttgtaa
ENO2_watermark    1261 ....................a...............................
```

Figure S3 – Watermarking of *FBA1* and *ENO2*.

Alignment of the sequence of the coding region of native versus watermarked *FBA1* **(A)** and *ENO2* **(B)**. Watermarks are highlighted in orange.

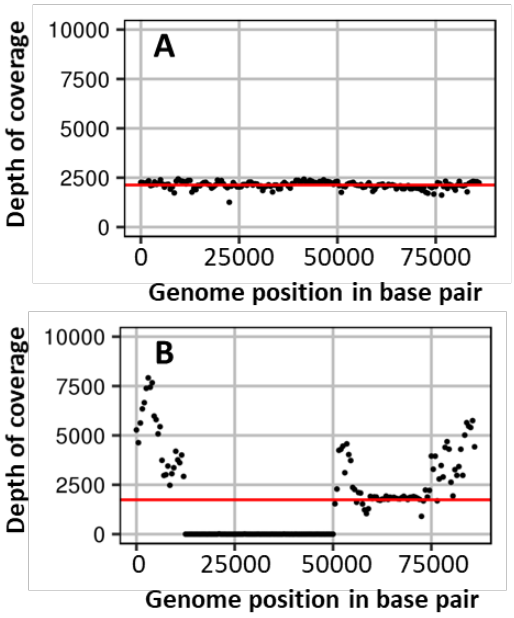


Figure S4 – Genetic characterization of IMX2028.

A) Coverage plot of mapped whole genome sequence reads of IMX1771 against the reference strain CEN.PK113-7D, only representing the mitochondrial genome. The uniform coverage around 2500 shows that IMX1771 mitochondrial genome is intact. **B)** Same plot as A showing the coverage plot of the mitochondrial DNA of IMX2028 mapped against the reference strain CEN.PK113-7D. A large fraction of IMX2028 mitochondrial genome is missing. Black dots are averages over 500 bp non-overlapping windows and the red line depicts the median of the 500 bp non-overlapping windows.

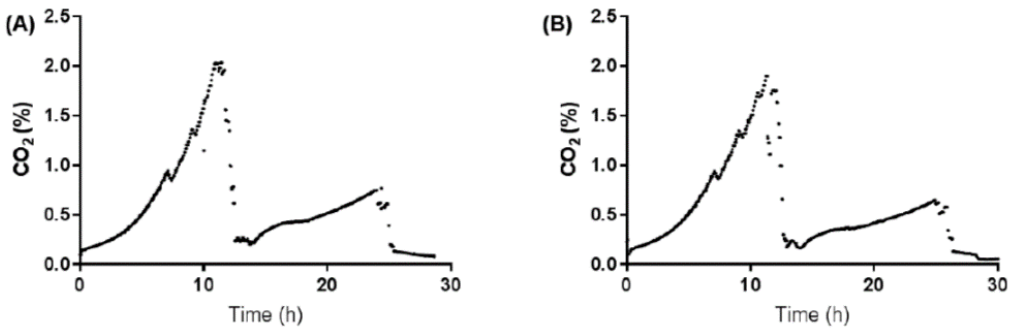


Figure S5 – CO_2 profiles of batch cultures with IMX1770 and IMX1771.

Growth profiles based on the CO_2 off-gas data of IMX1770 **A)** and IMX1771 **B)** when grown in aerobic batch cultures in bioreactors with glucose as sole carbon source. After approximately ten hours, glucose was depleted and the strains switched to consumption of the fermentation product ethanol.

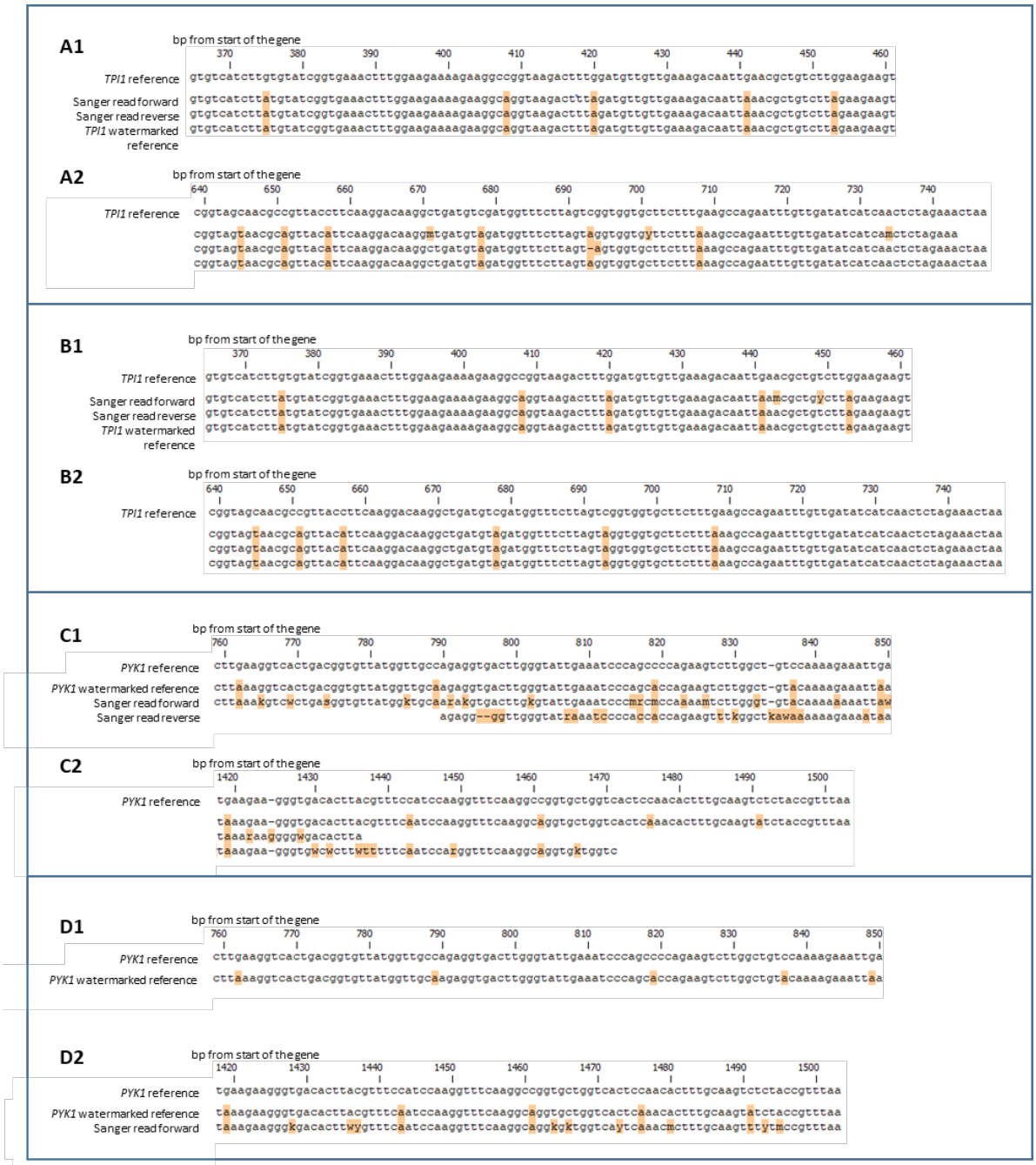


Figure S6 – Confirmation of selective DNA editing by Sanger sequencing. For *TPI1* (Fig. 7) the band of 1125 bp in lane 9 and 10 was Sanger sequenced. **A1** and **B1:** watermark in the middle of *TPI1* for colony 9 and 10 respectively. **A2** and **B2**, sequence at the end of *TPI1* until the stop codon for colony 9 and 10 respectively. For *PYK1*, the band of 2177 bp for colonies number 1 and 5 (Fig. 7A) was Sanger sequenced. **C1** and **D1:** watermark region in the middle of *PYK1* for colony 1 and 5, respectively. **C2** and **D2**, watermark region at the end of *PYK1* until the stop codon for colony 1 and 5 respectively.

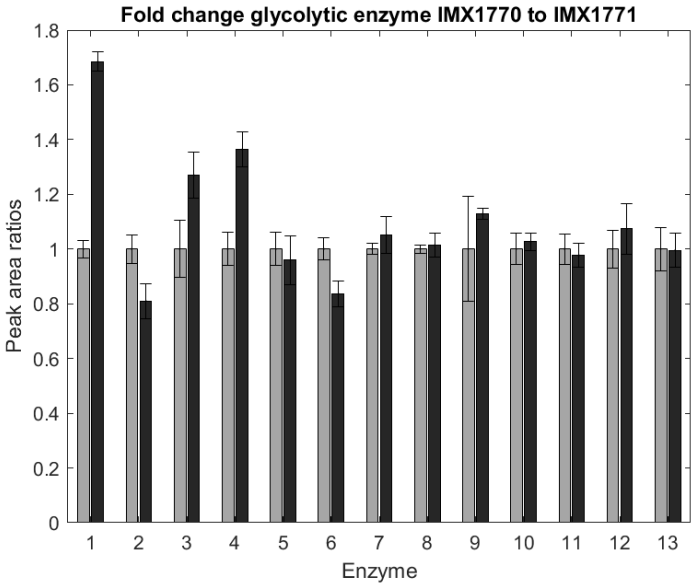


Figure S7 – Label free quantification (LFQ) of glycolytic protein abundance of watermarked (IMX1770, grey bars) and non-watermarked strain (IMX1771, black bars). Data represent the average and standard deviation for three biological replicates. 1=Gpm1, 2=Tdh3, 3=Eno2, 4=Pyk1, 5=Tpi1, 6=Pfk2, 7=Hxk2, 8=Pfk1, 9=Fba1, 10=Pdc1, 11=Pgk1, 12=Pgi1, 13=Adh1. The fold change for Gpm1 is significantly different between the strains (Student t-test, p-value threshold 0.05, two-tailed test, homoscedastic).

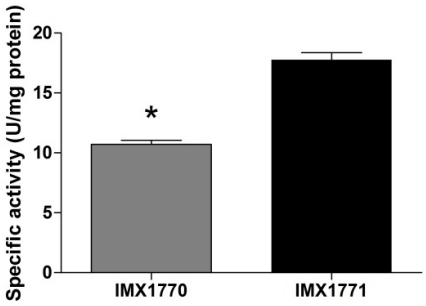


Figure S8 – Specific activity of Gpm1. Specific enzyme activity of the Gpm1 enzyme from IMX1770 (watermarked) and IMX1770 (native) strains. Samples were taken in mid-exponential phase. Bars represent the average and standard deviation of measurements from three independent shake flasks for each strain. Stars indicate enzyme activities that are significantly different between the two strains (Student t-test, p-value threshold 0.05, two-tailed test, homoscedastic).

References

1. S. Jiao, R. Goutte, in *International Conference on Signal Processing Proceedings, ICSP*. (IEEE, 2008), pp. 2166-2169.
2. M. Arita, Y. Ohashi, Secret signatures inside genomic DNA. *Biotechnol. Prog.* **20**, 1605-1607 (2004).
3. P. C. Wong, K. Wong, H. Foote, Organic data memory using the DNA approach. *Commun. ACM* **46**, 95-98 (2003).
4. C. T. Clelland, V. Risca, C. Bancroft, Hiding messages in DNA microdots. *Nature* **399**, 533-534 (1999).
5. D. G. Gibson *et al.*, Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52-56 (2010).
6. D. Heider, A. Barnekow, DNA watermarks: a proof of concept. *BMC Mol. Biol.* **9**, 40 (2008).
7. M. Liss *et al.*, Embedding permanent watermarks in synthetic genes. *PLoS One* **7**, e42465 (2012).
8. N. Yamamoto, H. Kajiura, S. Takeno, N. Suzuki, Y. Nakazawa, A watermarking system for labeling genomic DNA. *Plant Biotechnol.* **31**, 241-248 (2014).
9. N. Annaluru *et al.*, Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58 (2014).
10. J. Fredens *et al.*, Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514-518 (2019).
11. I. S. Pretorius, J. D. Boeke, Yeast 2.0-connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Res.* **18**, (2018).
12. J. S. Dymond *et al.*, Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476 (2011).
13. N. G. Kuijpers *et al.*, Pathway swapping: Toward modular engineering of essential cellular processes. *Proc Natl Acad Sci U S A* **113**, 15060-15065 (2016).
14. D. Solis-Escalante *et al.*, A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell* **14**, 804-816 (2015).
15. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
16. A. C. Roth, Decoding properties of tRNA leave a detectable signal in codon usage bias. *Bioinformatics* **28**, i340-i348 (2012).
17. S. Kanaya, Y. Yamada, Y. Kudo, T. Ikemura, Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* **238**, 143-155 (1999).
18. E. M. Novoa, L. Ribas de Pouplana, Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* **28**, 574-581 (2012).
19. P. M. Sharp, W. H. Li, An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**, 28-38 (1986).
20. V. Presnyak *et al.*, Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111-1124 (2015).
21. S. Pechmann, J. Frydman, Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* **20**, 237-243 (2013).
22. P. Cortazzo *et al.*, Silent mutations affect *in vivo* protein folding in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **293**, 537-541 (2002).
23. G. Cannarozzi *et al.*, A role for codon order in translation dynamics. *Cell* **141**, 355-367 (2010).
24. G. Hanson, J. Coller, Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* **19**, 20-30 (2018).
25. G. Kudla, A. W. Murray, D. Tollervey, J. B. Plotkin, Coding-sequence determinants of gene expression in *Escherichia coli. Science* **324**, 255-258 (2009).
26. J. V. Chamary, L. D. Hurst, Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol* **6**, R75 (2005).
27. J. V. Chamary, L. D. Hurst, Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else? *Trends Genet.* **21**, 256-259 (2005).
28. L. Cartegni, S. L. Chew, A. R. Krainer, Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* **3**, 285-298 (2002).

29. S. K. Nandy, R. K. Srivastava, A review on sustainable yeast biotechnological processes and applications. *Microbiol. Res.* **207**, 83-90 (2018).
30. Z. Liu, Y. Zhang, J. Nielsen, Synthetic Biology of Yeast. *Biochemistry* **58**, 1511-1520 (2019).
31. S. M. Richardson *et al.*, Design of a synthetic yeast genome. *Science* **355**, 1040-1044 (2017).
32. Y. Shen *et al.*, Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355**, (2017).
33. Z. X. Xie *et al.*, "Perfect" designer chromosome V and behavior of a ring derivative. *Science* **355**, (2017).
34. L. A. Mitchell *et al.*, Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355**, (2017).
35. Y. Wu *et al.*, Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355**, (2017).
36. W. Zhang *et al.*, Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355**, (2017).
37. M. A. Martinez, A. Jordan-Paiz, S. Franco, M. Nevot, Synonymous genome recoding: a tool to explore microbial biology and new therapeutic strategies. *Nucleic Acids Res.* **47**, 10506-10519 (2019).
38. T. Gehrmann *et al.*, Nucleus-specific expression in the multinuclear mushroom-forming fungus Agaricus bisporus reveals different nuclear regulatory programs. *Proc Natl Acad Sci U S A* **115**, 4429-4434 (2018).
39. K. D. Entian, P. Kötter, in *Yeast Gene Analysis,* I. Stansfield, M. J. R. Stark, Eds. (Academic Press, Elsevier, Amsterdam, 2007), vol. 36, chap. 25, pp. 629-666.
40. A. N. Salazar *et al.*, Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.* **17**, 10.1093/femsyr/fox1074 (2017).
41. J. F. Nijkamp *et al.*, *De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* **11**, (2012).
42. J. T. Pronk, Auxotrophic yeast strains in fundamental and applied research. *Appl. Environ. Microbiol.* **68**, 2095-2100 (2002).
43. D. Solis-Escalante *et al.*, amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res* **13**, 126-139 (2013).
44. M. E. Lee, W. C. DeLoache, B. Cervantes, J. E. Dueber, A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol* **4**, 975-986 (2015).
45. R. Mans *et al.*, CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **15**, (2015).
46. R. D. Gietz, R. A. Woods, Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**, 87-96 (2002).
47. J. E. DiCarlo *et al.*, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336-4343 (2013).
48. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
49. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
50. J. F. Nijkamp *et al.*, *De novo* detection of copy number variation by co-assembly. *Bioinformatics* **28**, 3195-3202 (2012).
51. E. Postma, C. Verduyn, W. A. Scheffers, J. P. van Dijken, Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol* **55**, 468-477 (1989).
52. M. D. Piper *et al.*, Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of Saccharomyces cerevisiae. *J Biol Chem* **277**, 37001-37008 (2002).
53. M. L. A. Jansen *et al.*, Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity. *Microbiology-SGM* **151**, 1657-1669 (2005).
54. A. L. Cruz *et al.*, Use of sequential-batch fermentations to characterize the impact of mild hypothermic temperatures on the anaerobic stoichiometry and kinetics of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng* **109**, 1735-1744 (2012).

55. O. H. Lowry, H. J. Rosebrough, A. L. Farr, R. J. Randall, Protein measurement with the folin phenol reagent. *J. Biol. Chem* **193**, 265-275 (1951).
56. A. C. Frazee, A. E. Jaffe, B. Langmead, J. T. Leek, Polyester: simulating RNA-seq datasets with differential transcript expression. *Bioinformatics* **31**, 2778-2784 (2015).
57. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
58. J. Koster, S. Rahmann, Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520-2522 (2012).
59. T. Kocher, P. Pichler, R. Swart, K. Mechtler, Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. *Nat Protoc* **7**, 882-890 (2012).
60. T. Valikangas, T. Suomi, L. L. Elo, A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief Bioinform* **19**, 1344-1355 (2018).
61. D. Chu *et al.*, Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.* **33**, 21-34 (2014).
62. T. Tuller, Y. Y. Waldman, M. Kupiec, E. Ruppin, Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A* **107**, 3645-3650 (2010).
63. P. Shah, Y. Ding, M. Niemczyk, G. Kudla, J. B. Plotkin, Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589-1601 (2013).
64. G. Kopperschlager, J. Bar, K. Nissler, E. Hofmann, Physicochemical parameters and subunit composition of yeast phosphofructokinase. *Eur. J. Biochem.* **81**, 317-325 (1977).
65. A. R. Gorter de Vries *et al.*, Allele-specific genome editing using CRISPR–Cas9 is associated with loss of heterozygosity in diploid yeast. *Nucleic Acids Res.* **47**, 1362-1372 (2019).
66. S. M. Richardson, S. J. Wheelan, R. M. Yarrington, J. D. Boeke, GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res* **16**, 550-556 (2006).
67. A. Riba *et al.*, Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc Natl Acad Sci U S A* **116**, 15023-15032 (2019).
68. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
69. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

Chapter 5

A yeast with muscle doesn't run faster: full humanization of the glycolytic pathway in *Saccharomyces cerevisiae*

Francine J. Boonekamp, Ewout Knibbe, Marcel A. Vieira-Lara, Melanie Wijsman, Karen van Eunen, Reinier Bron, Ana Maria Almonacid Suarez, Martin C. Harmsen, Patrick van Rijn, Karin Wolters, Barbara Bakker, Jean-Marc Daran, Pascale Daran-Lapujade

Manuscript in preparation for submission

Abstract

The yeast *Saccharomyces cerevisiae* is a popular model organism to study human cellular processes and diseases. Humanization of genes in yeast is a widely used strategy to explore gene functionality and test drugs, thereby improving yeast as metazoan model. Hindered by the high genetic redundancy of eukaryotic genomes and the lack of molecular tools for large scale genome remodelling, to date humanization studies have mostly focused on single gene complementation. Recent synthetic biology advances can overcome these challenges and make humanizing a full pathways or processes within reach. As proof of principle, the present study demonstrates the full humanization of the glycolytic pathway. Combining single gene complementation, full pathway humanisation and laboratory evolution, the functionality of 25 human enzymes in *S. cerevisiae* was explored. Out of 25 tested human genes, all but the hexokinases *Hs*HK1, *Hs*HK2 and *Hs*HK3 were able to complement the catalytic function of their yeast orthologs and aldolase and enolases also complemented their moonlighting functions outside glycolysis. Laboratory evolution suggested a remarkable variety of cellular mechanisms deployed to optimize the growth of strains with fully humanized glycolysis, such as the release of actin-bound aldolase. Comparison with skeletal muscle cells showed that, for most tested human enzymes, transplantation in yeast did not affect their turnover number (k_{cat}). Enabling to study the enzymes in a context closer to their native environment, yeast strains with fully humanized glycolytic pathways are promising models for metazoans.

Introduction

Due to its tractability and genetic accessibility, *S. cerevisiae* has played and still plays a key role as simplified model organism for higher eukaryotes. Many discoveries in yeast native processes such as the cell cycle and ribosome biogenesis were pivotal for understanding their mammalian equivalents (*1, 2*). Yeast is also a powerful model to study a wide range of diseases such as cancer, diabetes and neurodegenerative diseases (*3*). Heterologous expression of human genes in yeast enables the detailed investigation of human biology and disease-specific variations of human genes (*4*). The yeast and human genomes share over 2000 groups of orthologs (*4*), and several large initiatives have explored the complementarity of human genes in yeast, showing a high degree of functional conservation (*4-11*). These studies are however complicated by the genetic redundancy of eukaryotic genomes (*12*), which is even more prominent for genes encoding proteins with metabolic function (*13, 14*).

While individual gene complementation in yeast is an interesting approach to characterize single human proteins, the usefulness of yeast models would be greatly enhanced if endowed of entire humanized pathways or processes . Such 'next level' humanized yeast models hold the potential to capture the native functional context of the humanized proteins and to enable the study of more complex, multigene phenotypes, and epistatic interactions between genes. While the feasibility of such extensive humanization projects largely depends on the replaceability of yeast genes by their human orthologs, recent large scale humanization and bacterialization efforts of the yeast genome suggest that replaceability was better predicted on pathway- or process-basis than by sequence conservation (*9, 15*). To date, reports of full humanization of pathway or protein complexes are scarce (*8, 16-19*). However, the rapid developments in synthetic biology have tremendously expanded the molecular toolbox for extensive remodelling of microbial genomes, and promises to bring more examples of large scale humanization in the future.

The Embden-Meyerhof-Parnas (EMP) pathway of glycolysis, which is near ubiquitous in eukaryotes, plays a central role in carbon metabolism and is involved in a wide range of diseases in mammals, including cancer with the well-known Warburg effect (*20*). Mammals and *S. cerevisiae* both harbour the EMP pathway of glycolysis, however only a few single human glycolytic enzymes have been tested for complementation in yeast, mostly in large-scale studies (*6, 9, 10, 21-23*). Whether human glycolytic enzymes can complement their yeast orthologs is largely unknown. It is a particularly fascinating question as glycolytic enzymes, both in yeast and human are characterized by their versatility in moonlighting capabilities (*24, 25*). The degree of conservation of these

moonlighting functions between these two distant organisms has little been explored to date, with the exception of the human aldolase B (*Hs*ALDOB) and the glucokinase (*Hs*HK4) (*21, 22*). So far no attempt has been made to humanize more than one glycolytic step in yeast or any other eukaryote, let alone the entire glycolytic pathway.

To overcome the difficulty caused by genetic redundancy, a yeast strain in which the set of genes encoding glycolytic enzymes has been minimized from 19 to 11 was previously constructed (*26*). This minimal glycolysis (MG) strain is a perfect platform for single glycolytic gene complementation. Furthermore, a strain in which this minimized set of yeast glycolytic genes has been fully relocalized to a single chromosomal locus (SwYG strain), enables to swap the entire yeast glycolytic pathway in two transformations by any designer glycolysis (*27*). In the present study, the SwYG strain was used to fully humanize the yeast glycolytic pathway, thereby demonstrating the functionality of the entire human glycolytic pathway in yeast. Furthermore, combining single gene complementation in the MG strain, full humanization in the SwYG strain and adaptive laboratory evolution, this study offers an in-depth exploration of the functionality of human glycolytic genes in *S. cerevisiae*. Finally, the validity of strains with humanized glycolysis was evaluated by comparing the protein turnover number (k_{cat}) of the human glycolytic enzymes expressed in the humanized yeast model and in their native environment in human skeletal muscle myotube cell cultures.

Results

All human glycolytic genes complement their yeast ortholog except for *HK1 and HK3*

The human glycolytic enzymes share between 43% and 65% identity at protein level with their closest yeast ortholog, with the exception of the hexokinases and the fructose 1,6-bisphosphate (F-1,6-bP) aldolases (Fig. 1A), revealing a higher degree of conservation as compared to the whole proteome (genome-wide protein identity around 32%, (*9*)). The human and yeast F-1,6-bP aldolases belong to two different classes of enzymes and do not share homology at all at protein level (*Sc*Fba1 belongs to class II while the human *Hs*ALDOA, *Hs*ALDOB and *Hs*ALDOC belong to class I (*28*)), Fig. 1A and Table S1). Among the four human hexokinases (*Hs*HK1 to *Hs*HK4), *Hs*HK4 is closest in size and sequence to *Sc*Hxk2 (ca. 30% protein identity). *Hs*HK1, *Hs*HK2 and *Hs*HK3 underwent a duplication event and are twice as big as their yeast orthologs, however, each subunit shares ca. 30% identity with *Sc*Hxk2 (Table S1). So far, complementation in *S. cerevisiae* was only tested for eight human glycolytic genes (*6, 9, 10, 21, 23*), of which only PGAM2 was unsuccessful (Table S1, (*10*)). The genetic redundancy of metabolic pathways in eukaryotes (*13, 14*), and the inherent difficulty to perform complementation studies largely explain this knowledge gap. To overcome this problem, this study used the MG strain in which each of the ten reactions enabling the conversion of glucose to pyruvate is carried out by a single isoenzyme, with the exception of the phosphorylation of fructose-6-phosphate (fructose-6P) which requires both *Sc*Pfk1 and *Sc*Pfk2 that operate as hetero-octamer (*26*) (Fig. 1). In the presence of multiple splicing variants, the canonical version was used (Fig. 1 and Table S1). However, as the two pyruvate kinase genes *HsPKLR* and *HsPKM* have tissue-specific splicing variants (*HsPKL* and *HsPKR for HsPKLR*, and *HsPKM1* and *HsPKM2* for *HsPKM*), all four variants were tested (Fig. 1, Table S1 (*29*)).

In total, the ability of 25 human glycolytic genes to complement their yeast ortholog(s) was systematically explored by individual gene complementation in yeast. These 25 genes were codon-optimized, cloned downstream strong, constitutive promoters (Table S2) and individually cloned in the MG strain, after which the yeast ortholog(s) were removed (Fig. S1). With the exception of *Hs*PFKM and *Hs*GPI, expressed with the strong and constitutive *TEF1* and *TEF2* promoters respectively, the human enzymes were cloned behind the promoter of their yeast ortholog (Table S2).

Remarkably, 23 out of the 25 resulting strains were able to grow in chemically defined medium with 2% glucose as sole carbon source (SMG), thereby demonstrating complementation of the yeast genes by their human ortholog (Fig. 1). The human

HsHK1 and *HsHK3* gene failed to complement their yeast ortholog. While most strains were only marginally affected by single humanization of the glycolytic genes (with growth rates above 80% of the MG strain), strains harbouring a human hexokinase 2, the aldolases, phosphoglycerate mutases and the glyceraldehyde-3P dehydrogenase GAPDH variant S had a strongly reduced growth rate (strongest decrease (30%) with *Hs*ALDOB, Fig. 1). No clear correlation could be found between growth rate and conservation between human and yeast gene sequences or promoter strength (Fig. S3). Except for the aldolase A and C genes all human genes were Sanger sequenced, revealing that all besides *HsHK2* had the expected sequence. This study therefore demonstrated the absence of complementation of the native human *HsHK1* and *HsHK3* and the remarkable complementation by 22 human genes of their yeast orthologs.

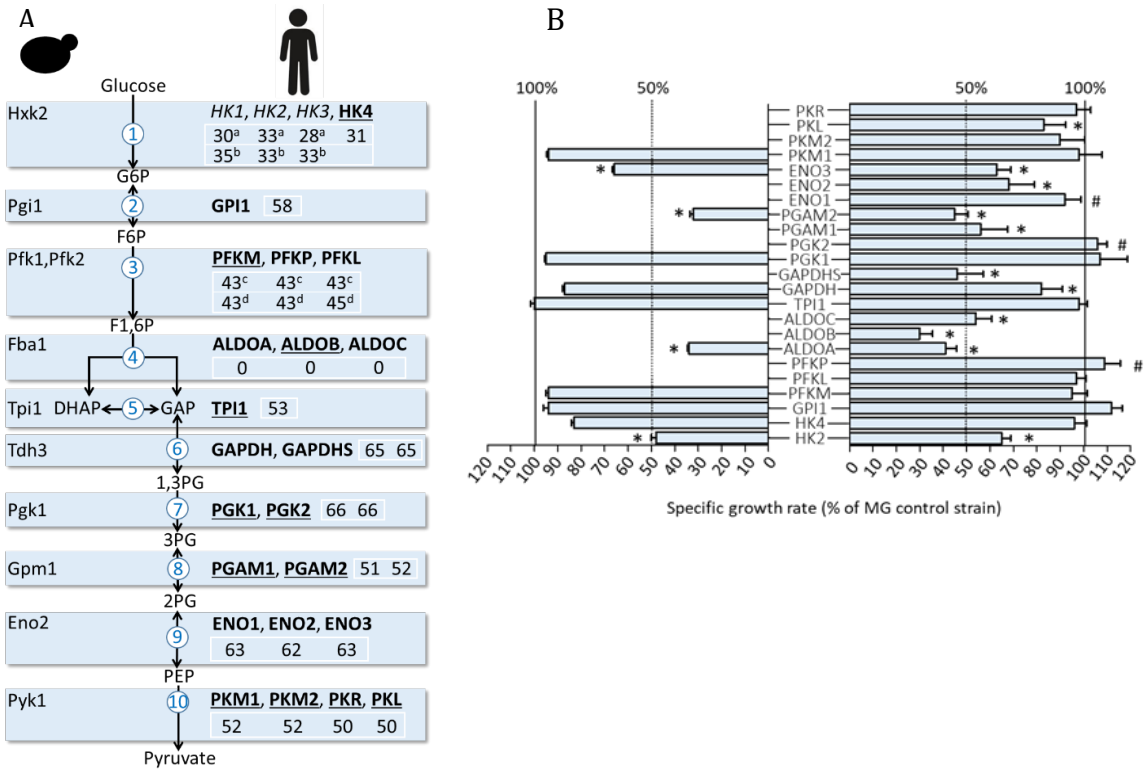


Figure 1 - Glycolytic pathway and the human and yeast enzymes relevant for this study.
A) Major yeast glycolytic enzymes that are present in the MG and SwYG strains (left side) and human enzymes used to complement their yeast orthologs in this study (right side). Human enzymes in bold indicate that they could complement their yeast counterpart in this study. Underlined human enzymes were previously shown to complement their yeast ortholog. The numbers represent the percentage identity at protein level of the human enzymes as compared to their yeast ortholog. a: % identity of the 1st subunit of human hexokinases vs *Sc*Hxk2, b: % identity of the 2nd subunit of human hexokinases vs *Sc*Hxk2, c: % identity of human phosphofructokinases vs *Sc*Pfk1 and d: % identity of human phosphofructokinases vs *Sc*Pfk2. A more complete overview of the human glycolytic enzymes is available in Table S1. **B)** Specific growth rates of the single gene complementation strains grown in synthetic medium with glucose as carbon source at 30°C. Right, measurement in 96-well plates in Growth Profiler GP), left confirmation with measurement in shake-flasks for a selection of the strains. The MG strain (IMX372) was used as control. *HsHK2* and *HsHK4* are expressed with the *PDC1* promoter. The error bars represent at least three independent replicates for GP measurements and two for shake-flasks. * indicates changes in growth rates in the mutants as compared to the control strain with p-values below 0.01 and # between 0.01 and 0.05 (Student t-test, two-tailed, homoscedastic).

The human *HsHK2* can only complement the yeast hexokinases upon mutation

The *HsHK2* gene contained a mutation leading to an amino acid substitution, suggesting that this mutation might be required for growth on glucose. Galactose utilization proceeds via galactokinase in the Leloir pathway and does not require hexokinase in *S. cerevisiae*. When solely exposed to galactose since transformation, strains expressing the human *HsHK2* as sole hexokinase grew well on this hexose and maintained *HsHK2* native sequence. However, when the same strains were exposed to glucose, they displayed a lag phase of 4 to 5 days before resuming growth. Sequencing of the *HsHK2* gene in several strains after growth completion on glucose revealed the systematic occurrence of a single mutation leading to an amino acid substitution or loss (Fig. 2A). Heterologous expression of the human *HsHK2* in yeast might cause misfolding or lack of a post-translational modification leading to a catalytically inactive *Hs*HK2. However, *in vitro* enzyme assays performed with a complementation strain solely exposed to galactose and unable to grow on glucose (IMX2419) revealed that the native HK2 enzyme was active (Fig. 2). Additionally, the native *Hs*HK2 displayed the same specific activity as the mutated variant found in IM1690, a strain able to grow on glucose (Fig. 2). As the catalytic activity of native and mutated *HsHK2* alleles was conserved, the inability to grow on glucose for strains harbouring the native *HsHK2* allele might result from inhibition of the native enzyme *in vivo*. The yeast hexokinase 2 is strongly inhibited by trehalose-6P (*30*). Accordingly, 1 mM trehalose-6P resulted in 90% inhibition of the yeast *Sc*Hxk2 in *in vitro* assays, while it only marginally affected the human *Hs*HK2 (Fig. 2). As concentrations of Trehalose-6P above 1 mM are not often encountered in yeast cells (*30-32*), trehalose-6P inhibition was probably not responsible for the lack of activity of the native *Hs*HK2 in yeast cells. This theory was also supported by the similar tolerance to trehalose-6P of the native and mutated *HsHK2* alleles (Fig. 2C, inhibition of *Hs*HK2 and *Hs*HK2* not significantly different, t-test, two-tailed, homoscedastic, $p>0.05$). *Hs*HK2 activity *in vivo* responds of course to its substrates concentration (K_m for glucose ca. 0.2 mM and K_m for ATP ca. 1 mM (*33, 34*)), but it is also sensitive to allosteric inhibition by its product glucose-6P (*33, 35*). The catalytic and glucose-6P binding sites of *Hs*HK2 are spatially close ((*34*), Fig. 2D). The sequence of three independent *Hs*HK2 complementation strains revealed that the mutations occurred in the vicinity of these two binding sites, suggesting that they might alter *Hs*HK2 affinity for its substrates or its sensitivity to glucose-6P. The specific activity of the native and mutated *Hs*HK2 enzymes was identical with 1, 10 and 50 mM glucose, suggesting that the affinity for glucose was unaltered by the mutation (Fig. 2B). For testing the role of glucose-6P, purification and fine biochemical characterization of native and mutated *Hs*HK2 will be required. In addition, a glycolytic kinetic model

developed for yeast (*36*) could be modified to mimic the sensitivity of the glucose phosphorylation to glucose-6P instead of trehalose-6P and to take into account the low specific activity of the human *Hs*HK2.

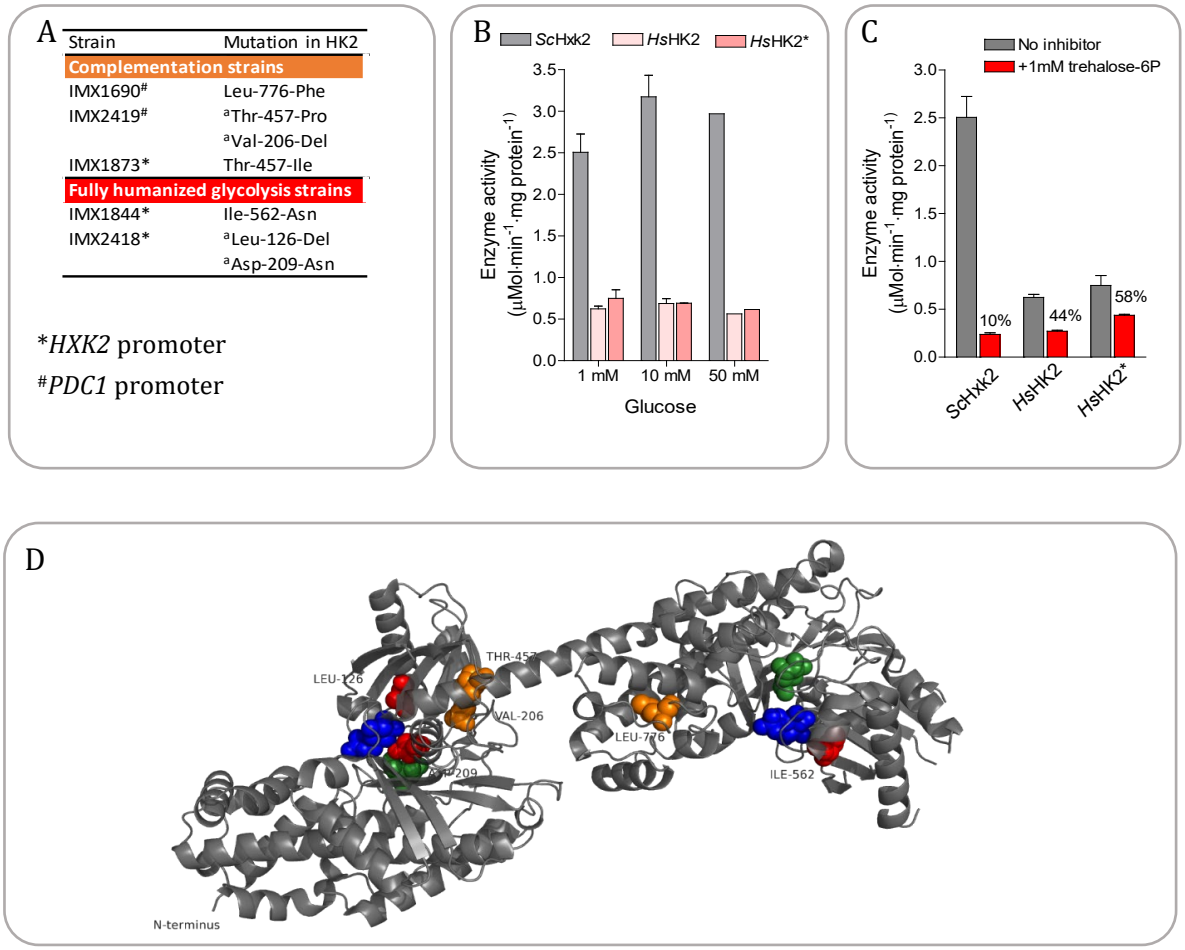


Figure 2 – Characterization of *HK2* mutants.

A) list of amino acid substitutions found in *Hs*HK2 in fully humanized and complementation strains grown on glucose. The red and orange colours correspond to the colour of the mutations in panel D. **B)** and **C)** *In vitro* assay of hexokinase activity from *S. cerevisiae* cell extracts grown on SM-galactose. Control *Sc*Hxk2 activity was measured from IMX215, native *Hs*HXK2 activity was assayed with cell extracts of IMX2419 never exposed to glucose. *Hs*HK2* carries the Leu-776-Phe substitution and was assayed with cell extracts from IMX1690. **B)** assaying with various glucose concentrations. **C)** assaying with 1 mM glucose, testing the inhibitory effect of 1 mM trehalose-6P. In these strains *HsHK2*, *HsHK2** and *ScHXK2* are expressed with the *ScPDC1* promoter. **D)** localization of the amino acid substitutions found in *Hs*HK2 (identity and colour coding in panel A) according to the crystal structure resolved in Nawaz et al. 2018 (*34*). Green represents the glucose binding site in the catalytic domain, blue indicates the allosteric binding site for glucose-6P.

Successful humanization of the entire glycolytic pathway in yeast

The complementation study revealed that individual human glycolytic enzymes could complement the glycolytic function of their yeast orthologs, meaning that they were active in yeast and functional in a yeast glycolytic context. To what extent the simultaneous complementation of all yeast glycolytic enzymes by their human orthologs can be predicted by individual complementation is however unknown. Several human enzymes did not restore yeast growth rate, and their simultaneous expression might synergistically impair growth. This effect could result from suboptimal catalytic activity, but might also be caused by cumulative moonlighting function of the human enzymes. Furthermore, the operation of all human enzymes in the yeast metabolic network might lead to a metabolic imbalance (*36, 37*), more particularly as two key regulatory steps were altered in the humanized pathway, with human hexokinases insensitive to inhibition by trehalose-6P ((*30*) and Fig. 2C), *Hs*HK2 inhibited by glucose-6P and *Hs*PKM1 insensitive to the feed-forward activation by fructose-1,6-bisP (Fig. S4, (*38*)).

To increase the likelihood of a functional humanized glycolytic pathway, we selected the set of enzymes highly expressed in a single tissue, and selected skeletal muscle cells for their high glycolytic flux (*39-41*). *Hs*HK2 is in human skeletal muscle cells together with *Hs*HK1 the most abundant hexokinase (*42*), however complementation experiments revealed the inability of *Hs*HK1 and of *Hs*HK2 in its native form to complement its yeast ortholog (Fig. 2). This made *Hs*HK2 particularly interesting, to test whether a fully humanized glycolytic context would impose the same selective pressure and result in the same mutation in *HsHK2* as in complementation strains with a yeast glycolytic context. As *HsHK4* was highly functional in *S. cerevisiae* (Fig. 1), a second humanized glycolysis was constructed with *Hs*HK4 as sole glucose phosphorylation enzyme. *Hs*HK4, also known as glucokinase, strongly differs from the three other hexokinases in sequence and kinetic and regulatory properties, with a substantially lower affinity for glucose and an insensitivity to glucose-6P (*43, 44*). *HsHK4* is the dominant hexokinase isoform in liver cells and in pancreatic cells in which it plays a key role in glucose sensing (*43, 44*). Pathway transplantation was performed in a SwYG background in two transformation steps, resulting in two strains with a fully humanized glycolysis, *Hs*Gly-HK2 (strain IMX1844) with *Hs*HK2 as hexokinase and *Hs*Gly-HK4 (strain IMX1814) with *Hs*HK4 (Fig. 3). In parallel, *Sc*Gly (IMX1821), a control strain expressing the native but minimized and relocalized glycolytic pathway with the native promoters and terminators and same integration site, was also constructed (Fig. 3). It is noteworthy that in these fully humanized strains, *HsHK2* and *HsHK4* were expressed with the yeast *ScHXK2* promoter. This promoter led to a slower growth rate in

complementation strains than shown above with the *ScPDC1* promoter (Fig. S7). All other human genes were expressed with the same promoters as those used for the complementation study (Table S2).

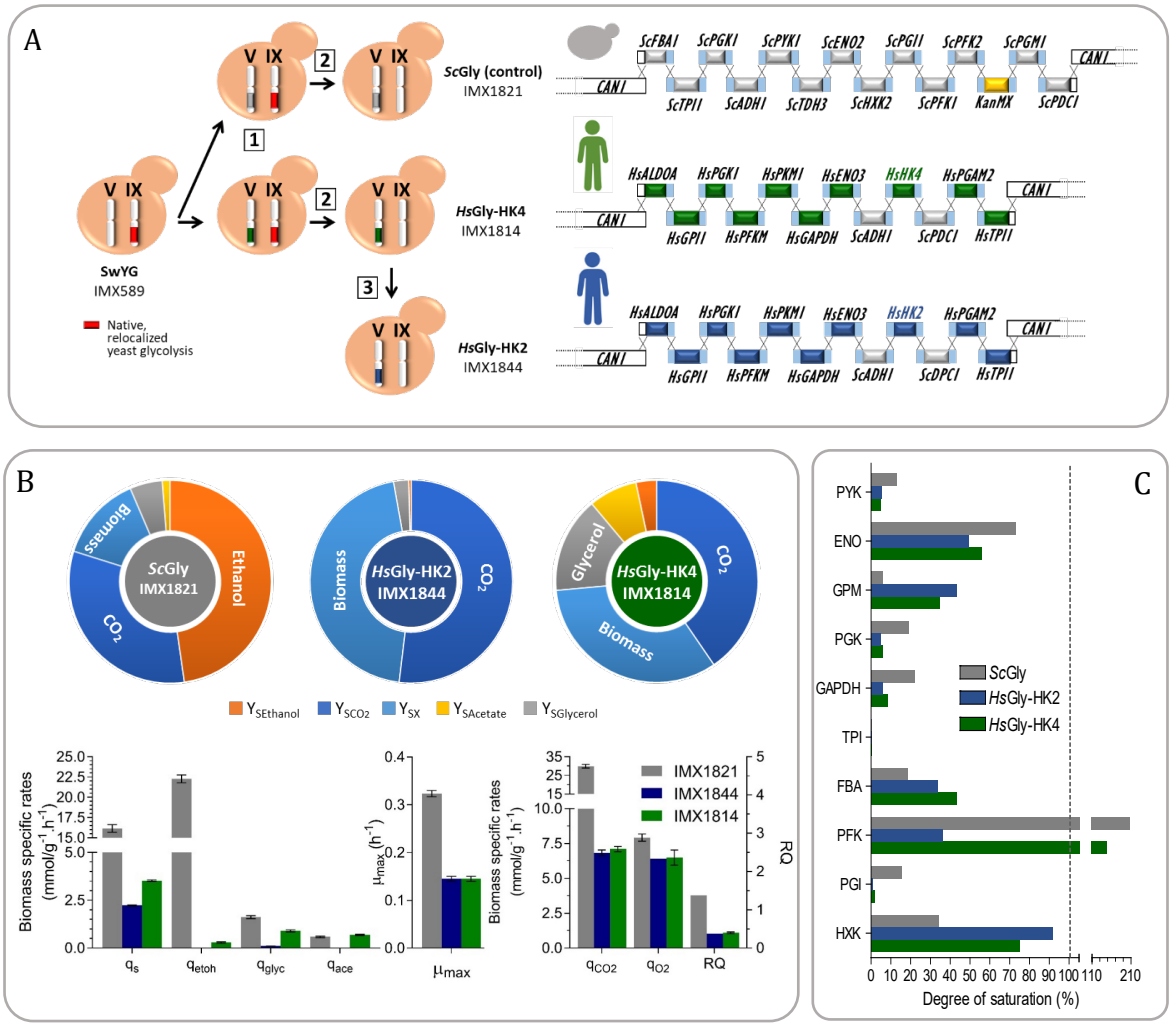


Figure 3 – Construction and physiological characterization of strains with fully humanized glycolysis. A) Overview of the strain construction strategy and of the composition of the glycolytic pathway in the strains with native, co-localized glycolysis (IMX1821) and fully humanized glycolysis (IMX1844 and IMX1814). **B)** Physiological characterization of strains with fully humanized glycolysis IMX1844 (HsGly-HK2), IMX1814 (HsGly-HK4) and their control strain with yeast glycolysis IMX1821. On the top are indicated the yields on glucose (CMol/CMol) $Y_{SEthanol}$, ethanol yield, Y_{SCo2}, CO_2 yield, Y_{SX}, biomass yield, $Y_{SAcetate}$, acetate yield and $Y_{SGlycerol}$, glycerol yield. On the bottom are shown the specific rate for glucose and oxygen uptake (q_{glu} and q_{O2}), and for ethanol (q_{Eth}), glycerol (q_{gly}), acetate (q_{acet}) and CO_2 (q_{CO2}) production. μ_{max}, specific growth rate, RQ, respiratory quotient (q_{CO2}/q_{O2}). **C)** Estimation of the degree of saturation of glycolytic enzymes based on *in vitro* assays from cell extracts of the control strain with yeast glycolysis (*Sc*Gly, IMX1821) and fully humanized glycolysis (*Hs*Gly-HK2, IMX1844 and *Hs*Gly-HK4, IMX1814). The *in vivo* local fluxes were approximated from the q_{glu}. The horizontal dashed line indicates 100% saturation.

Both *Hs*Gly-HK2 and *Hs*Gly-HK4 were able to grow on glucose as sole carbon source in minimal, chemically defined medium, with a specific growth rate of ca. 0.15 h^{-1} (Fig. 3). As observed with the complementation strains, *HsHK2* was systematically mutated in the fully humanized yeast strains. Two transformants were sequenced and revealed a single mutation in the vicinity of the catalytic and glucose-6P binding sites (Fig. 2, Table S4). The human glycolytic context was therefore not able to offer conditions enabling the maintenance of the native *HsHK2*. *Hs*Gly-HK2 mostly respired glucose, with traces of ethanol and glycerol, while *Hs*Gly-HK4 displayed a respiro-fermentative metabolism, more similar to that of the control strain (Fig. 3 and Table S3). Similar physiological responses were observed in their respective *Hs*HK2 and *Hs*HK4 complementation strains, revealing that the human hexokinases differentially affected yeast metabolism (Fig. 4A). *S. cerevisiae* favours a mixed respiro-fermentative metabolism when glucose is present in excess (see IMX1821 in Fig. 3), a phenomenon known as the Crabtree effect, reminiscent of the Warburg effect (*45*).

*Sc*Hxk2 is involved in this response by partially localizing to the nucleus in the presence of excess glucose where it represses the expression of genes involved in respiration and the utilization of alternative carbon sources such as *SUC2* encoding invertase that hydrolyses sucrose *(46, 47).* Accordingly, in the presence of excess glucose, *SUC2* was repressed and invertase activity was not detected (Fig. 4B). Conversely, when *S. cerevisiae* is grown on ethanol supplemented with a low glucose concentration, *SUC2* expression is induced and the invertase activity becomes detectable, as observed in the present study (Fig. 4B). In a yeast mutant carrying a double *hxk1hxk2* deletion, alleviation of glucose repression enables expression of *SUC2* and invertase activity in the presence of excess glucose (Fig. 4B). The activity of invertase and the physiological data of the *Hs*HK2 and *Hs*HK4 complementation strains (Fig. 4A,B) suggested that *Hs*HK4 complemented *Sc*Hxk2 function in glucose repression, in agreement with an earlier report (*21*) while *Hs*HK2* did not. For both fully humanized and complementation strains, while the correlation between the ethanol production and glucose consumption rates was in line with previous reports (Fig. 3 and Fig. 4A) (*48*)), ethanol production was not expected at these slow growth rates (*49, 50*). Overall, the humanized yeast strains grew remarkably well.

While the physiological function of the allosteric regulations of *Sc*Hxk2, *Sc*Pfk1-*Sc*Pfk2 and *Sc*Pyk1 have not been fully elucidated, they appear to play a role in the cellular adaption to transitions, and more particularly sugar transitions (*36, 51*). Metabolic imbalances can be detected by monitoring population heterogeneity (i.e. differences among cells in ability to grow) during switches between glucose and alternative carbon source (*52*). No population heterogeneity was observed during galactose to glucose

switches for both humanized yeast strains (Fig. S5), revealing that the humanized glycolytic pathways were fully capable of fine tuning the carbon flux in response to sudden changes from alternative to favourite and repressive carbon source.

Complementation of moonlighting functions

The specific growth rate of fully humanized yeast strains was clearly suboptimal as compared to yeast strains with a native glycolysis (ca. 60% slower, Fig. 3). While several factors could explain this response, an important aspect to consider is the involvement of many eukaryotic glycolytic enzymes, next to their glycolytic functions, in other cellular activities. These moonlighting functions might not be conserved across species (*25*) and failure of the human orthologs to complement the yeast moonlighting activities might have a strong impact on the physiology of the humanized yeast strains. The function of yeast and human hexokinases in glucose catabolite repression was addressed above, but two other yeast glycolytic enzymes are involved in non-glycolytic functions. Remarkably, the important role of aldolase in assembly of vacuolar proton-translocating ATPases (V-ATPases) is highly conserved between the yeast Fba1 and the human *Hs*ALDOB (*22, 53*). However, whether the other human aldolases (*Hs*ALDOA and *Hs*ALDOC) can complement the moonlighting functions of *Sc*Fba1 is yet unknown. Aldolase-deficient strains cannot grow at alkaline pH (*53*). As previously demonstrated, *Hs*ALDOB supports growth at pH 7.5 and fully complements the yeast aldolase, including its vacuolar function (Fig. 4C, (*22*)). The present work demonstrates that the human aldolases A and C, sharing ca. 70% identity with *Hs*ALDOB, also fully complement the vacuolar function of the yeast aldolase. (Fig. 4C).

Furthermore, the yeast enolases *Sc*Eno1 and *Sc*Eno2, and more particularly *Sc*Eno2, are involved in vacuolar fusion and protein transport to the vacuole (*54*). Whether the human enolases can take over this function in yeast is yet unknown. While enolase-deficient mutants display a fragmented vacuole phenotype and growth defects (*54*), this phenotype was not observed for the MG strain expressing *Sc*Eno2 only ((*26*), (Fig. 4D) and was also not observed for complementation strains expressing either of the three human enolases (Fig. 4D). This vacuolar moonlighting function is therefore conserved between yeast and human enolases when expressed in yeast.

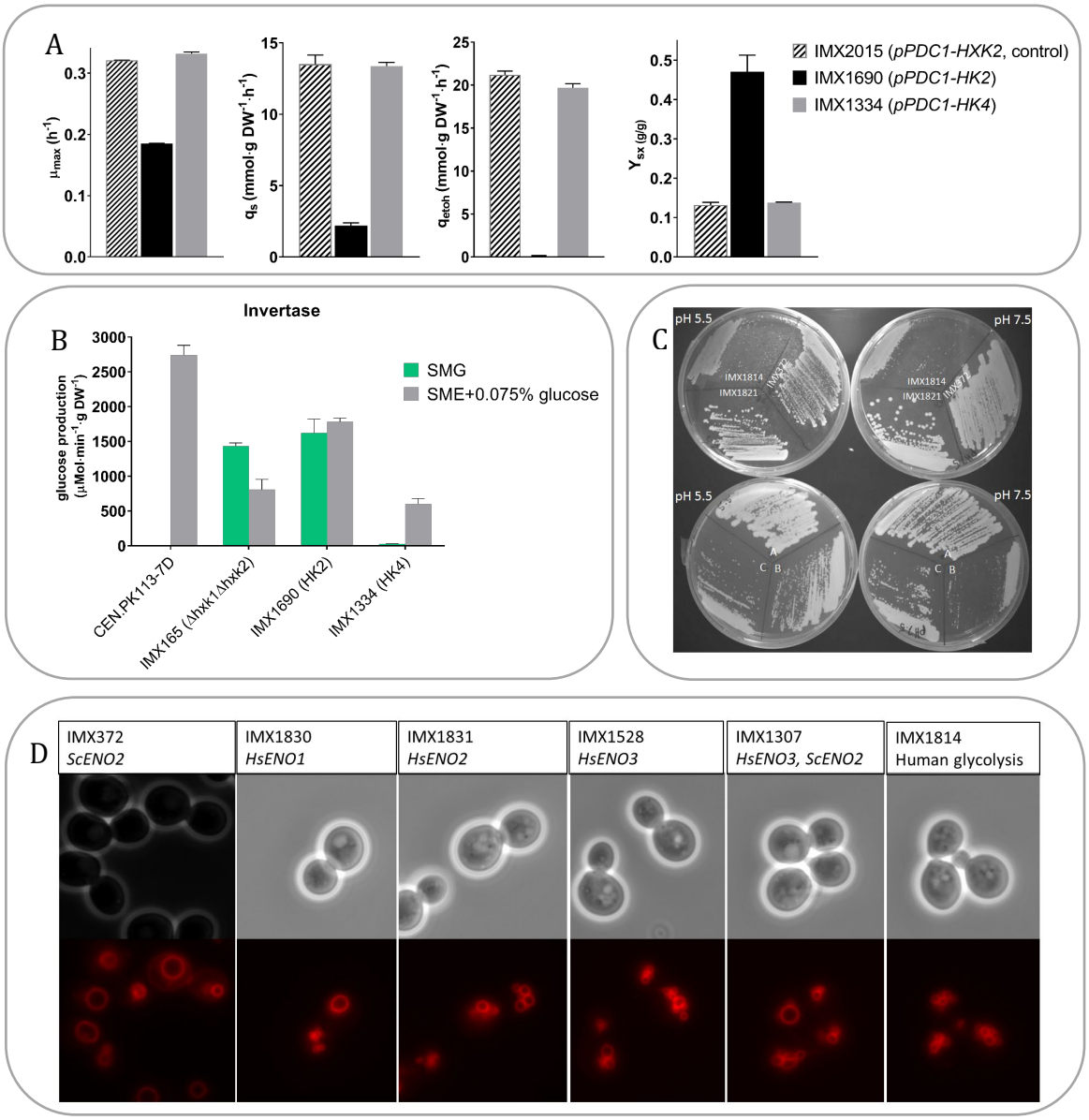


Figure 4 – Complementation of moonlighting functions.

A) Growth rate, specific glucose consumption rate and ethanol production rate and biomass yield of individual hexokinase complementation strains grown in shake flask in SMG. Error bars represent the SEM of two biological replicates. **B)** Extracellular invertase activity of cells incubated for 2h in SMG (repressing condition) or SM ethanol + 0.075% glucose (inducing condition). Invertase activity is presented as µM glucose production per minute per gram biomass dry weight. Error bars represent the SEM of two biological replicates. CEN.PK113-7D and IMX165, a *hxk1* and *hxk2* deletion mutant were used as control. **C)** Growth on YPD at pH 5.5 and pH 7.5 of two strains with yeast *Sc*Fba1 (IMX372 MG control, IMX1821 SwYG control), of fully humanized strain with *Hs*ALDOA (IMX1814, *Hs*Gly-HK4) and of three complementation strains with human *Hs*ALDOA, *Hs*ALDOB and *Hs*ALDOC (indicated by A, B and C, IMX1720, IMX2116 and IMX2018, respectively). **D)** Staining of vacuoles with FM4-64 in *S. cerevisiae* control strains with *Sc*Eno2 (in MG background, IMX372) and humanized strains with *Hs*Eno1, *Hs*Eno2 and *Hs*Eno3 single complementations (IMX1830, IMX1831, IMX1528) and with fully humanized glycolysis with *Hs*Eno3 (IMX1814, *Hs*Gly-HK4). IMX1307 carries one copy of the human *HsENO3* and the yeast *ScENO2* genes.

Engineering and evolutionary approaches to increase the slow growth phenotype of humanized glycolysis strains

The lack of complementation of moonlighting functions did not appear to be the major determinant of the slow growth phenotype of the humanized yeast strains. This slow growth could be explained by many other factors such as a low capacity of the human enzymes (resulting from low expression, improper folding or inappropriate post-translational modifications), (allosteric) inhibition of the enzyme activity *in vivo*, affinities for substrate and co-factors not adapted to the yeast cellular environment, or deleterious moonlighting activities of the human orthologs. *In vitro* assays confirmed that, with the exception of *Hs*GPI1 the capacity (V_{max}) of the human enzymes was two to fifty times lower than the enzyme capacity of their yeast ortholog (Fig. S6). With the notable exception of phosphofructokinase, sensitive *in vivo* to many effectors, the yeast glycolytic enzymes operate at overcapacity ((*55-57*) and Fig. 3). In the humanized yeast strains, with the exception of HXK, FBA and GPM, the degree of saturation of most glycolytic enzymes was lower as compared to the control strain with the native yeast glycolysis, suggesting that the activity of these enzymes did not limit the glycolytic flux (Fig. 3). This hypothesis is supported by the similarity in growth rate between most complementation strains and the *S. cerevisiae* control with a full set of native glycolytic genes (Fig. 1). Conversely, the degree of saturation of HXK, FBA and GPM was significantly higher in the humanized strains (two to six-fold higher in *Hs*Gly-HK2 and *Hs*Gly-HK4 as compared to *Sc*Gly), suggesting that their activity might be limiting the glycolytic flux. Accordingly, the replacement of the yeast enzymes by these orthologs resulted in a decrease in growth rate in complementation strains (Fig. 1 and Fig. S7). Furthermore, simultaneous overexpression of *HsHK2*, *HsALDOA* and *HsPGAM2* in *Hs*Gly-HK2, as well as well as overexpression of *HsHK4*, *HsALDOA* and *HsPGAM2* in *Hs*Gly-HK4 successfully increased the specific growth rate by 63% and 48% respectively (Fig. 5A). These optimized, humanized yeast strains still grew 30% to 40% lower than the control strain with native, minimized yeast glycolysis (Fig. 5A). It is noteworthy that the enzyme capacity of *Hs*PFKM, while expressed with the same expression cassette and harbouring no mutations between *Hs*Gly-HK2 and *Hs*Gly-HK4, was 2.6-fold lower in *Hs*Gly-HK4 than in *Hs*Gly-HK2 (Student t-test, p-value 0.02, Fig. S6). Consequently, while *Hs*PFKM operated above capacity in *Hs*Gly-HK4, similarly to what is typically observed in *S. cerevisiae* and in IMX1821, the flux through *Hs*PFKM in *Hs*Gly-HK2 was only at ca. 30% of its *in vitro* capacity (Fig. 3).

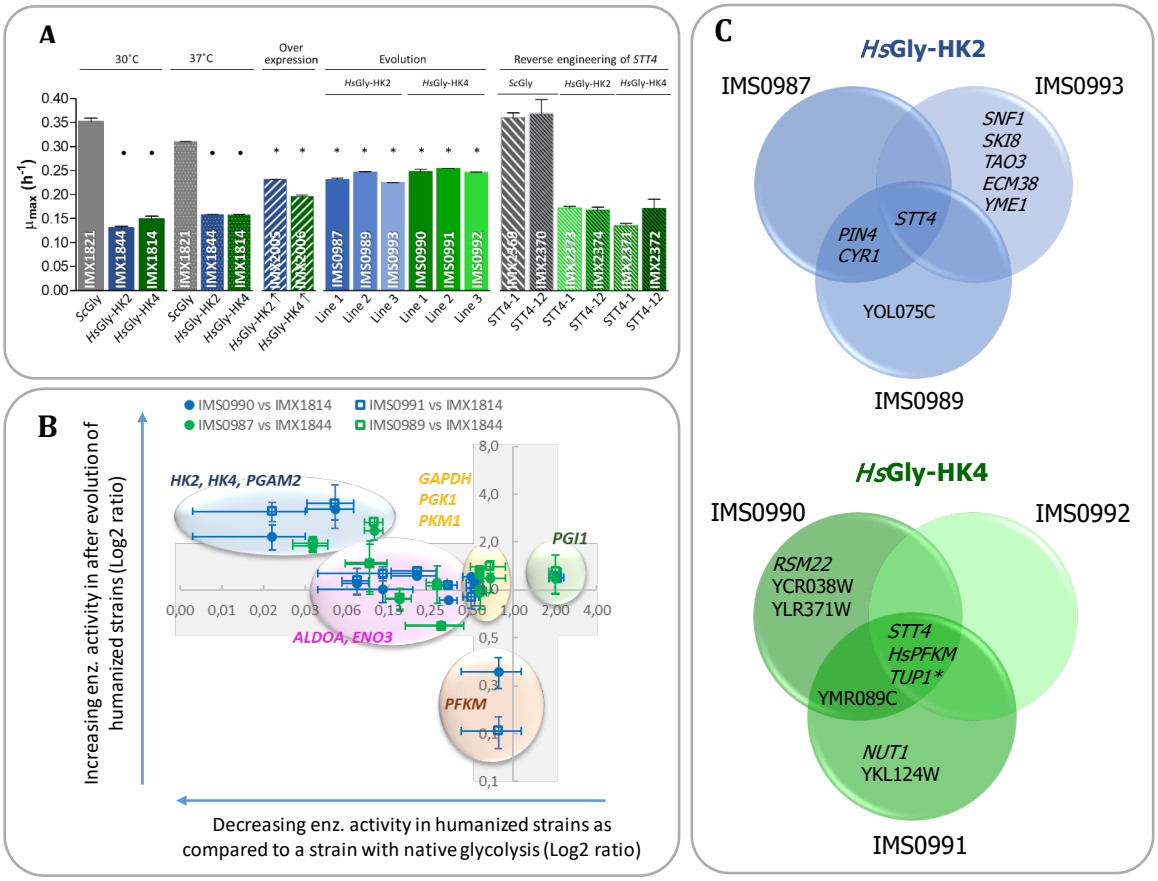


Figure 5 – Strategies to improve the growth rate of fully humanized strains.
A) Specific growth rate measured in aerobic shake-flasks on SMG. Asterisks indicate growth rates of the evolved strains that are significantly different (Student t-test, two-tailed, homoscedastic, p-value<0.05) from their parental strain (*Hs*Gly-HK2 or *Hs*Gly-HK4). Dots indicate when the growth rate of *Hs*Gly-HK2 or *Hs*Gly-HK4 is significantly different from the growth rate of the control strain IMX1821 (Student t-test, two-tailed, homoscedastic, p-value<0.05). **B)** Comparison between the change in growth rate caused by single complementation of the glycolytic enzymes and the change in activity of the glycolytic enzymes during evolution. Blue dots and square represent strains IMS0987 and IMS0989 evolved from *Hs*Gly-HK2 and the green dots represent IMS0990 and IMS0991 evolved from *Hs*Gly-HK4. **C)** Mutations found in the genomes of humanized strains after evolution for 630 generations. The strains are single colony isolates from the different evolution lines. The asterisk indicates when the mutation is exactly the same between evolved strains.

As human enzymes likely have evolved to operate optimally at body temperature (*58*), humanized yeast strains might perform better at 37°C, a temperature permissive for *S. cerevisiae*. However no growth improvement was observed by increasing the growth temperature for the humanized yeast strains (Fig. 5A). Many mechanisms could explain the slow growth phenotype of the humanized strains, more that can be tested by rational, design-build-test-learn approaches. Adaptive laboratory evolution (ALE) has been successfully used for a broad range of research areas from exploring fundamental evolutionary theories to improving the biotechnological properties of industrial microbes (*59*). ALE is a particularly powerful tool to elucidate complex phenotypes and was used to improve the fitness of the humanized strains in an attempt to identify the molecular basis of their slow growth phenotype. After approximately 630 generations in glucose medium, evolved populations of humanized yeast grew ca. two-fold faster than their *Hs*Gly-HK2 and *Hs*Gly-HK4 ancestors (Fig. S8). Single colony isolates from six independent evolution lines, three per humanized yeast strain, confirmed the increased growth rate of the evolved humanized yeast strains (strains IMS0987 to IMS0993, Fig. 5, Table S8D). With faster growth, these strains evolved towards a more fermentative metabolism, although the relatively high biomass yield and low ethanol yield as compared to the control strain with native yeast glycolysis indicated that the evolved strains were still largely respiring (Fig. 5 and Fig. S9).

The enzyme capacity (V_{max}) of several human glycolytic enzymes was affected by evolution, and overall, changes in enzyme capacity were correlated to the impact on specific growth rate of the complementation of the yeast enzymes by their human orthologs (Fig. 1, Fig. 5 and Fig. S10). Across the six evolution lines, the capacity of both hexokinases (*Hs*HK2 and *Hs*HK4) and *Hs*PGAM2, which complementation leads to the strongest impact on growth rate, was increased two to three-fold during evolution. Remarkably, *Hs*ALDOA, which also led to a strong decrease in growth rate upon complementation, was not affected by evolution. *Hs*PFKM displayed a particularly interesting response (Fig. 5 and Fig. S10). While its capacity (V_{max}) was already lower in the humanized yeast strains than in the *Hs*Gly strain, it is the only activity that was substantially decreased during evolution. In the evolved humanized strains *Hs*PFKM activity was 20 to 50 times lower than in the *Sc*Gly strain IMX1821.

Exploring the causes of the slow growth phenotype of humanized glycolysis strains

The genome sequence of the evolved strains offered little clarification regarding the mechanisms leading to the lower *in vitro* specific activities of the glycolytic enzymes. With the exception of *Hs*PFKM, which carried mutations in its coding region in all three evolution lines of *Hs*Gly-HK4 (Fig. 5), neither the promoter, coding or terminator

regions of the human glycolytic genes displayed mutations. The transcription factors involved in the regulation of the activity of the yeast glycolytic promoters (Rap1, Abf1, Gcr1, Gcr2) did not harbour mutations either. Overall, few mutations were conserved between the evolution lines of the two humanized strains, and a single mutation, in *STT4*, was conserved for all six evolution lines (Fig 5). Remarkably the six identified mutations were located within 164 amino acids, in the C-terminus of the protein harbouring its catalytic domain (Table S5). *STT4* encodes a phosphatidylinositol-4P (PI4P) kinase that catalyses the phosphorylation of PI4P into PI4,5P_2. As Stt4 is essential (*60*), the mutations present in the evolved strains could not cause a loss of function. Phosphoinositides are important signalling molecules in eukaryotes, involved in vacuole morphology and cytoskeleton organisation via actin remodelling (*61*). Reverse engineering of two of the mutations found in the evolved strains IMS0990 and IMS0992 was performed in non-evolved strain backgrounds with native yeast glycolysis and humanized glycolysis, by mutating the native *STT4* gene. No change in specific growth rate was observed in these reverse engineering strains irrespective of the strain background (Fig. 5), however in these strains the mutations resulted in a fragmented vacuole phenotype (Fig. S11), confirming that the mutations interfered with Stt4 activity and PI4P signalling. Such a phenotype was not observed in the evolved strains, however in these strains vacuoles also displayed abnormal morphologies with collapsed structures, indicating that specific mechanisms might have evolved in parallel to mitigate the effect of *STT4* mutations on vacuolar morphology (Fig. S11).

Evolution of the two humanized yeast strains, differing only in the hexokinase step, resulted in different genotypes. All three evolved strains from the *Hs*Gly-HK4 variant harboured a mutation in *Hs*PFKM (Fig. 5, Table S5), one in the N-terminal catalytic domain of the protein and the two others in the C-term regulatory domain where several allosteric effectors can bind (F2,6bisP, ATP, ADP, citrate, etc.). The impact of these mutations cannot be inferred from the location of the mutation, but it might be involved in the strong decrease in *in vitro* activity of PFKM measured in the strains evolved from *Hs*Gly-HK4. All three evolved strains from the *Hs*Gly-HK2 strain were mutated in *TUP1* with the exact same non-synonymous mutation resulting in the replacement of histidine in position 489 by a tyrosine. The Tup1 protein, which is conserved in eukaryotes, forms together with Cyc8 a repressor complex which is involved in the regulation of at least 150 genes with various functions in *S. cerevisiae* (*62, 63*). Deletion of *TUP1* is associated with several phenotypes, including loss of some aspects of glucose repression (*63*). This can be explained by the observations that the Cyc8-Tup1 complex interacts with the repressor Mig1 which is involved in the glucose repression mechanism (*64*). The mutations in *TUP1* occurred in the *Hs*Gly-HK2 strain

background, containing the *HsHK2* gene which could not complement the *ScHXK2* glucose repression moonlighting function (Fig. 4). The mutation might result in a (partial) restoration of glucose repression.

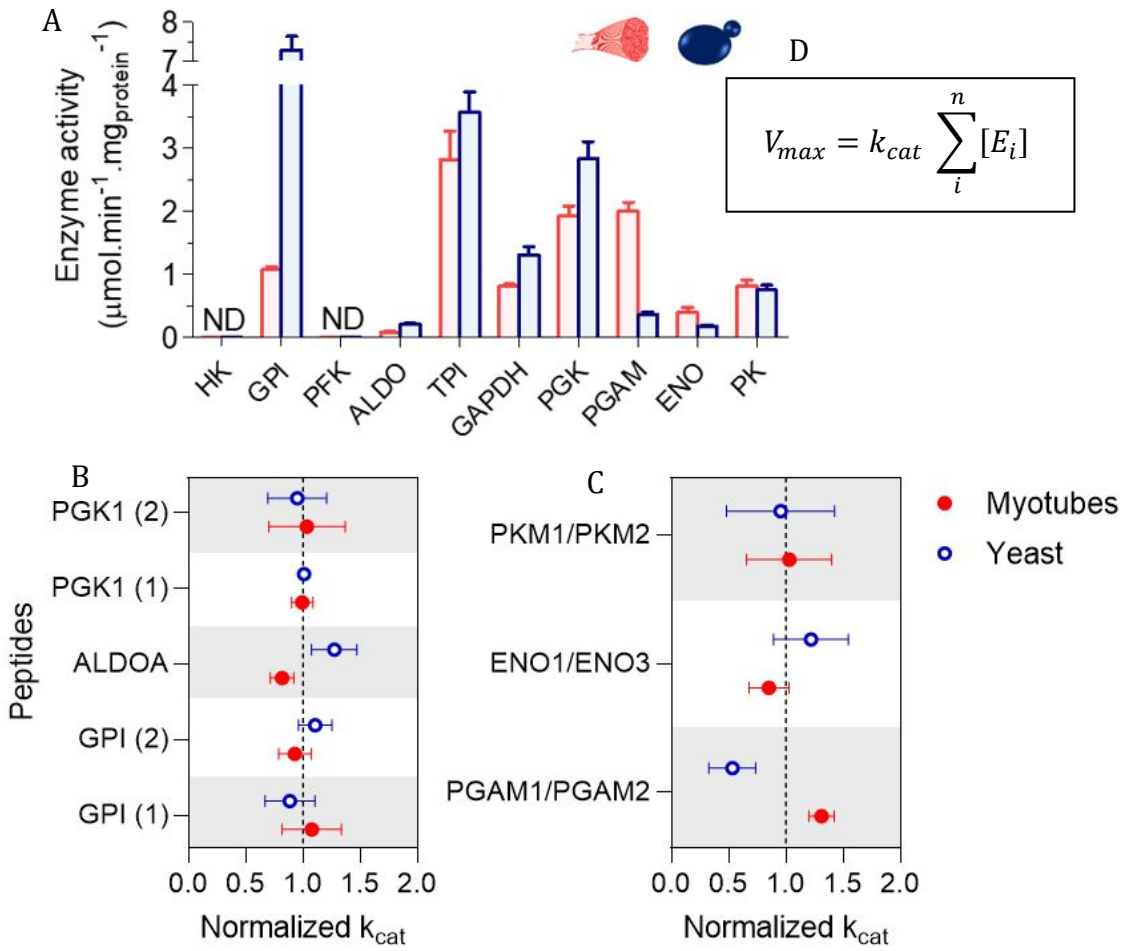


Figure 6 – Enzyme activity and k_{cat} of human glycolytic enzymes in human myotubes and humanized yeast.

A) Specific activity of human glycolytic enzymes measured *in vitro* with *in-vivo* like assaying conditions. Cell-free extracts came from the yeast strain *Hs*Gly-HK2 (blue) and from muscle myotube (red) cultures. **B)** k_{cat} values were calculated as the maximal enzyme activity divided by the concentration of each individual peptide detected by proteomics when no other isoforms were detected. **C)** k_{cat} values calculated as the maximal enzyme activity divided by the sum of the concentration of all isoforms that catalyse the specific reaction. Peptide data used for calculations can be found in supplemental figure S13-A. When more than one peptide was detected per protein, protein concentration was determined as the average of the peptide levels. For both cases, k_{cat} values were normalized by the average of the values from yeast and myotubes. **D)** Equations used to estimate k_{cat}, $[E_i]$ represents the concentration of each isoform, number of subunits per enzyme complex not included. Data represent mean values and error bars show the standard deviation of three and two independent culture replicates for the myotubes and yeast cultures, respectively. ND: not detected.

Relevance of yeast as model for human glycolysis

Humanized yeast strains can serve as a powerful models to study human diseases (*4*). To get a better insight in the similarities and differences between glycolysis in its native, human environment and in humanized yeast, we quantified the catalytic turnover rates (k_{cat}) of the human glycolytic proteins in yeast and in human myotube cultures. If the intracellular milieu of the yeast cells would affect posttranslational modifications that affect the enzyme activity, it should be reflected in their k_{cat} values. First, catalytic capacities of all enzymes (V_{max}) were measured in cell extracts using *in vivo*-like assay conditions mimicking the intracellular environment of mammalian cells in term of pH, temperature, and ion concentrations. In these assay conditions phosphofructokinase and hexokinase activities were too low for detection, although both proteins were present in both yeast and myotube samples (Fig S13A). Overall the V_{max} values of glycolytic enzymes were of the same order of magnitude in humanized yeast and muscle cells (Fig. 6A). *Hs*GPI and *Hs*PGK1 activity was higher in yeast cells than in muscle cells, particularly for *Hs*GPI (seven-fold) while the activity of PGAM was 5.5-fold lower in yeast than in muscle cells (Fig. 6A). To estimate the k_{cat} values, protein concentrations were quantified by mass spectrometry based on ^{13}C labelled peptide standards. The differences in activity were mirrored in the peptide abundance for these proteins (Fig. S13A), suggesting that the k_{cat} values of the human proteins expressed in human and yeast cells were not substantially different. For *Hs*GPI, *Hs*ALDOA and *Hs*PGK1, the k_{cat} values were calculated by dividing the V_{max} values by the respective protein concentrations (Fig 6D). This revealed no differences in the turnover rate between yeast and myotubes, irrespective of which of the standard peptides was used for protein quantification (Fig. 6B). For the remaining enzymes, calculation of the turnover rate was complicated by the presence of isoenzymes other than the canonical muscle glycolytic enzymes in the myotube cultures. The isoenzymes *Hs*PFKL, *Hs*PFKP, *Hs*PGAM1, *Hs*ENO1 and *Hs*PKM2 were present at concentrations equivalent to that of their canonical muscle isoenzymes or higher (Fig. S13A). This has been reported before for *in vitro* muscle cultures and, to a lower extent, for muscle biopsies (*65*). Therefore, we assumed that the apparent k_{cat} in these cases would be the V_{max} divided by the sum of all detected isoforms catalysing the specific reaction (Fig. 6D). The apparent k_{cat} values for pyruvate kinase and enolase were similar between humanized yeast and myotubes while the k_{cat} of PGAM was significantly lower in humanized yeast (40% of the value for myotubes) (Fig. 6C). This may suggest that the yeast milieu has a negative impact on posttranslational processing of the enzyme. To complement the analysis, we examined the available literature on the k_{cat} of glycolytic isozymes. Although no differences were found between the k_{cat} values of *Hs*ENO1 and *Hs*ENO3 (*66*), we found

evidence of a higher k_{cat} of the isoforms *Hs*PGAM2 and *Hs*PKM1 versus their respective isozymes (*67, 68*). Taking this into consideration, expression in yeast had no major effect on the apparent k_{cat} values of PKM and PGAM, with the k_{cat} for *Hs*PGAM2 remaining significantly lower in yeast (Fig. S13B). Altogether, these results demonstrate that out of the seven enzymes for which a turnover rate could be estimated, six were not catalytically altered by the yeast environment, with *Hs*PGAM2 as the only exception.

Discussion

This study demonstrates the remarkable complementation at single gene and at entire pathway level of the human glycolytic genes in *S. cerevisiae.* Despite many genetic, biochemical and regulatory differences between yeast and human enzymes (e.g. expression levels, enzyme kinetic properties, (allosteric) regulation by metabolites *in vivo*, post-translational modifications, etc.) the human enzymes can carry a high glycolytic flux in an organism as evolutionary distant as the unicellular *S. cerevisiae*. The large-scale study by Kachroo and co-workers (2015) proposed that belonging to the same process or pathway, is a better predictor for replaceability between yeast and human genes than sequence conservation (*9*). This concept is confirmed in the present study in which 10 of the 13 newly tested human genes successfully complemented their yeast ortholog. More remarkable was the conservation of moonlighting functions between yeast and human. Combined with the similarity in turnover numbers of the glycolytic enzymes expressed in yeast and in human myotubes, the present results reveal that yeast strains with fully humanized glycolytic pathways are promising models for metazoans.

Complementation or not: *Hs*HK2

Of all human glycolytic genes tested in the present study, only hexokinases proved difficult to transplant in yeast. Interestingly, only mutated variants of *Hs*HK2 complemented the glycolytic function of *Sc*Hxk2. Biochemical and genetic characterization showed that the mutations did not affect *Hs*HK2 k_{cat} and suggested that binding to mitochondria was also not hindered, as no mutations occurred in the 15 first amino acids of the N-terminus, known to be essential to this binding (*69*). While the presented data are not sufficient to identify the precise mechanisms making *Hs*HK2 functional in a yeast context, the fact that the native variant of *Hs*HK2 is functional *in vitro* but not *in vivo* demonstrates that the yeast environment is responsible for the impairment of *Hs*HK2 activity *in vivo*. Considering that mutations are localized in the vicinity of the catalytic and allosteric sites, and that yeast and human intracellular environments largely differ in metabolite concentrations (*70-72*), it is tempting to

speculate that the mutations altered the sensitivity of *Hs*HK2 to substrates or effectors. Sensitivity to trehalose-6P, a metabolite present in yeast but not in mammals, was most likely not involved in the gain of function of *Hs*HK2 *in vivo* in yeast, however binding of glucose-6P to the allosteric site, or of glucose, ATP, ADP and Mg2+ to the catalytic site might be. *Hs*HK2 is formed of two repeated, highly similar and functional domains. As mutations were observed in both domains, reported differences in affinity for ATP and glucose-6P between the two domains (*73*) cannot bring further insight in the molecular mechanisms underlying *Hs*HK2 gain of functionality in yeast. Glucose-6P is a potent inhibitor of *Hs*HK2 (K_i ca. 0.02 mM (*74*)), and intracellular concentrations typically found in yeast might fully inhibit *Hs*HK2 activity *in vivo* (2.45-12 mM (*70, 75*)). The mechanisms for glucose-6P inhibition have been mostly studied in *Hs*HK1, but for this enzyme mutating the glucose-6P binding site of only one half of the protein hardly affected the K_i for glucose-6P (2-fold at most (*76*)). If the same mechanism occurred in *Hs*HK2, a single mutation would not be sufficient to substantially modify the inhibition by glucose-6P. Additionally, among the reported mutations known to affect glucose-6P binding, none occurred on the amino acids substituted in our study. Mammalian HK2 is also reported to be inhibited by ADP (*77*). Since the concentration range of ADP in resting human skeletal muscle cells is much lower (0.09-0.11 mM (*72*)) than in yeast cells (0.87-1.32 mM (*70, 78*)) and ADP is reported to be inhibitory at a concentration of 1 mM (*77*), the mutations in *Hs*HK2 might lead to a lower sensitivity to ADP. To test further the potential role of glucose-6P, purification and fine biochemical characterization of native and mutated *Hs*HK2 will be required. In addition, "humanization" of available yeast glycolysis kinetic models should prove useful testbeds to explore the functionality of human enzymes in a yeast context (*36*). *Hs*HK4, structurally distinct from its human paralogs readily complemented *Sc*Hxk2, while its three human paralogs did not. *Hs*HK1 and *Hs*HK2 share 74% similarity at protein level, while *Hs*HK3 is more distant (53% and 56% homology with *Hs*HK1 and *Hs*HK2 respectively), and the three paralogs are functionally distinct (*33, 35*). These structural and biochemical differences between the human hexokinase paralogs most likely play a key role in their ability be transplanted in yeast and should be further explored.

From a practical perspective, the spontaneous occurrence of mutated *Hs*HK2 alleles revealed the importance to check the sequence of the gene after complementation. In most high throughput (and many small scale) studies the sequence of the gene is only verified before transformation, and not after complementation, which raises the question of how many of the genes that have been concluded to complement based on high throughput approaches can really complement upon transplantation. The present study is therefore a strong advocate for systematic sequencing upon complementation.

Remarkable conservation of moonlighting functions

Large scale, single complementation studies rarely explore the ability of the transplanted human genes to complement multiple (moonlighting) functions, as it usually requires to test the humanized strains under a broad range of specific conditions. There are several reported examples of conservation of primary and secondary functions of proteins, such as the human and yeast *HsALDOB* and *ScFBA1,* and *S. cerevisiae* and *Kluyveromyces lactis* galactokinase as transcriptional regulator of the *GAL* genes *(22, 25)*. Identifying conservation of multiple functions is particularly interesting as it has the potential to reveal unexpected connections between cellular processes and the necessity for cells to coordinate their activities. While ubiquitously present in abundance and highly conserved across kingdoms (40% to 90% sequence homology at protein level among bacterial and eukaryotic species (*79*)), enolases are particularly versatile enzymes in term of localization and moonlighting functions (*80*). The present study demonstrates for the first time that the human and yeast enolases fulfil a similar secondary function in vacuolar fusion and transport. More remarkably, yeast enolases have yet another moonlighting function, in the import of $tRNA^{Lys}$ (CUU) (called tRK1) into mitochondria (*81*). In mammals and *S. cerevisiae*, growth above 37°C requires the translocation of the cytosolic tRK1 into mitochondria, a conserved mechanism facilitated by enolases (*81*). It has been shown that *S. cerevisiae* tRK1 is imported *in vitro* and *in vivo* in human mitochondria, in an enolase-dependent manner (*82, 83*). The ability of fully humanized yeast strains to grow at 37°C suggests that the human enolase 3 is also able to fully take over the yeast enolase function in mitochondrial tRK1 import. This conservation in functionality might be explained by the high degree of sequence similarity between human and yeast enolases (62-63%). In the present study the muscle enolase variant *Hs*Eno3 was more particularly investigated, however *Hs*Eno1, the predominantly expressed enolase form across mammalian tissues, showed a higher level of complementation in terms of growth rate than enolase 2 and enolase 3 in yeast (92% of the yeast control strain with native *Sc*Eno2, Fig. 1). Further experiments involving growth on respiratory carbon sources at 37°C, a condition leading to a stronger phenotype in tRK1-deficient strains for mitochondrial import, of the set of complementation strains with all human enolases will reveal the degree of conservation of this moonlighting function among human enolases (*81*). The remarkably portability of human enolase moonlighting functions in *S. cerevisiae* offers new avenues to study its involvement in human diseases.

The current data suggest that *Hs*HK2 did not complement the yeast *Sc*HXK2 moonlighting function. However, considering the sensitivity of invertase derepression to growth rate (*84, 85*) and the slow growth rate of the *Hs*HK2* complementation strain

(IMX1690, 0.18 h^{-1} ± 0.003), the absence of invertase activity and of fermentative metabolism in this strain is not sufficient to discard the potential role of *Hs*HK2 in glucose repression. Nevertheless, considering the complexity of the mechanisms enabling the yeast Hxk2 to respond to glucose availability, migrate to the nucleus and repress gene expression, it appears highly unlikely that the structurally different *Hs*HK2 fulfils the same function (*47, 86, 87*). Additionally, *Hs*HK2 is not localized in the nucleus in mammals and does not harbour a nuclear localization signal (*74*). The systematic mutation in the carbon catabolite repression general regulator Tup1 in all evolution lines of the *Hs*Gly-HK2 strain might be further evidence of the inability of *Hs*HK2 to contribute to glucose repression in *S. cerevisiae*. Measurements of invertase activity in the fully humanized yeast strains and in the evolved strains might bring more insight in the involvement of *Hs*Hk2 in glucose repression. *Hs*HK4 is structurally closer to *Sc*Hxk2 than the other human hexokinases, as it consist of a single subunit. However, the observation that *Hs*HK4 is able to complement the role of *Sc*Hxk2 in glucose repression, in line with an earlier report (*21*), is confounding. Firstly, the degree of conservation between these two proteins is low (30%), secondly *Hs*HK4 lacks the Lys^{7}-Met^{16} decapeptide indispensable for Mig1 binding and translocation to the nucleus and glucose repression (*47, 87*). While *Hs*HK4 localizes to the nucleus in human cells, its translocation is mediated by the glucokinase regulatory protein (GCKR) that has no homolog in yeast, its nuclear localization is induced under low glucose conditions and it is not involved in transcriptional regulation (*74, 88*). The mechanisms underlying *Hs*HK4 ability to repress invertase and enable alcoholic fermentation therefore remain to be elucidated. Localization studies will be required to bring further insight in the moonlighting functions of these two hexokinases.

Considering the key cellular role of glycolysis, it is not surprizing that many glycolytic proteins have alternative functions and localizations that enable to directly connect energy metabolism and carbon source availability to other cellular functions. Their level of conservation between yeast and human is however remarkable and offers new yeast simplified models to study complex human processes.

Combining single complementation, full pathway transplantation and evolution

This work presents a first demonstration of the power of combining single gene complementation, full pathway transplantation and ALE to study pathway or process transplantation. In such large scale studies at full process or pathway level, it can be expected that the transplantation in a new cellular environment will likely negatively affect the host physiology. Investigating the fully humanized strains can only give a partial understanding of the impact of the humanized proteins on the host. In the present study the individual complementation strains played a key role in

understanding the impact of the various human enzymes in the slow growth rate phenotype of the fully humanized strains. For future study, combination of complementation and fully humanized strains should be further explored to identify potential synergetic effect between the different human enzymes. An intriguing observation was the substantially higher (at least two-fold) specific activity of *Hs*PFKM expressed in *Hs*Gly-HK2 as compared to *Hs*Gly-HK4, while *Hs*PFKM is expressed with the same promoter in both strains. This difference is even more pronounced in the evolved strains (Fig. S10) and hints towards a potential synergy between hexokinases and phosphofructokinase, for which the underlying mechanism should be further explored with simple complementation strains expressing *Hs*PFKM with either *Hs*HK2 or *Hs*HK4.

Nevertheless, as exemplified by this study, in complex engineered strains harbouring the humanization of a large set of genes, fully humanized and complementation strains were also not sufficient to unravel the full range of mechanisms involved in the host physiological responses. The present study demonstrates the power of ALE to fill in this knowledge gap, by identifying mechanisms that are too laborious to explore experimentally by targeted strain engineering or are simply not predictable. For instance ALE identified a set of glycolytic enzymes which activity was too low to sustain a high glycolytic flux (Fig. 5). The molecular mechanisms involved in this increased activity are still unclear, as no mutations were found in the coding or non-coding regions of the glycolytic genes and their known transcriptional regulators, and no change in gene, chromosomal region or full chromosome copy number was found (Fig. S12). Previous studies have shown that the abundance of glycolytic enzymes can be regulated at translation level by yet unknown mechanisms (*89, 90*). Monitoring transcript and protein levels in the evolved strains and their ancestor might bring new light on how *S. cerevisiae* regulate the translation of glycolytic genes.

Remarkably, while comparing the single complementation and fully humanized strains suggested that *Hs*ALDOA low activity contributed to their slow growth rate, *Hs*ALDOA activity was not increased in the evolved strains. The systematic mutation of *STT4* in all evolved strains suggested that yeast cells found an alternative strategy to enhance *Hs*ALDOA activity *in vivo*. PI4P kinases are conserved eukaryotic proteins (*91*), and Stt4 has a human ortholog, PI3K. In mammals, activation of PI3K remodels actin, thereby releasing aldolase A trapped in the actin cytoskeleton in an inactive state and increasing cellular Aldolase A activity (*92, 93*). As aldolase of rabbit skeletal muscle which is 98.4% identical to human aldolase A was shown to bind to yeast actin filaments (*94*) and yeast and human forms of actin are highly conserved (89% identity at protein level) it is likely that a similar mechanism is active in yeast and enables the evolved, humanized

yeast strains to increase Aldolase A activity *in vivo* without increasing its concentration. Reverse engineering of *SST4* mutations in *Hs*Gly-HK2 and *Hs*Gly-HK4 resulted in defects in vacuoles morphologies (Fig. 4 and S11). The absence of growth rate increase in these reversed engineered strains and the intact morphology of vacuoles in evolved strains suggest that specific mechanisms might have evolved in parallel to mitigate the negative effect of *STT4* mutations on vacuolar morphology. As other glycolytic enzymes such as *Hs*GAPDH bind actin in mammals and yeast (*94*), it is conceivable that a similar Stt4-mediated regulation occurs for other human enzymes in the evolved yeast strains.

Outlook

In this study we showed that it is possible to fully humanize the glycolytic pathway using the SwyG strain. These results open up the way to construct any tissue-specific glycolysis and combinatorial approaches can be used to construct many human glycolytic variants with native genes or with alleles across several genes to study polygenic diseases. In human cells hexose transport plays an important role in regulation of glycolytic flux . Recently a CRISPR-toolkit was published that can be used to eliminate all glucose transporters in three transformations (*95*). Using this toolkit hexose transport could subsequently be humanized leading to an improved yeast model with full humanization of sugar import and catabolism.

Material and methods

Strains, media and laboratory evolution

All strains used in this study are derived from a CEN.PK background (*96*) and are listed in table S8. Yeast strains were propagated on YP medium containing 10g L^{-1} Bacto Yeast extract, 20 g L^{-1} Bacto Peptone or synthetic medium containing 5 g L^{-1} $(NH_4)_2SO_4$, 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4 \cdot 7 \cdot H_2O$, and 1 mL L^{-1} of a trace elements and vitamin solution (*97*). Media were supplemented with 20 g L^{-1} glucose or galactose or 2% (v/v) ethanol. For the characterization of the individual hexokinase complementation strains (Fig. 4A), $(NH_4)_2SO_4$ was replaced with 6.6 g L^{-1} K_2SO_4 and 2.3 g L^{-1} urea to reduce acidification of the medium. Urea was filter sterilized and added after heat sterilization of the medium at 121°C. When indicated, 125 mg L^{-1} histidine was added. For solid media 2% (w/v) agar was added to the medium prior to heat sterilization. The pH of SM was adjusted to pH 6 by addition of 2 M KOH, which was also used for preparation of YPD medium at 7.5 pH. For selection, YP medium was supplemented with 200 mg L^{-1} G418 (KanMX) or 100 mg L^{-1} nourseothricin (Clonat). For removal of the native yeast glycolysis cassette from the *sga1* locus the SM glucose (SMG) medium was supplemented with 2.3 g L^{-1} fluoracetamide to counter select for the *AmdS* marker present in the cassette (*98*). For plasmid propagation chemically competent *Escherichia*

coli XL1-Blue (Agilent Technologies, Santa Clara, CA) cells were used which were grown in lysogeny broth (LB) supplemented with 100 mg L^{-1} ampicillin, 25 mg L^{-1} chloramphenicol or 50 mg L^{-1} kanamycin when required (*99, 100*). Yeast and *E. coli* strains were stored at -80 °C after addition of 30% (v/v) glycerol to an overnight grown culture.

For all growth experiments in shake flask, 100 mL medium in a 500 mL shake flask was used except for the growth study with the individual complementation strain for which 20 mL in a 100 mL volume shake flaks was used. Strains were incubated with constant shaking at 200 rpm and at 30°C unless stated otherwise. Strains were inoculated from glycerol stocks in YPD and grown overnight. This culture was used to inoculate the pre-culture (SMG) from which the exponentially growing cells were transferred to new shake flasks to start a growth study.

Growth studies in microtiter plate were performed at 30 °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 YPD or YPGal medium in a 50 mL volume shake flask. This culture was used to inoculate a preculture in a 24-wells plate with a 1 mL working volume (EnzyScreen, type CR1424f), which was cultivated until mid/late-exponential growth in the conditions of interest in the growth profiler. From this culture the growth study was started in a 96-wells microtiter plate (EnzyScreen, type CR1496dl), with final working volumes of 250 µL and approximate starting OD_{660} of 0.1-0.2. Microtiter plates were closed with a sandwich cover (EnzyScreen, type CR1296). Images of cultures were made at 30 min intervals. Corrected green-values were obtained with software supplied and installed by the manufacturer. These values were used for conversion to OD-values based on a 16-point calibration, based on the following equation:

$$\text{OD-equivalent} = 0.0692\times(GV(t) - GV_{med}) + 1.52\times10^{-4}\times(GV(t) - GV_{med})^{2.63} + 4.4\times10^{-14}\times(GV(t)-GV_{med})^{7.2},$$

in which GV(t) is the corrected green-value measured in a well at time point 't', and GV_{med} is a green-value obtained from a measurement of medium without cells. Growth rates were calculated over a time frame in which OD doubled at least twice.

Adaptive laboratory evolution of IMX1814 and IMX1844 was performed in SMG at 30°C in 100 mL volume shake flasks with a working volume of 20 mL. Initially every 48 hours 200 µL of the culture was transferred to a new shake flask with fresh medium, after 22 transfers (approximately 170 generations) this was done every 24h. For both strains

three evolution lines were run in parallel. At the end of the experiment single colony isolates were obtained by restreaking three times on YPD plates (Table S8D).

Molecular techniques, gene synthesis and Golden Gate plasmid construction

PCR amplification for cloning purposes was performed with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA) according to the manufacturers recommendations except that the primer concentration was lowered to 0.2 µM. PCR products for cloning and Sanger sequencing were purified using the Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine, CA) or the GeneJET PCR Purification kit (Thermo Fisher Scientific). Sanger sequencing was performed at Baseclear BV (Baseclear, Leiden, The Netherlands). Diagnostic PCR to confirm correct assembly, integration of the constructs and sequence verification by Sanger sequencing was done with DreamTaq mastermix (Thermo Fisher Scientific) according to the manufacturers recommendations. To obtain template DNA, cells of single colonies were suspended in 0.02 M NaOH, boiled for 5 min and spun down to use the supernatant. All primers used in this study are listed in Tables S9A-S9G. Primers for cloning purposes were ordered PAGE purified, the others desalted. To obtain gRNA and repair fragments the designed forward and reverse primers were incubated at 95°C for 5 min to obtain a double stranded piece of DNA. PCR products were separated in gels containing 1% agarose (Sigma) in Tris-acetate buffer (TAE). Genomic DNA from CEN.PK113-7D was extracted using the YeaStar™ Genomic DNA kit (Zymo Research Corporation, Irvine, CA, USA). Cloning of promoters, genes and terminators was done using Golden Gate assembly. Per reaction volume of 10 µl, 1 µl T4 buffer (Thermo Fisher Scientific), 0.5 µl T7 DNA ligase (NEB New England Biolabs, Ipswich, MA) and 0.5 µl BsaI (Eco31I) (Thermo Fisher Scientific) or BsmBI (NEB) was used and DNA parts were added in equimolar amounts of 20 fmol as previously described (*101*). First a plasmid backbone was constructed from parts of the yeast toolkit (*101*) using a kanamycin marker, *URA3* marker, bacterial origin of replication, 3'and 5' *ura3* integration flanks and a *GFP* marker resulting in pGGKd002 (Table S10C). In a second assembly, the *GFP* gene in this plasmid was replaced by a transcriptional unit containing a *S. cerevisiae* promoter and terminator and a human glycolytic gene. The sequences of the human glycolytic genes were obtained from the Uniprot data base (www.uniprot.org), codon optimized for *S. cerevisiae* and ordered from GeneArt Gene Synthesis (Thermo Fisher Scientific). Genes were synthetized flanked with BsaI restriction sites to use them directly in Golden Gate assembly (Table S10A). The *PKL* gene which is a shorter splicing variant of *PKR* was obtained by amplifying it from the *PKR* plasmid pGGKp024 using primers containing BsaI restriction site flanks (Table S9G). *S. cerevisiae* promoters and terminators were PCR amplified from genomic DNA using primers flanked with BsaI and BsmBI

restriction sites (Table S9A) (*102*). The resulting PCR product was directly used for Golden Gate assembly. For long term storage of the fragments, the promoters and terminators were cloned into the pUD565 entry vector using BsmBI Golden Gate cloning resulting in the plasmids pGGKp025-048 listed in Table S10B. For the *HXK2* and *TEF2* promoters and *HXK2* and *ENO2* terminators already existing plasmids were used (Table S10B). For the construction of pUDE750 which was used as PCR template for the amplification of the *HK4* fragment used in IMX1814, first a dropout vector (pGGKd003) was constructed from the yeast toolkit parts pYTK002, 47, 67, 74, 82 and 84 (Table S10C). In this backbone, *ScHXK2p, HK4* and *ScHXK2t* were assembled as described above (Table S10E). Plasmid isolation was done with the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MO). Yeast transformations were performed according to the lithium acetate method (*103*).

Construction of individual gene complementation strains

To enable CRISPR/Cas9 mediated gene editing, *Cas9* and the *NatNT1* marker were integrated in the *SGA1* locus of the minimal glycolysis strain IMX370 by homologous recombination, resulting in strain IMX1076 (*26*). *Cas9* was PCR amplified from p414-*TEF1p-Cas9-CYC1t* and *NatNT1* from pUG-*natNT1* (Tables S9G and S10E) and 750 ng of both fragments were, after gel purification, used for transformation.

For the individual gene complementation study, 400 ng of the constructed plasmids containing the human gene transcriptional units (Table S10D) were linearized by digestion with NotI (FastDigest, Thermo Fisher Scientific) according to the manufacturer's protocol for 30 min and subsequently the digestion mix was directly transformed to IMX1076. The linearized plasmids were integrated by homologous recombination in the disrupted *ura3-52* locus of strain IMX1076 and the transformants were plated on SMG. After confirmation of correct integration by PCR (Table S9B), in a second transformation the orthologous yeast gene (or genes, in case of *PFK1* and *PFK2*) was removed using CRISPR/Cas9 according to the protocol of Mans *et al.* (*104*). Since only the yeast gene and not the human ortholog should be targeted, the gRNAs were designed manually (Table S9D). For deletion of *FBA1*, *GPM1*, and *PFK1* and *PFK2,* the plasmids containing the gRNA were preassembled as previously described (*104*) using Gibson assembly and a PCR amplified pROS13 backbone containing the KanMX marker (Tables S9D and S10F). For *HXK2* deletion, the double stranded gRNA and a PCR amplified pMel13 backbone were assembled using Gibson assembly (Table S9D). The constructed plasmids were verified by PCR. The rest of the gRNA plasmids for yeast gene deletion were assembled *in vivo* in yeast and were not stored as individual plasmid afterwards. For the *in vivo* assembly approach the strains were co-transformed with 100 ng of the PCR amplified backbone of pMel13 (Table S9D and S10F), 300 ng of the

double stranded gRNA of interest (Table S9D) and 1 µg repair fragment to repair the double stranded break (Table S9E). For pre-assembled plasmids, strains were co-transformed with 0.6-1 µg of plasmid (Table S10F) and 1 µg repair fragment (Table S9E). Transformants were plated on YPD + G418 and for the *HsHK1-HK3* strains on YPGal + G418. Successful gene deletion was confirmed with diagnostic PCR (Table S9B, Fig. S14). gRNA plasmids were afterwards removed by several restreaks on non-selective medium. To test if the complementation was successful, the strains were tested for growth in SMG. For *ScHK2,* three complementation strains were made. IMX1690 (*pScPDC1-HsHK2)* and IMX1873 (*pScHXK2-HsHK2*) which were grown on glucose medium and contain a mutation in *HsHK2* and IMX2419 (*pScPDC1-HsHK2)* which was never exposed to glucose and does not contain mutations. *HsHK4* was also expressed both with the *ScHXK2* and *ScPDC1* promoter, resulting in IMX1874 and IMX1334 respectively (Table S8A,B). An overview of the workflow is provided in Fig. S1. To test for the occurrence of mutations, the human gene transcriptional units were PCR amplified and send for Sanger sequencing.

Full human glycolysis strain construction

For the construction of the strains containing a full human glycolysis, the transcriptional units of the *HsHK2, HsHK4, HsGPI, HsPFKM, HsALDOA, HsTPI1, HsGAPDH, HsPGAM2, HsENO3,* and *HsPKM1* gene were PCR amplified from the same plasmids as were used for the individual gene complementation using primers with flanks containing synthetic homologous recombination (SHR) sequences (Table S9C and S10D). An exception was made for the *HsHK2* and *HsHK4* gene for which pUDE750 and pUDI207 were used as template, which contain the Sc*HXK2* promoter and terminator. An overview of the promoters used for the human gene expression is provided in Table S2. The yeast *PDC1* and *ADH1* genes were amplified with their corresponding promoter and terminator regions from genomic DNA from CEN.PK113-7D (Table S8E). The fragments were gel purified and the fragments were assembled in the *CAN1* locus of strain IMX589 by *in vivo* assembly. 160 fmol per fragment and 1 µg of the pMel13 plasmid targeting *CAN1* was used. Transformation mix was plated on YPD + G418 and correct assembly was checked by PCR and resulted in strain IMX1658. In a second transformation, the cassette in the *SGA1* locus containing the native *S. cerevisiae* glycolytic genes and the *AmdS* marker was removed. To do this IMX1658 was transformed with 1 µg of the gRNA plasmid pUDE342 (Table S10F) and 2 µg repair fragment (counter select oligo) (Table S9E) and plated on SMG medium with fluoracetamide to counter select for the *AmdS* marker. From the resulting strain the pUDE342 plasmid was removed and it was stored as IMX1668. To replace the *HsHK4* gene with *HsHK2,* the *HsHK2* gene was PCR amplified from pGGKp002 using primers

flanked with sequences homologous to the *ScHXK2* promoter and terminator to allow for recombination (Table S9G). IMX1668 was co-transformed with this fragment and pUDR387 containing the gRNA targeting the *HsHK4* gene and the cells were plated on YPD + G418 (Table S10F). After confirmation of correct integration by PCR and plasmid removal, the strain was stored as IMX1785. The pUDR387 gRNA plasmid was constructed with Gibson Assembly from a pMel13 backbone and double stranded *HsHK4* gRNA fragment (Table S9D). To make the constructed yeast strains prototrophic, the *ScURA3* marker was PCR amplified from CEN.PK113-7D genomic DNA using primers with flanks homologous to the *TDH1* region (Table S9G) and integrated in the *tdh1* locus of IMX1785 and IMX1668, by transforming the strains with 500 ng of the fragment and plating on SMG. This resulted in IMX1844 and IMX1814 respectively. IMX2418 (*Hs*Gly *HK2* strain without mutation in Hs*HK2*) was constructed by transforming IMX1814 with pUDR387 and the *HsHK2* fragment amplified as described above. The cells were plated on YPGal + G418 and later restreaked on YPGal plates to remove the plasmid. For the overexpression of *HsALDOA, HsPGAM2* and *HsHK2/HsHK4* resulting in IMX2005 and IMX2006, the expression cassettes were PCR amplified from pUDI141, pUDI150, pUDI134 and pUDI136 respectively using primer sets 12446/12650, 12467/14542 and 14540/14541 (Table S9C, S10D). IMX1844 and IMX1814 were transformed with 160 fmol per fragment and 1 µg of the plasmid pUDR376 containing a gRNA targeting the X2 locus (*105*) and plated on SMG-acetamide plates. To obtain the reference strain IMX1821 which contains a yeast glycolysis cassette integrated in *CAN1,* the pUDE342 plasmid was removed from the previously described strain IMX605 (*27*) and the *URA3* fragment was integrated in *tdh1* in the manner as described above. An overview of strain construction is provided in Fig. S15.

STT4 reverse engineering

The single nucleotide polymorphisms (SNPs) which were found in the *STT4* gene of evolved strains IMS0990 and IMS0992 resulting in amino acid changes G1766R and F1775I respectively, were introduced in the *STT4* genes of the non-evolved strains IMX1814, IMX1844 and IMX1822 using CRISPR/Cas9 editing (*104*) (Table S8D). Two gRNA plasmids pUDR666 and pUDR667 were constructed using Gibson Assembly of a backbone amplified from pMel13 (Table S9D, S10F) and a gRNA fragment consisting of oligo 16748+16749 and 16755+16756 respectively (Table S9D). For introduction of the G1766R mutation, strains were transformed with 500 ng of pUDR666 and 1 µg of repair fragment (oligo 16750+16751) and for introduction of F1775I with 500 ng of pUDR667 and 1 µg of repair fragment (oligo 16757+16758) (Table S9E, S10F). Strains were plated on YPD + G418 and introduction of the mutation was verified by Sanger sequencing. The control strain IMX1822 containing the native yeast minimal glycolysis

in the *SGA1* locus originates from strain IMX589 (*27*). From this strain the *AmdS* marker was removed by transforming the strain with 1 µg repair fragment (oligo 11590+11591) and 300 ng of a gRNA fragment (oligo 11588+11589, Table S9D) targeting *AmdS* and 100 ng of backbone amplified from pMel10 resulting in a *in vivo* assembled gRNA plasmid. After removal of the plasmid by restreaking on non-selective medium, this strain, IMX1769, was made prototrophic by integrating *ScURA3* in *tdh1* (Table S9G, S8E), resulting in IMX1822.

Characterization of *ScHK2* mutants

Sanger sequencing of the *HsHK2* complementation strains showed the presence of mutations in all strains after growth on glucose. All found mutations in *HsHK2* were mapped unto the protein sequence and visualized on the structural model with PDB code 2NZT (*34*) using the PyMOL Molecular Graphics System, version 1.8.6 (Schrödinger LLC).

Construction of *Δhxk1Δhxk2* strain IMX165 and control strain IMX2015

The *Δhxk1Δhxk2* strain IMX165 which was used as control in the invertase assay was constructed in three steps. The *HXK1* and *HXK2* deletion cassettes were PCR amplified from pUG73 and pUG6 respectively using the primers listed in Table S9F. First, *HXK1* was removed from CEN.PK102-12A by transformation with the *HXK1* deletion cassette containing the *Kluyveromyces lactis LEU2* marker flanked with loxP sites and *HXK1* recombination flanks resulting in strain IMX075. To remove the *LEU2* marker from this strain, it was transformed with the plasmid pSH47 containing the galactose inducible Cre recombinase (*106*). Transformants were plated on SMG with histidine and were transferred to YPGal for Cre recombinase induction to remove *LEU2,* resulting in strain IMS0336. Subsequently, this strain was transformed with the *HXK2* deletion cassette containing the *KanMX* marker flanked with LoxP sites and *HXK2* recombination flanks, resulting in IMX165. IMX2015 was constructed as control strain for the characterization of the human hexokinase complementation strains. In this strain Sc*HXK2* is expressed with the p*PDC1* promoter instead of the native *HXK2* promoter. *pPDC1* was PCR amplified from genomic DNA from CEN.PK113-7D with primer 14670 and 14671 containing *HXK2* recombination flanks. 500 ng of this fragment was transformed to IMX1076 together with 800 ng of pUDE327 containing a gRNA targeting the *HXK2* promoter (table S10F).

Illumina whole genome sequencing

Genomic DNA for sequencing was isolated with the the Qiagen 100/G kit according to the manufacturer's description (Qiagen, Hilden, Germany) and library preparation and sequencing was done as described previously using Illumina Miseq sequencing

(Illumina, San Diego, CA) (*102*). A list of mutations is provided in Table S4. For the mutation found in the *SBE2* gene which is involved in bud growth, it is unlikely to have an effect since it has a functionally redundant paralog *SBE22 (107)*. No abnormalities were observed under the microscope.

Quantitative aerobic batch cultivations

Quantitative characterization of strain IMX1821, IMX1814 and IMX1844 was done in 2 L bioreactors with a working volume of 1.4 L (Applikon, Schiedam, The Netherlands). The cultivation was done in synthetic medium supplemented with 20 g L^{-1} glucose, 1.4 mL of a vitamin solution (*97*) and 1.4 mL of 20% (v/v) Antifoam emulsion C (Sigma, St. Louise, USA). During the fermentation 0.5 mL extra antifoam was added when necessary. The salt and antifoam solution were autoclaved separately at 121°C and the glucose solution at 110°C for 20 min. During the fermentation the temperature was kept constant at 30°C and the pH at 5 by automatic addition of 2 M KOH. The stirring speed was set at 800 rpm. The medium was flushed with 700 mL min^{-1} of air (Linde, Gas Benelux, The Netherlands).

For preparation of the inoculum, freezer stocks were inoculated in 100 mL YPD and grown overnight. From this culture the pre-culture was inoculated in 100 mL SMG which was incubated till mid-exponential growth phase. This culture was used to inoculate the inoculum flasks which were incubated till OD 4.5. The cells were centrifuged for 10 min at 3000g and the pellet was suspended in 100 mL demineralized water and added to the fermenter to start the fermentation with an OD of 0.25-0.4.

Biomass dry weight determination was done as previously described (*97*) by filtering 10 mL of culture on a filter with pore-size 0.45 mm (Whatman/GE Healthcare Life Sciences, Little Chalfont, United Kingdom) in technical duplicate. For extracellular metabolite analysis 1 mL of culture was centrifuged for 3 min at 20000g and the supernatant was analysed using high performance liquid chromatography (HPLC) using an Aminex HPX-87H ion-exchange column operated at 60°C with 5 mM H_2SO_4 as the mobile phase with a flow rate of 0.6 mL min^{-1} (Agilent, Santa Clara). The OD_{660} was measured with a Jenway 7200 spectrophotometer (Jenway, Staffordshire, UK) at 660 nm. Per strain at least two independent fermentations were performed. The carbon balances for all reactors closed within 5%.

Sample preparation and enzymatic assays for comparison of yeast and humanized yeast samples

Yeast samples were prepared as previously described (*108*), from exponentially growing cultures (62 mg dry weight per sample) from bioreactor and for testing of

allosteric effectors and for comparison of the evolved strains from shake flask. All determinations were performed at 30°C and 340 nm (εNAD(P)H at 340 nm/6.33mM^{-1}).

In most cases glycolytic V_{max} enzyme activities were determined in 1 mL reaction volume (in 2 mL cuvettes), using a Hitachi model 100-60 spectrophotometer, using previously described assays (*57*), except for phosphofructokinase activity which was determined according to Cruz *et al.(109)*. To increase throughput, the specific activities of the evolved strains were assayed using a TECAN infinite M200 Pro. (Tecan, Männedorf, Switzerland) microtiter plate reader. Samples were prepared manually in microtiter plates (transparent flat-bottom Costar plates; 96 wells) using a reaction volume of 300 µl per well. The assays were the same as for the cuvette-based assays.

The reported data are based on at least two independent biological replicate samples, with at least two analytic replicates per sample per assay, including two different cell free extract concentrations.

The protein concentration was determined using the Lowry method with bovine serum albumin as a standard (*110*). Enzyme activities are expressed as µmol substrate converted (mg protein)$^{-1}$ h^{-1}.

To calculate the degree of saturation of glycolytic enzymes, the specific activity in µmol.mg$_{protein}^{-1}$.h^{-1} was converted into mmol.g$_{DW}^{-1}$.h^{-1} considering that soluble proteins represent 30% of cell dry weight. This value represents the maximal enzyme flux capacity. The *in vivo* flux in the glycolytic reactions were approximated from the glucose specific uptake rate (q_{glu}). Reactions in the top of glycolysis (hexokinase to triosephosphate isomerase) were assumed to equal the q_{glu}, while reactions in the bottom of glycolysis (glyceraldhyde-3P dehydrogenase to pyruvate kinase) were calculated as the q_{glu} times two. The degree of saturation was calculated as follows:

$$degree\ of\ saturation = \frac{approximated\ in\ vivo\ flux}{maximal\ flux\ capacity} \times 100$$

Invertase enzyme assay

The invertase assay was performed on whole cells previously described (*111*). Exponentially growing cells in SMG were washed with sterile dH_2O, transferred to shake flasks (at OD 3) with 100 mL fresh SMG or SME+0.075% glucose and incubated for 2h at 30°C and shaking at 200 rpm. Afterwards the dry weight of the cultures was determined and the cells were washed in 50mM sodium acetate buffer wit 50mM NaF to block the metabolism and were then suspended till a concentration of 2.5-7.5 mg dry weight per mL. 4 mL of this cell suspension were added to a dedicated vessel thermostated at 30°C, and kept under constant aeration by flushing with air (Linde, Gas

Benelux, The Netherlands) and stirring with a magnetic stirrer. The reaction was started by addition of 1 mL 1M sucrose and 1 mL reaction mix was taken at 0, 1, 2, 3 and 5 minutes, directly filtered using 13 mm diameter 0.22 µm pore size nylon syringe filters to remove cells and put on ice. Afterwards the glucose concentration resulting from sucrose hydrolysis by invertase was determined using a D-Glucose assay kit (Megazyme, Ireland). The glucose production rate was calculated in $\mu Mol \cdot min^{-1} \cdot g$ dry $weight^{-1}$.

Staining of vacuoles

Yeast strains were stained with the red fluorescent dye FM4-64 (excitation/emission, 515/640 nm) (Thermo Fisher Scientific). Exponentially growing cells were incubated at an OD of 0.5-1 in YPD with 2 µM FM4-64 in the dark for 30 min at 30˚C. Afterwards cells were spun down, washed and incubated for 2-3 h in 5 mL YPD. For analysis, cells were spun down and suspended in SMG medium. Yeast cells and vacuoles were visualized with an Imager-Z1 microscope equipped with an AxioCam MR camera, an EC Plan-Neofluar 100x/1.3 oil Ph3 M27 objective, and the filter set BP 535/25, FT 580, and LP 590 (Carl-Zeiss, Oberkochen, Germany).

Ploidy determination by flow cytometry

Samples of culture broth (equivalent to circa 10^7 cells) were taken from mid-exponential shake-flask cultures on YPD and centrifuged (5 min, 4700g). The pellet was washed once with H_2O_d, and centrifuged again (5 min, 4700g) and suspended in 800 µL 70% ethanol while vortexing. After addition of another 800 µL 70% ethanol, fixed cells were stored at 4˚C until further staining and analysis. Staining of cells with SYTOX® Green Nucleic Acid Stain (Invitrogen S7020) was performed as described (*112*). Samples were analyzed on a BD Accuri C6 flow cytometer equipped with a 488 nm laser (BD Biosciences, Breda, The Netherlands). The fluorescence intensity (DNA content) was represented using FlowJo (v. 10.6.1, FlowJo, LLC, Ashland, OR, USA), (Fig. S12).

Transition experiment

For testing transitioning between carbon sources, strains were grown overnight in SMGal medium till mid-exponential phase. These cultures were used to plate single cells on SMG and SMGal plates (96 cells per plate) using a BD FACSAriaII (Franklin Lakes, NJ). After 5 days the colonies were counted to determine the percentage of cells which was growing.

Comparison human and humanized yeast glycolytic enzymes

Human cell culture and harvest

Human myoblasts were obtained from orbicularis oculi muscle biopsies, as previously described (*113*). Briefly, subclone V49 expressed Pax7, MyoD and Myogenin and was used for the assays here described. Cells were maintained in high glucose Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich/Merck) in the presence of L-glutamine, 20% fetal bovine serum (FBS, Life Technologies Gibco/Merck) and 1% penicillin/streptomycin (p/s, Sigma-Aldrich/Merck). For differentiation, cells were seeded on 10 cm dishes covered with polydimethylsiloxane (PDMS) gradients at 5,000 cells/cm^2 and after reaching confluence, medium was changed to DMEM in the presence of 2% FBS, 1% p/s, 1% Insulin-Transferrin-Selenium (Life Technologies Gibco/Merck) and 1% dexamethasone (Sigma-Aldrich/Merck). The presence of PDMS gradients allows cells to grow aligned, which in turn improves myotube maturity and functionality. Cells were harvested after 5 days in differentiation medium. In short, cells were washed twice with ice-cold Dulbecco's Phosphate Buffered Saline (DPBS, Gibco) and scraped in DPBS in the presence of Complete Protease Inhibitor Cocktail (Merck, 11836145001, 1:25 v/v after resuspension according to manufacturer's guidelines). Cells were frozen at -80 °C.

Cell-free extract preparation and V_{max} enzyme assays

Human cells stored at -80°C were thawed, centrifuged at 20000 g for 10 minutes at 4 °C and the pellet was discarded to obtain cell-free extracts.

Yeast samples (IMX1844) were harvested as previously described (*108*) from exponentially growing cultures (62 mg dry weight per sample) from bioreactor. Cell-free extract preparation for yeast cells was done using YeastBuster™ Protein Extraction Reagent supplemented with 1% of 100x THP solution according to the description (Novagen, San Diego, CA, USA). To a pellet with a wet weight of 0.3g, 3.5 mL YeastBuster and 35 µl THP solution was added. The pellet was suspended and incubated for 20 min at room temperature. Afterwards the cell debris was removed by centrifugation at 20000g for 15 min at 4°C and the supernatant was used for the assays.

Prior to experimentation, YeastBuster™ Protein Extraction Reagent with 1% THP (Novagen) was added to the human cell samples and DPBS supplemented with protease inhibitor was added to the yeast samples (both as 50% of final volume). This strategy was taken in order to equalize the buffer composition of yeast and human culture samples to perform enzyme kinetics assays and proteomics.

V_{max} assays for comparison of yeast and human cell extracts were carried out with freshly prepared extracts via NAD(P)H-linked assays at 37 °C in a Synergy H4 plate reader (BioTek™). The reported V_{max} values represent total capacity of all isoenzymes in the cell at saturating concentrations of all substrates and expressed per extracted cell protein. Four different dilutions of extract were used to check for linearity. Unless otherwise stated, at least 2 dilutions were proportional to each other and these were used for further calculation. All enzymes were expressed as µmoles of substrate converted per minute per mg of extracted protein. Protein determination was carried out with the Bicinchoninic Acid kit (BCA™ Protein Assay kit, Pierce) with BSA (2 mg/ml stock solution of bovine serum albumin, Pierce) as standard.

Based on the cytosolic concentrations described in literature, we have designed an assay medium that was as close as possible to the *in vivo* situation, whilst at the same time experimentally feasible. The standardized *in vivo*-like assay medium contained 150 mM potassium(*114-117*), 5 mM phosphate (*114, 118*), 15 mM sodium (*114, 119*), 155 mM chloride (*120, 121*), 0.5 mM calcium, 0.5 mM free magnesium (*114, 122, 123*) and 0.5-10.5 mM sulfate. For the addition of magnesium, it was taken into account that ATP and ADP bind magnesium with a high affinity. The amount of magnesium added equalled the concentration of either ATP or ADP plus 0.5 mM, such that the free magnesium concentration was 0.5 mM. Since the sulfate salt of magnesium was used the sulfate concentration in the final assay medium varied in a range between 0.5 and 10.5 mM. The assay medium was buffered at a pH of 7.0 (*124-129*) by using a final concentration of 100 mM Tris-HCl (pH 7.0). To end up with the above concentrations, an assay mixture containing 100 mM Tris-HCl (pH 7.0), 15 mM NaCl, 0.5 mM $CaCl_2$, 140 mM KCl, and 0.5-10.5 mM $MgSO_4$ was prepared.

In addition to the assay medium, the concentrations of the coupling enzymes, allosteric activators and substrates for each enzyme were as follows:

Hexokinase (HK; EC2.7.1.1) – 1.2 mM $NADP^+$, 10 mM Glucose, 1.8 U/mL glucose-6-phosphate dehydrogenase (EC1.1.1.49), and 10 mM ATP as start reagent.

Phosphoglucose isomerase (GPI; EC5.3.1.9) – 0.4 mM $NADP^+$, 1.8 U/mL glucose-6-phosphate dehydrogenase (EC1.1.1.49), and 2 mM fructose 6-phosphate as start reagent.

Phosphofructokinase (PFK; EC2.7.1.11) – 0.15 mM NADH, 1 mM ATP, 0.5 U/mL aldolase (EC4.1.2.13), 0.6 U/mL glycerol-3P-dehydrogenase (EC1.1.1.8), 1.8 U/mL triosephosphate isomerase (EC5.3.1.1), 65 µM fructose 2,6-bisphosphate as activator

(synthesized as previously described (*130*)), and 10 mM fructose 6-phosphate as start reagent.

Aldolase (ALDO; EC4.1.2.13) – 0.15 mM NADH, 0.6 U/mL glycerol-3P-dehydrogenase (EC1.1.1.8), 1.8 U/mL triosephosphate isomerase (EC5.3.1.1), and 2 mM fructose 1,6-bisphosphate as start reagent.

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH; EC1.2.1.12) – 0.15 mM NADH, 1 mM ATP, 24 U/mL 3-phosphoglycerate kinase (EC2.7.2.3), and 5 mM 3-phosphoglyceric acid as start reagent.

3-phosphoglycerate kinase (PGK; EC2.7.2.3) – 0.15 mM NADH, 1 mM ATP, 8 U/mL glyceraldehyde-3-phosphate dehydrogenase (EC1.2.1.12), and 5 mM 3-phosphoglyceric acid as start reagent.

Phosphoglycerate mutase (PGAM; EC5.4.2.1) – 0.15 mM NADH, 1 mM ADP, 2.5 mM 2,3-diphospho-glyceric acid, 5 U/mL enolase (EC4.2.1.11), 50 U/mL pyruvate kinase (EC2.7.1.40), 60 U/mL L-lactate dehydrogenase (EC1.1.1.27), and 5 mM 3-phosphoglyceric acid as start reagent.

Enolase (ENO; EC4.2.1.11) – 0.15 mM NADH, 1 mM ADP, 50 U/ml pyruvate kinase (EC2.7.1.40), 15 U/mL L-lactate dehydrogenase (EC1.1.1.27), and 1 mM 2-phosphoglyceric acid as start reagent.

Pyruvate kinase (PK; 2.7.1.40) – 0.15 mM NADH, 1 mM ADP, 1 mM fructose 1,6-bisphosphate, 60 U/mL L-lactate dehydrogenase (EC1.1.1.27) and 2 mM phosphoenolpyruvate as start reagent.

Protein concentrations [E]

Absolute concentrations of glycolytic targets was performed by targeted proteomics (*131*). Isotopically labelled peptides with ^{13}C lysines and arginines were designed for human glucose metabolism and a list of peptides of interest detected in our samples can be found in Supl. Table S7.

Turnover number (k_{cat}) calculations

Turnover numbers were estimated based on the equation $V_{max} = k_{cat} \cdot [E]$. In the human skeletal muscle samples, more than one isoform was detected for certain proteins. In this case the sum of the concentrations of all isoforms was used to estimate the turnover number. Protein concentrations [E] were measured as $pmol \cdot mg\ protein^{-1}$ and V_{max}'s as $\mu mol \cdot min^{-1} \cdot mg\ protein^{-1}$. In order to obtain k_{cat} values in s^{-1}, the following equation was used for each enzymatic reaction in the dataset:

$$k_{cat} = \frac{V_{max}}{[E]} \cdot \frac{10^6}{60}$$

Acknowledgements

We thank Ingeborg van Lakwijk and Agnes Hol for their valuable work on the human hexokinase strains and Lycka Kamoen, Eveline Vreeburg and Daniel Solis-Escalante for their contribution to construction of reference strains. We thank Pilar de la Torre and Marcel van den Broek for whole genome sequencing and analysis, Marijke Luttik for help with the enzyme assays, Philip de Groot for assistance with bioreactor sampling and Jordi Geelhoed for sequencing of the complementation strains.

Supplementary data

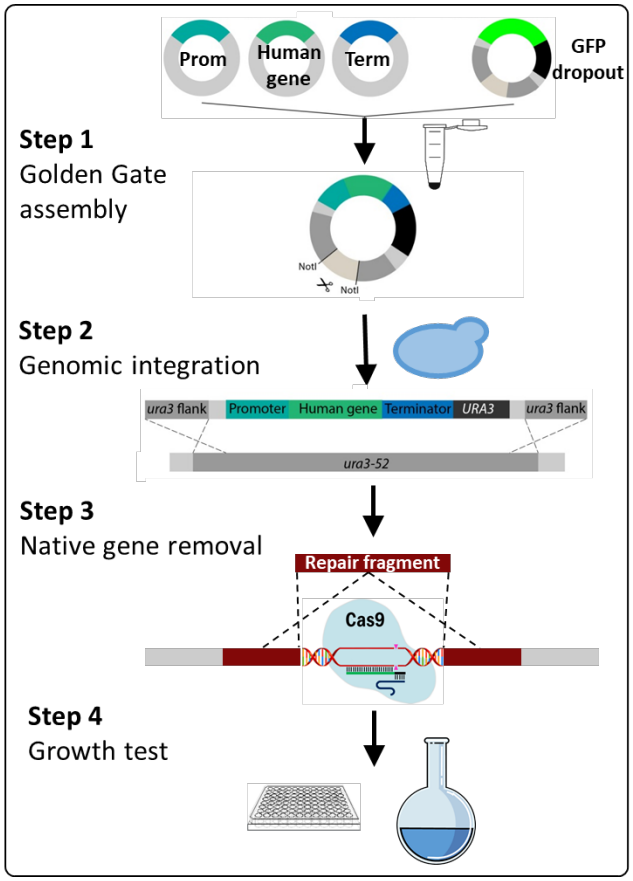


Figure S1 – Complementation strategy.
Codon-optimized human glycolytic genes were stitched to yeast promoters and terminators using Golden Gate assembly (step 1). The resulting plasmids were linearized by restriction with NotI and integrated in the *ura3-52* locus of the minimal glycolysis strain IMX1076 which was plated on SMG (Step 2). In a second transformation round, the yeast ortholog was selectively removed using CRISPR/Cas9-mediated DNA editing. Transformed cells were plated on YPD + G418, except for the *HsHK1-3* strains which were plated on YPGal + G418 (step 3). Strains were confirmed by PCR and Sanger sequencing. All strains were tested for growth on chemically defined medium with glucose as sole carbon source (SMG).

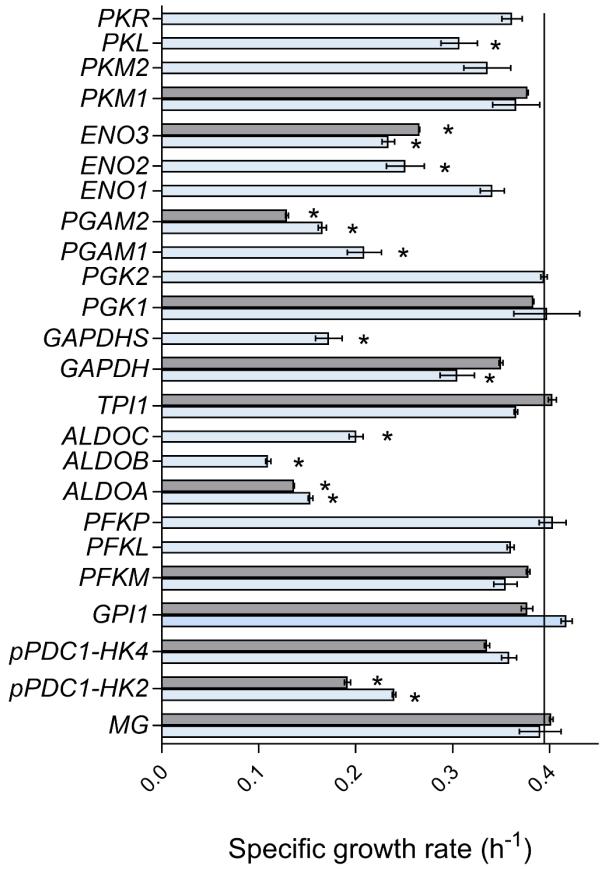


Figure S2 – Growth rate of complementation strains.

Specific growth rates of complementation strains and minimal glycolysis (MG) control strain determined in shake flask (grey) and growth profiler (light blue). * indicates significant difference from control strain ($P<0.01$, Student t-test, two-tailed, homoscedastic). The human genes from strains measured both in shake flask and growth profiler were used for construction of the full human glycolysis strains.

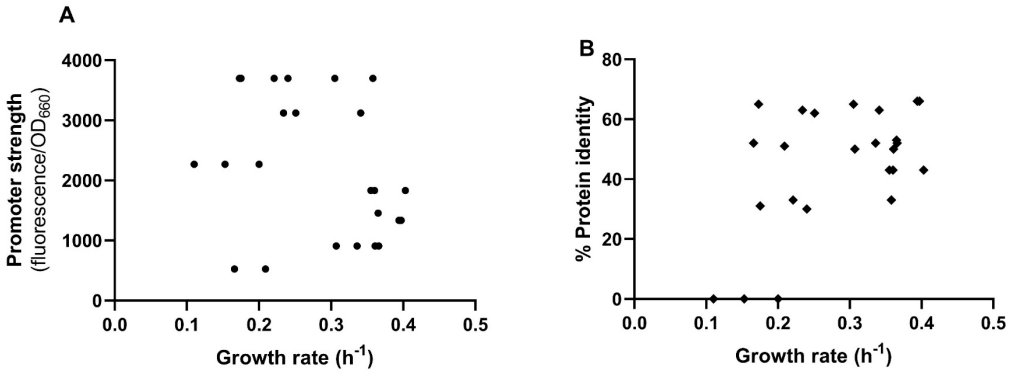


Figure S3 – Comparison of promoter strength, protein identity and growth rate in the complementation strains.

The growth rate of the complementation strains measured in the growth profiler was plotted against: **A)** the strength of the promoter used to express the human gene. Promoter strengths were obtained from Boonekamp *et al.* 2018 *(102)*. **B)** Percentage protein identity between the human gene and the corresponding yeast ortholog.

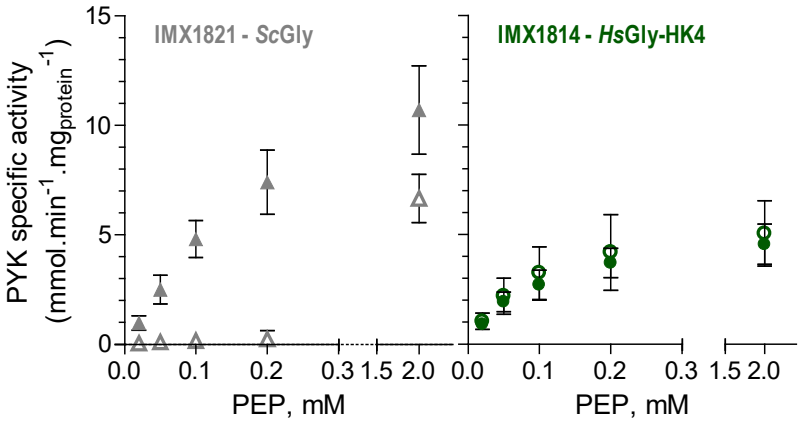


Figure S4 – Sensitivity of human and yeast pyruvate kinase to fructose-1,6bisP.

Yeast (left) and human (right) specific pyruvate kinase activity *in vitro* assayed with fructose-1,6bisP (closed symbols) and without fructose-1,6bisP (open symbols) in IMX1821 (*Sc*gly) and IMX1814 (*Hs*gly) with different concentrations of the substrate phosphoenolpyruvate (PEP). Symbols and error bars represent the average and SEM of three biological replicates.

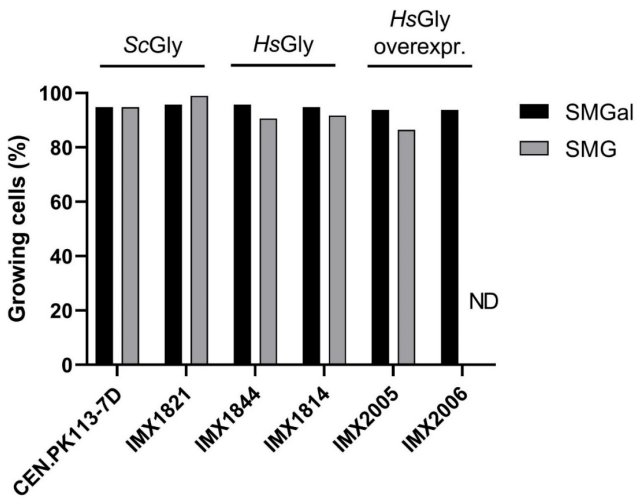


Figure S5 – Robustness to transitioning between carbon sources.

The strains with yeast and humanized glycolysis were cultivated in liquid synthetic medium with galactose as sole carbon source (SMGal), from these culture, single cells were plated on plates with fresh SM medium with glucose (SMG) or galactose as sole carbon source. Per condition 96 single cells were plated. 5 days after plating, the colonies growing on SMGal and SMG were counted. These data originate from a single experiment. ND; not determined.

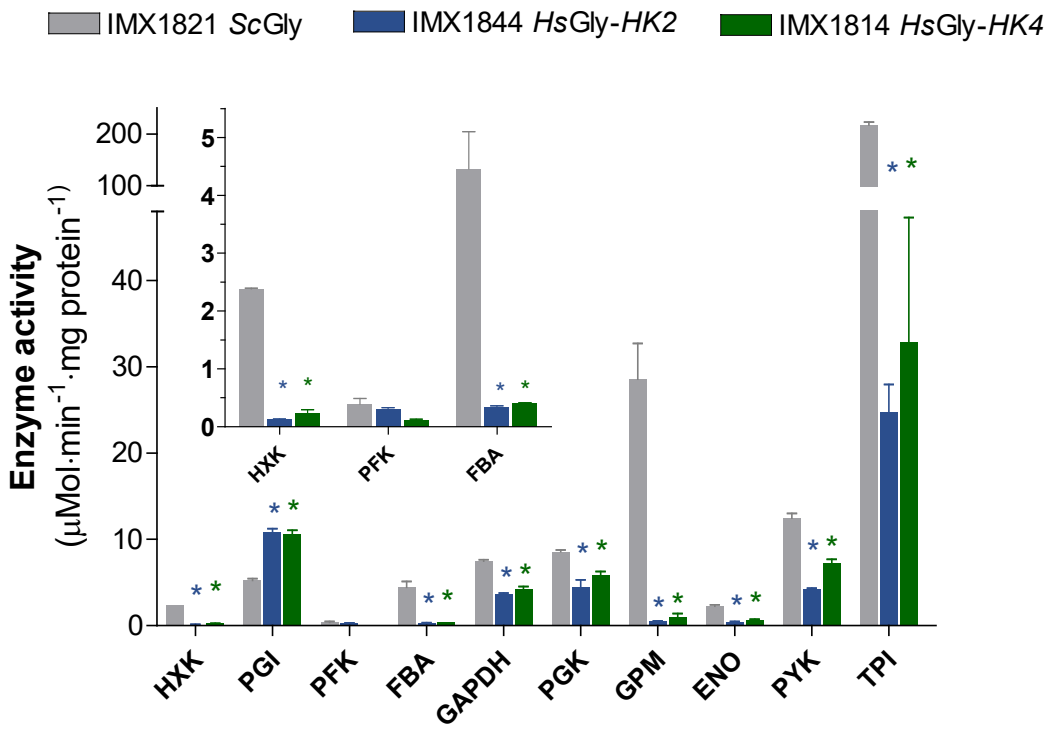


Figure S6 – Glycolytic enzyme activities of full human glycolysis strains .

Specific enzyme activities measured *in vitro* with cell free extracts from batch cultures in bioreactors. Error bars represent the SEM from two biological replicates. The asterisks indicate values that are significantly different from the control strain IMX1821 with *S. cerevisiae* glycolysis (Student t-test, two-tailed, homoscedastic, $P<0.05$).

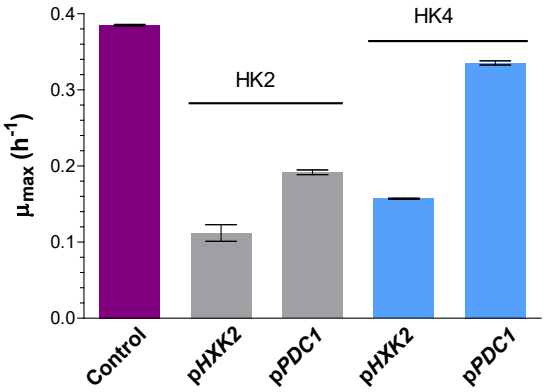


Figure S7 – Impact on growth of *HsHK2* and *HsHK4* expression levels in complementation strains. Growth in aerobic shake-flasks with glucose as carbon source of complementation strain expressing *Hs*HK2 and *HsHK4*. The control strain is MG (IMX372). The human *HsHK2* and *HsHK4* genes were expressed either with the *ScHXK2* promoter (IMX1873, IMX1874) or the stronger *ScPDC1* promoters (IMX1690, IMX1334). The data represent the average specific growth rate and SEM of two independent culture replicates.

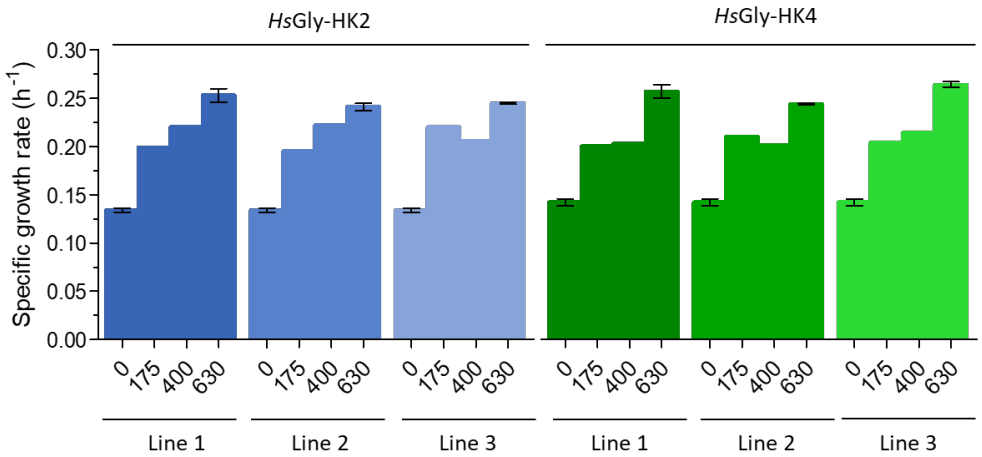


Figure S8 – Specific growth rate during evolution of IMX1844 and IMX1814.
The specific growth rates were measured in shake-flask with SMG. Independent duplicates were only done for 0 and 630 generations. Measurements at 400 and 630 generations are from single colony isolates, while at 175 generations they were done with the whole evolved population.

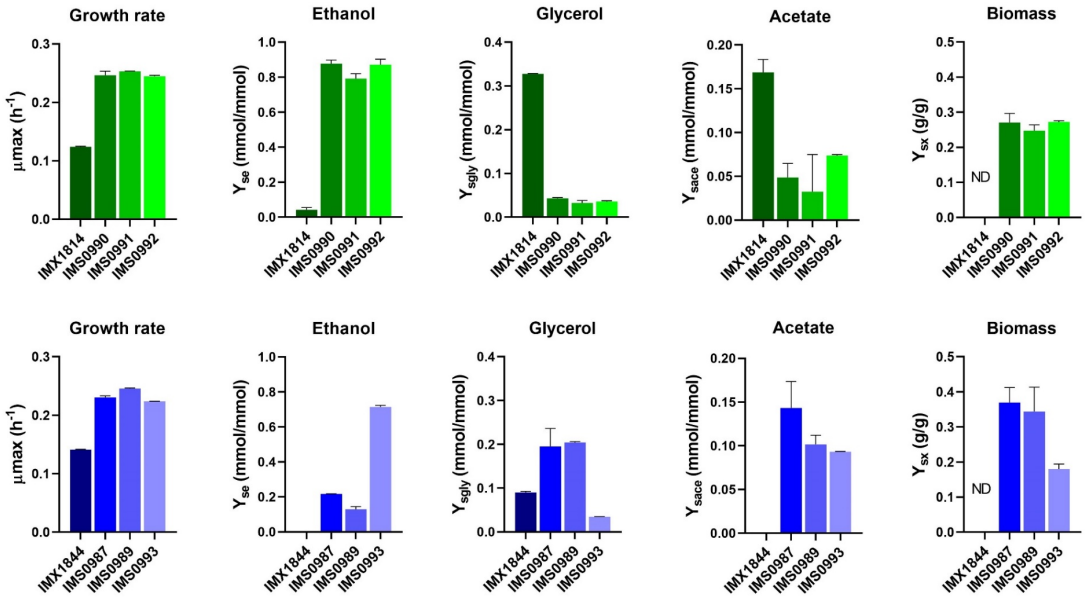


Figure S9 – Physiology of evolved, humanized glycolysis strains in shake flask.
Strain were grown in SMG at 30°C in shake-flask cultures. Data represent the average and SEM of two independent culture replicates. ND: Not determined.

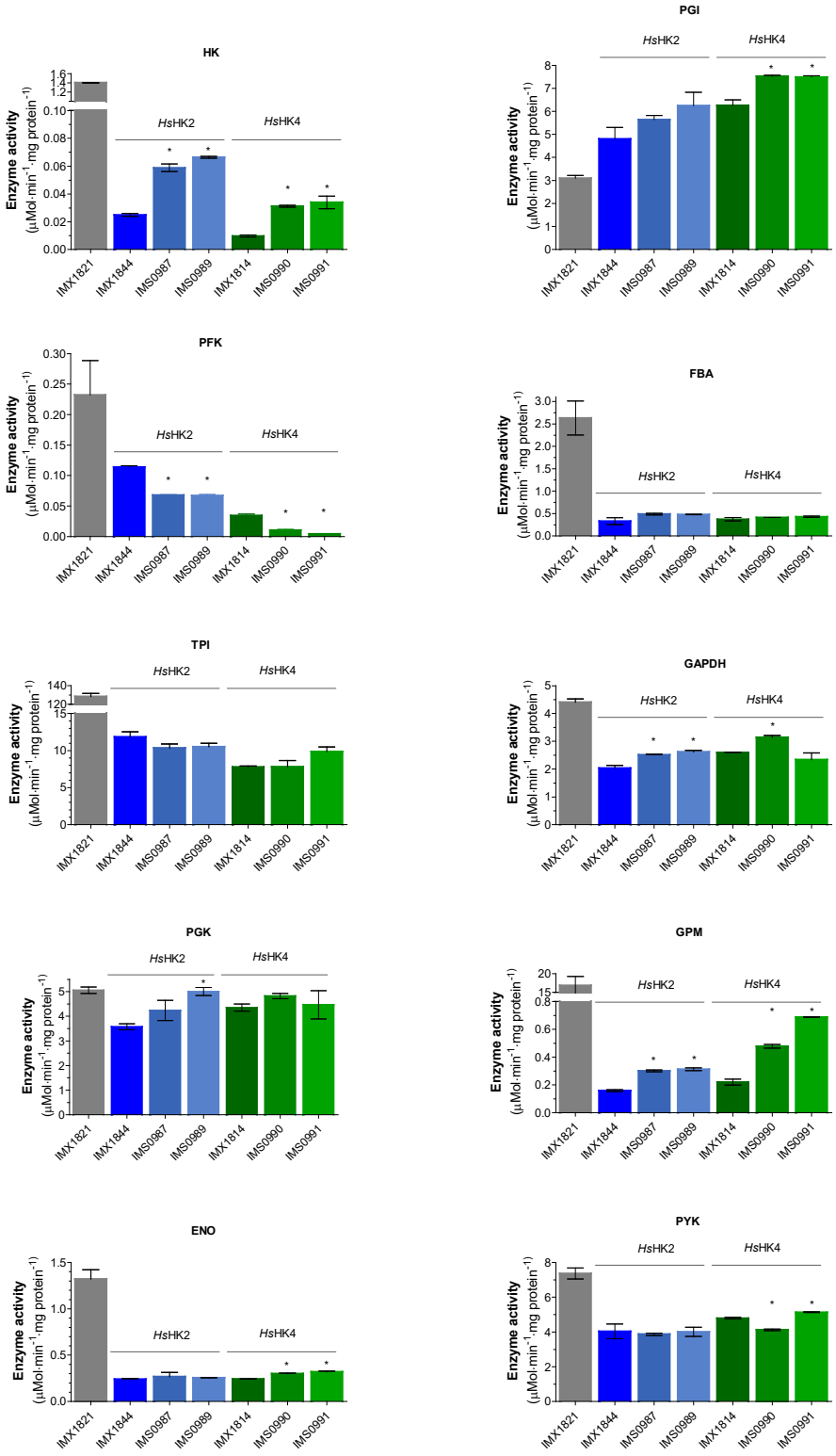


Figure S10 – Specific activity of the glycolytic enzymes in evolved humanized yeast strains.
Shake flask SMG, biological duplicates. Values for IMX1821 extrapolated from cuvette/bioreactor. Stars indicate t-test p-value<0.05.

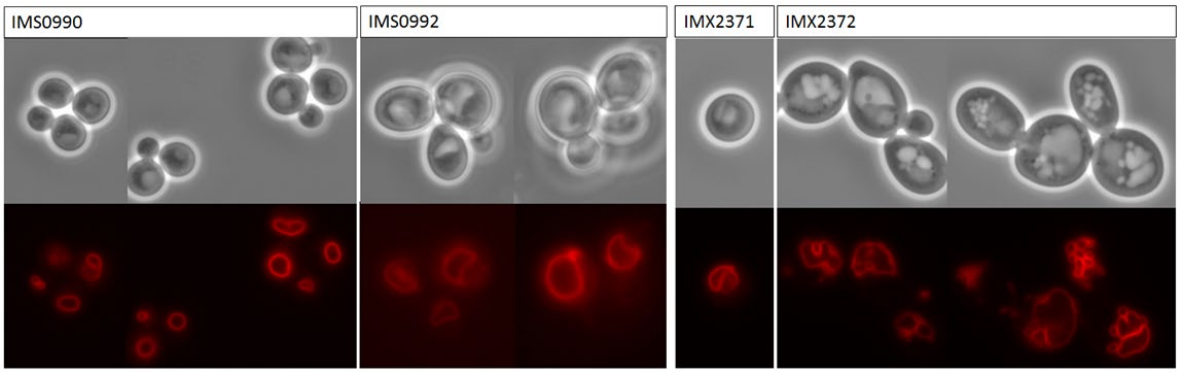


Figure S11 – Vacuole morphology in evolved humanized yeast strains.
Cells were stained with the red fluorescent dye FM4-64 and visualized with an Imager-Z1 microscope (Carl-Zeiss). Strain IMS0990 and IMS0992 are *Hs*Gly-HK4 evolved strains and IMX2371 and IMX2372 are the non-evolved variants in which the *STT4* gene is mutated.

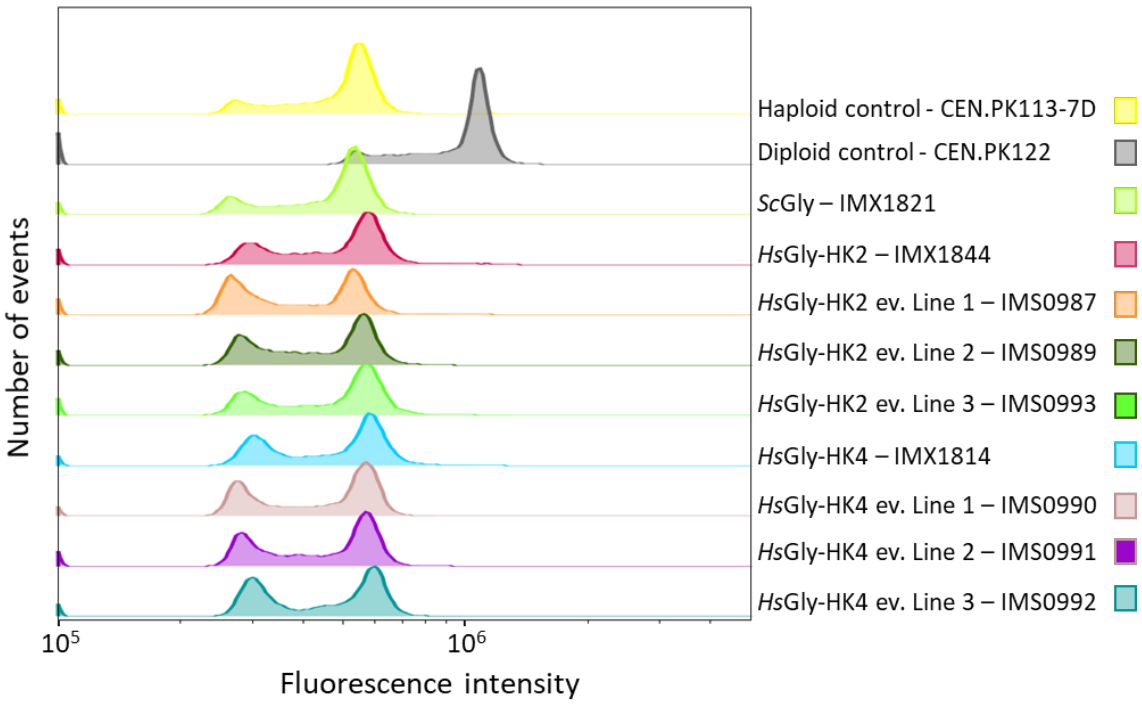
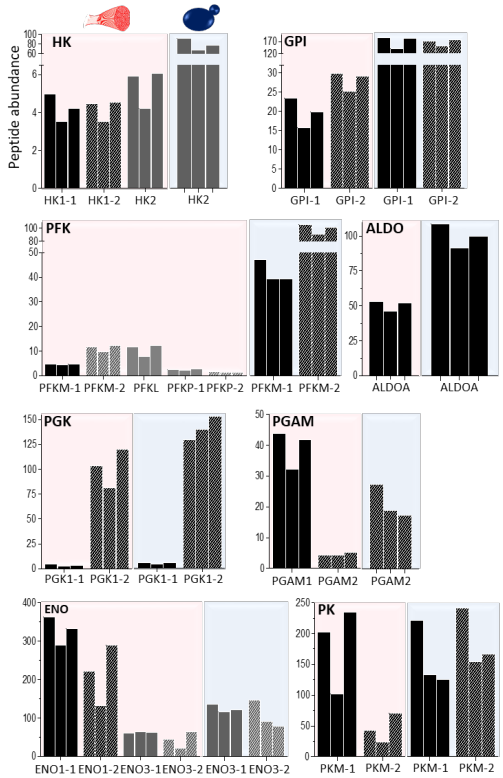
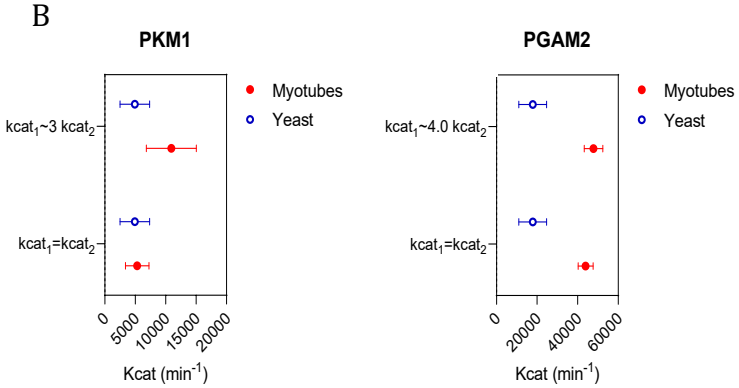


Figure S12 – Ploidy analysis of the strains used in this study.
Ploidy measurement based on DNA content determination by flow cytometry. The two top plots represent the haploid control and the diploid control.

A



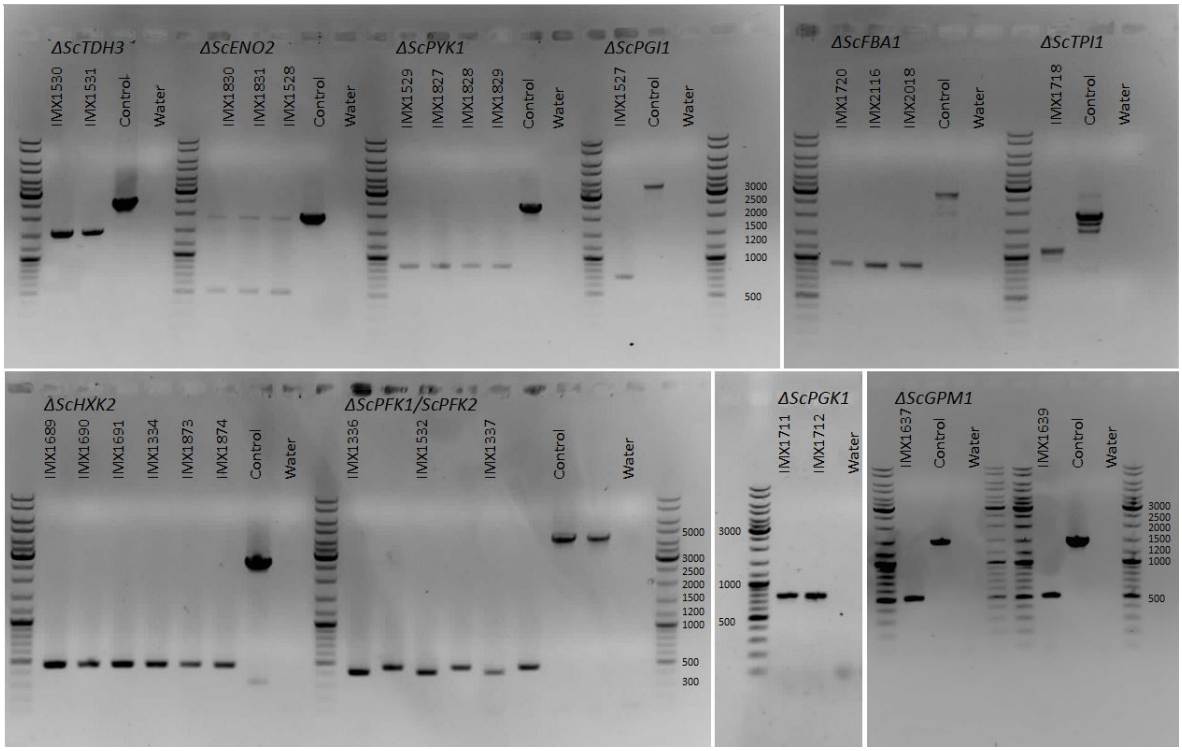
B



$$V_{max} = \sum_{i}^{n} k_{cat_i}[E_i]$$

Figure S13 – Peptide abundance and k_{cat}.

A) Peptide abundance in fmol.$mg_{protein}^{-1}$ for the human glycolytic proteins in yeast (strain *Hs*Gly-HK2, blue background) and myotube (pink background) cell extracts. On the x-axis are represented the peptides that were identified in the proteomics analysis. When several peptides were quantified for a protein they are indicated by -1 or -2 after the name of the protein (see Table S7 for the peptide sequence). For the myotubes, the bars represent the peptide concentration for three independent cultures, for yeast, the two leftmost bars represent the peptide concentration for two independent cultures, while the right bar represent an analytical replicate of the middle bar. No absolute quantification could be made for TPI and GAPDH due to the lack of standard peptides. **B)** Estimated k_{cat} values for PKM1 and PGAM2 based on the provided equation and k_{cat} proportions for isozymes derived from the literature. Data show means ± SD.



Yeast gene	Deletion (bp)	No deletion (bp)
ENO2	520	1839
PYK1	860	2373
TDH3	1530	2534
PGI1	710	3453
PGK1	755	2362
GPM1	512	1481
FBA1	865	2714
HXK2	460	2653
TPI1	1090	1840
PFK1	415	4418
PFK2	375	4302

Figure S14 – Diagnostic PCR of single complementation strains.
PCR confirmation of yeast gene deletions resulting in the individual human gene complementation strains which are indicated per well. The forward and reverse primer were chosen upstream and downstream of the region of gene deletion. The control shows the PCR product resulting from a strain (IMX1076) in which the gene is not deleted. In case of the *ENO2* deletion the primers were binding in the promoter and terminator which resulted in amplification of the human gene as well.

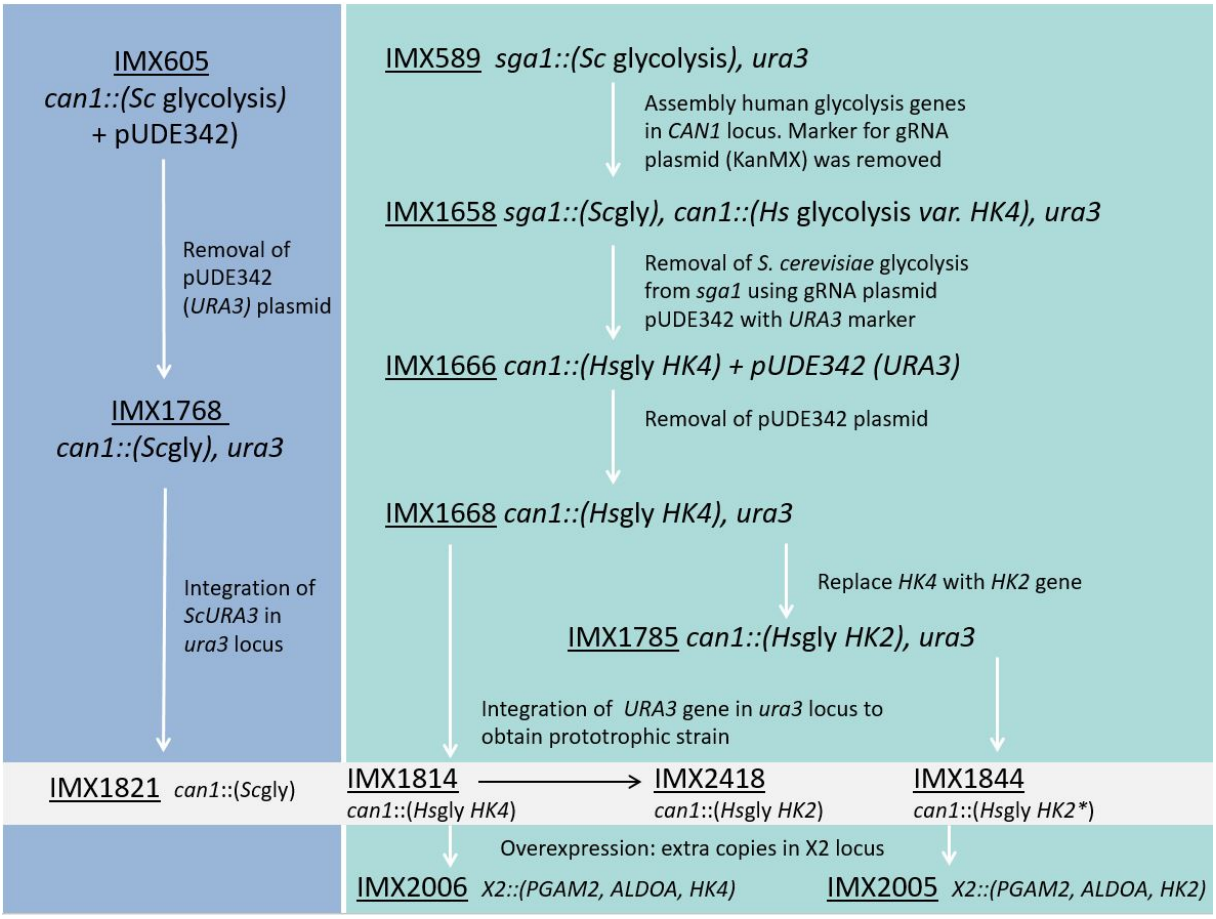


Figure S15 – Overview of single locus glycolysis strain construction.

Table S1 – Comparison of the human glycolytic proteins to their yeast orthologs.

Human enzyme	Size (aa)	Closest yeast enzyme	Size (aa)	% Identity at protein level	Comple-mentation	Shown previously?
HK1	917	HXK2	486	30% to subunit 1, 35% to subunit 2	Yes	No
HK2	917	HXK2	486	33% to both subunits	Yes	No
HK3	923	HXK2	486	28% to subunit 1 33% to subunit 2	No	No
HK4	465	HXK2	486	31%	Yes	Yes
GPI1	558	PGI1	554	58%	Yes	No
PFKM	780	PFK1, PFK2	987/959	43% to PFK1, 43% to PFK2	Yes	Yes
PFKP	784	PFK1, PFK2	987/959	43% to PFK1, 43% to PFK2	Yes	No
PFKL	780	PFK1, PFK2	987/959	43% to PFK1, 45% to PFK2	Yes	No
ALDOA	364	FBA1	359	-	Yes	No
ALDOB	364	FBA1	359	-	Yes	Yes
ALDOC	364	FBA1	359	-	Yes	No
TPI	286	TPI1	248	53%	Yes	Yes
GAPDH	335	TDH3	332	65%	Yes	No
GAPDHS	408	TDH3	332	65%	Yes	No
PGK1	417	PGK1	416	66%	Yes	Yes
PGK2	417	PGK1	416	66%	Yes	Yes
PGAM1	253	GPM1	247	51%	Yes	Yes
PGAM2	253	GPM1	247	52%	Yes	Tested, but negative
ENO1	434	ENO2	437	63%	Yes	No
ENO2	434	ENO2	437	62%	Yes	No
ENO3	434	ENO2	437	63%	Yes	No
PKM1*	531	PYK1	500	52%	Yes	Yes
PKM2*	531	PYK1	500	52%	Yes	No
PKR#	574	PYK1	500	50%	Yes	Yes
PKL#	553	PYK1	500	50%	Yes	Yes

*,# Splicing variants

Table S2 – Genetic composition of the glycolytic transcriptional units used for the single complementation strains and strains with fully humanized glycolysis.

Human gene	Yeast prom	Yeast term	Yeast gene	Yeast prom	Yeast term
HsHK1	*ScPDC1*	*ScPDC1*	*ScHXK2*	*ScHXK2*	*ScHXK2*
HsHK2#	*ScPDC1*	*ScPDC1*			
HsHK3	*ScPDC1*	*ScPDC1*			
HsHK4#	*ScPDC1*	*ScPDC1*			
HsGPI	*ScTEF2*	*ScTEF2*	*ScPGI1*	*ScPGI1*	*ScPGI1*
HsPFKM	*ScTEF1*	*ScTEF1*	*ScPFK1*	*ScPFK1*	*ScPFK1*
HsPFKP	*ScTEF1*	*ScTEF1*	*ScPFK2*	*ScPFK2*	*ScPFK2*
HsPFKL	*ScTEF1*	*ScTEF1*			
HsALDOA	*ScFBA1*	*ScFBA1*	*ScFBA1*	*ScFBA1*	*ScFBA1*
HsALDOB	*ScFBA1*	*ScFBA1*			
HsALDOC	*ScFBA1*	*ScFBA1*			
HsTPI	*ScTPI1*	*ScTPI1*	*ScTPI1*	*ScTPI1*	*ScTPI1*
HsGAPDH	*ScTDH3*	*ScTDH3*	*ScTDH3*	*ScTDH3*	*ScTDH3*
HsGAPDHS	*ScTDH3*	*ScTDH3*			
HsPGK1	*ScPGK1*	*ScPGK1*	*ScPGK1*	*ScPGK1*	*ScPGK1*
HsPGK2	*ScPGK1*	*ScPGK1*			
HsPGAM1	*ScGPM1*	*ScGPM1*	*ScGPM1*	*ScGPM1*	*ScGPM1*
HsPGAM2	*ScGPM1*	*ScGPM1*			
HsENO1	*ScENO2*	*ScENO2*	*ScENO2*	*ScENO2*	*ScENO2*
HsENO2	*ScENO2*	*ScENO2*			
HsENO3	*ScENO2*	*ScENO2*			
HsPKM1	*ScPYK1*	*ScPYK1*	*ScPYK1*	*ScPYK1*	*ScPYK1*
HsPKM2	*ScPYK1*	*ScPYK1*			
HsPKR	*ScPYK1*	*ScPYK1*			
HsPKL	*ScPYK1*	*ScPYK1*			

For hexokinase expression in full human glycolysis strains *HXK2* promoter and terminator was used.

Table S3 – Physiology of humanized glycolysis strains in bioreactors.

Yields and biomass specific conversion rates of IMX1844 (*Hs*Gly-*HK2*), IMX1814 (*Hs*Gly-*HK4*) and the control strain IMX1821 (*Sc*Gly). Strains were grown in bioreactor at 30°C under aerobic conditions in SMG. Yields on glucose for biomass (Y_{sx}), glycerol (Y_{sgly}), ethanol (Y_{setoh}) and acetate (Y_{sace}). Biomass specific rates of glucose and oxygen consumption (q_s and q_{O2}, respectively), of glycerol (q_{gly}), ethanol (q_{etoh}), acetate (q_{ace}) and carbon dioxide production (q_{CO2}) and maximum specific growth rate (μ_{max}). RQ: respiration quotient (q_{CO2}/q_{O2}).Values show the average and standard error of the mean from at least two independent bioreactors per strain. DW: biomass dry weight. BDL: below detection level.

	***Sc*Gly IMX1821**	**HsGly-HK2 IMX1844**	**HsGly-HK4 IMX1814**
Biomass specific rates			
μ_{max} (h^{-1})	0.32 ± 0.01	0.15 ± 0.00	0.15 ± 0.00
$-q_s$ ($mmol.g_{DW}^{-1}.h^{-1}$)	16.1 ± 0.8	2.2 ± 0.0	3.5 ± 0.0
q_{etoh} ($mmol.g_{DW}^{-1}.h^{-1}$)	22.3 ± 0.8	0.03 ± 0.00	0.29 ± 0.04
q_{gly} ($mmol.g_{DW}^{-1}.h^{-1}$)	1.6 ± 0.1	0.11 ± 0.00	0.90 ± 0.05
q_{ace} ($mmol.g_{DW}^{-1}.h^{-1}$)	0.59 ± 0.06	BDL	0.69 ± 0.03
Respiration			
q_{CO2} ($mmol.g_{DW}^{-1}.h^{-1}$)	29.9 ± 1.6	6.8 ± 0.3	7.1 ± 0.3
q_{O2} ($mmol.g_{DW}^{-1}.h^{-1}$)	7.9 ± 0.4	6.2 ± 0.3	6.5 ± 0.7
RQ	3.8 ± 0.01	1.1 ± 0.1	1.1 ± 0.1
Yields			
Y_{sx} (g_{DW}/g_{glu})	0.11 ± 0.01	0.37 ± 0.01	0.23 ± 0.01
Y_{setoh} ($mol_{ethanol}/mol_{glu}$)	1.4 ± 0.0	0.01 ± 0.00	0.08 ± 0.01
Y_{sgly} ($mol_{glycerol}/mol_{glu}$)	0.10 ± 0.00	0.05 ± 0.00	0.25 ± 0.01
Y_{sace} ($mol_{acetate}/mol_{glu}$)	0.04 ± 0.00	BDL	0.19 ± 0.01

Table S4 – Whole genome sequence analysis of the strains with fully humanized glycolysis.

Systematic name	Name	Type	Amino acid change
Mutations in IMX1666 (human glycolysis var. *HK4*) compared to IMX589 (SwYG)			
YDR351W	SBE2	NSY	Ser-412-Thr
YFL064C	uncharacterized protein	NSY	Asn-66-Asp
YHR219W	uncharacterized protein	NSY	Val-383-Leu
YJL223C	PAU1	SYN	Ala-13-Ala
	HsPFKM	SYN	Val-529-Val
-	tPDC1/SHR BC	4 bp gap	Last 2 bp missing and 2 bp from SHR BC
Mutations in IMX1844 (human glycolysis var. HK2) relative to IMX1666			
YJL225C	uncharacterized protein	NSY	Arg-267-Ser
-	HsHK2	NSY	Ile-562-Asn
Mutations in IMX2005 relative to IMX1844			
YEL074W	Putative protein	NSY	His-66-Pro
No mutations in IMX2006 relative to IMX1666			

Table S5 – Mutations in the coding regions of evolved strains.

Text in grey indicates synonymous mutation. The pink background indicates the mutations common to all six evolved strains and the blue background the mutations common to all evolution lines from a single humanized strain.

		IMS0990		IMS0991		IMS0992	
Evolved strains with IMX1814 (HsGly-HK4) background	TUP1	G to A	His-489-Tyr	G to A	His-489-Tyr	G to A	His-489-Tyr
	STT4	C to T	Gly-1766-Arg	C to G	Arg-1707-Pro	A to T	Phe-1775-Ile
	PFKM	A to G	Thr-81-Ala	G to C	Arg-623-Ser	G to T	Arg-673-Ile
	NUT1			T to G	Tyr-432-stop		
	RSM22	G to A	Arg-120-Cys				
	YMR089C	C to A	Pro-217-Thr	G to C	Asp-430-His		
	YCR038W	G to A	Leu-123-Leu				
	YLR371W	A to T	Ala-74-Ala				
	YKL124W			G to A	Thr-346-Thr		
	NMa111	G to A	Glu-40-Glu	G to A	Glu-40-Glu	G to A	Glu-40-Glu

		IMS0987		IMS0989		IMS0993	
Evolved strains with IMX1844 (Hx-Gly-HK2) background	STT4	C to A	Asp-1650-Tyr	A to G	Ile-1771-Thr	G to C	Ser-1611-Cys
	PIN4	A to T	Phe-130-Leu	G to T	Gln-482-stop		
	SNF1					T to G	Phe-261-Cys
	SKI8					G to A	Pro-219-Ser
	TAO3					G to T	His-167-Asn
	CYR1	G to T	Gly-1612-Val	G to A	Gly-1768-Ser		
	ECM38					T to G	Tyr-650-Asp
	YOL075C			G to C	Pro-246-Ala		
	YME1					G to T	STOP-748-Leu
	YOR114W	C to A	Pro-146-Pro				

Table S7 – Peptides standards for absolute quantification by targeted proteome analysis.

Protein	Peptide abbreviation	Identified peptide
HK2	HK2	HK2_HM_LDESFLVSWTK
HK1	HK1-1	HK1_HM_FLLSESGSGK
	HK1-2	HK1_HM_HIDLVEGDEGR
GPI	GPI-1	GPI_HM_TFTTQETITNAETAK
	GPI-2	GPI_HM_VFEGNRPTNSIVFTK
PFKM	PFKM-1	PFKM_HM_LLAHVRPPVSK
	PFKM-2	PFKM_HM_VLVVHDGFEGLAK
PFKL	PFKL	PFKL_HM_GQVQEVGWHDVAGWLGR
PFKP	PFKP-1	PFKP_HM_VTILGHVQR
	PFKP-2	PFKP_M_YLEHLSGDGK
ALDOA	ALDOA	ALDOA_H_ALQASALK
PGK1	PGK1	PGK1_HM_ALESPERPFLAILGGAK
	PGK1	PGK1_HM_ITLPVDFVTADK
PGAM1	PGAM1	PGAM1_HMmc_FSGWYDADLSPAGHEEAK
PGAM2	PGAM2	PGAM2_HM_FCGWFDAELSEK
ENO1	ENO1-1	ENO1_HM_IGAEVYHNLK
	ENO1-2	ENO1_HM_VNQIGSVTESLQACK
ENO3	ENO3-1	ENO3_HM_IGAEVYHHLK
	ENO3-2	ENO3_HM_VNQIGSVTESIQACK
PKM	PKM	PKM_HMalliso_CDENILWLDYK
	PKM	PKM_HMiso2_CLAAALIVLTESGR

Table S8A – Strains with integrated human gene expression cassette (native yeast ortholog still present).

Strain	Genotype	Integration plasmid
IMX1316	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::(pPDC1-HK1(hs)-tPDC1, URA3)*	pUDI133
IMX1317	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::(pPDC1-HK2(hs)-tPDC1, URA3)*	pUDI134
IMX1318	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::(pPDC1-HK3(hs)-tPDC1, URA3)*	pUDI135
IMX1319	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT) , ura3::(pPDC1-HK4(hs)-tPDC1, URA3)*	pUDI136
IMX1838	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::(pHXK2-HK2(hs)-tHXK2, URA3)*	pUDI206
IMX1839	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT) , ura3::(pHXK2-HK4(hs)-tHXK2, URA3)*	pUDI207
IMX1320	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT) , ura3::(pTEF2-GPI1(hs)-tTEF2, URA3)*	pUDI137
IMX1321	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT) , ura3::(pTEF1-PFKM(hs)-tTEF1, URA3)*	pUDI138
IMX1322	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT) , ura3::(pTEF1-PFKP(hs)-tTEF1, URA3)*	pUDI139
IMX1323	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT) , ura3::(pTEF1-PFKL(hs)-tTEF1, URA3)*	pUDI140
IMX1353	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pFBA1-ALDOA(hs)-tFBA1, URA3)*	pUDI141
IMX1354	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pFBA1-ALDOB(hs)-tFBA1, URA3)*	pUDI142
IMX1355	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pFBA1-ALDOC(hs)-tFBA1, URA3)*	pUDI143
IMX1692	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::(pTPI1-TPI1(hs)-tTPI1, URA3)*	pUDI144
IMX1356	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pTDH3-GAPDH(hs)-tTDH3, URA3)*	pUDI145
IMX1300	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pTDH3-GAPDHS(hs)-tTDH3, URA3)*	pUDI146

Strain	Genotype	Integration plasmid
IMX1301	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pPGK1-PGK1(hs)-tPGK1, URA3)*	pUDI147
IMX1302	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pPGK1-PGK2(hs)-tPGK1, URA3)*	pUDI148
IMX1303	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pGPM1-PGAM1(hs)-tGPM1, URA3)*	pUDI149
IMX1304	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pGPM1-PGAM2(hs)-tGPM1, URA3)*	pUDI150
IMX1305	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pENO2-ENO1(hs)-tENO2, URA3)*	pUDI151
IMX1306	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pENO2-ENO2(hs)-tENO2, URA3)*	pUDI152
IMX1307	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pENO2-ENO3(hs)-tENO2, URA3)*	pUDI153
IMX1387	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pPYK1-PKM1(hs)-tPYK1, URA3)*	pUDI154
IMX1388	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pPYK1-PKM2(hs)-tPYK1, URA3)*	pUDI155
IMX1389	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::pPYK1-PKL(hs)-tPYK1, URA3)*	pUDI156
IMX1693	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ura3::(pPGK1-PKR(hs)-tPGK1, URA3)*	pUDI157

Table S8C – Single locus glycolysis strains.

Strain	Genotype	Description/ Modification
IMX605	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(FBA1$_H$TPI1$_P$PGK1$_Q$ADH1$_N$PYK1$_O$ TDH3$_A$ENO2$_B$HXK2$_C$PGI1$_D$ PFK1$_J$PFK2$_K$KanMX$_L$GPM1$_M$PDC1) pUDE342*	Yeast glycolysis in *can1 + pUDE342 (27)*
IMX1768	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(FBA1$_H$TPI1$_P$PGK1$_Q$ADH1$_N$PYK1$_O$ TDH3$_A$ENO2$_B$HXK2$_C$ PGI1$_D$ PFK1$_J$PFK2$_K$KanMX$_L$GPM1$_M$PDC1)*	Yeast glycolysis in *can1*
IMX1821	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1::ScURA3 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(FBA1$_H$TPI1$_P$PGK1$_Q$ADH1$_N$PYK1$_O$ TDH3$_A$ENO2$_B$HXK2$_C$ PGI1$_D$PFK1$_J$PFK2$_K$KanMX$_L$GPM1$_M$PDC1)*	Yeast glycolysis in *can1 + ScURA3* i
IMX589	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1::URA3 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_H$TPI1$_P$PGK1$_Q$ADH1$_N$PYK1$_O$ TDH3$_A$ENO2$_B$ HXK2$_C$PGI1$_D$PFK1$_J$PFK2$_K$AmdSYM$_L$ GPM1$_M$PDC1) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*	Yeast glycolysis in *sga1* (SwYG) (*27*)
IMX1658	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1::URA3 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(ENO2(long promoter) FBA1 PGI1 TPI1 PGK1 PFK1 PFK2 HXK2 TDH3 PGK1 GPM1 AmdSYM PYK1 ADH1) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$HsPFKM$_{BH}$HsPKM1$_{BJ}$HsGAPDH $_{BE}$HsENO3$_{BF}$ScADH1$_{BB}$HsHK4 $_{BC}$ScPDC1$_{BD}$ HsPGAM2$_{BK}$HsTPI1)*	Integration human glycolysis
IMX1666	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ ScADH1$_{BB}$ HsHK4 $_{BC}$ScPDC1$_{BD}$ HsPGAM2$_{BK}$HsTPI1) pUDE342*	Removal yeast glycolysis
IMX1668	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ ScADH1$_{BB}$HsHK4 $_{BC}$ScPDC1$_{BD}$ HsPGAM2$_{BK}$HsTPI1)*	Removal plasmid
IMX1785	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ ScADH1$_{BB}$HsHK2*$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1)*	Replace *HK4* with *HK2*

Strain	Genotype	Modification
IMX1814	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1::URA3 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$ HsENO3$_{BF}$ScADH1$_{BB}$HsHK4$_{BC}$ScPDC1$_{BD}$ HsPGAM2$_{BK}$HsTPI1)*	Integration *ScURA3* in *tdh1*
IMX1844	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1::URA3 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ScADH1$_{BB}$HsHK2*$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1)*	Integration *ScURA3* in *tdh1*
IMX2418	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5, hxk1::LEU2, tdh1:URA3sc, tdh2::AB, gpm2::LoxP, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1, pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh, pdc1, eno2, can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ScADH1$_{BB}$HsHK2$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1)*	Replace *HK4* from IMX1814 with *HK2*
IMX2005	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1::URA3 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$ HsENO3$_{BF}$ScADH1$_{BB}$HsHK2*$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1), X2::(ALDOA$_{BA}$HK2$_{BD}$PGAM2)*	Integration extra copies in X2 locus
IMX2006	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5 hxk1::LEU2 tdh1::URA3 tdh2 gpm2::LoxP gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1 pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2 can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$ HsENO3$_{BF}$ScADH1$_{BB}$HsHK4$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1), X2::(ALDOA$_{BA}$HK4$_{BD}$PGAM2)*	Integration extra copies in X2 locus

Table S8D – Evolved strains and reverse engineered *STT4* strains

Strain	Genotype	Description
IMS0987	Single colony isolate of evolved IMX1844 (Hsgly var. HK2)	
IMS0989	Single colony isolate of evolved IMX1844 (Hsgly var. HK2)	
IMS0993	Single colony isolate of evolved IMX1844 (Hsgly var. HK2)	
IMS0990	Single colony isolate of evolved IMX1814 (Hsgly var. HK4)	
IMS0991	Single colony isolate of evolved IMX1814 (Hsgly var. HK4)	
IMS0992	Single colony isolate of evolved IMX1814 (Hsgly var. HK4)	
IMX2369	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1:URA3 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{H}$TPI1$_{P}$ PGK1$_{Q}$ ADH1$_{N}$ PYK1$_{O}$ TDH3$_{A}$ ENO2$_{B}$ HXK2$_{C}$ PGI1$_{D}$ PFK1J PFK2$_{JK}$ GPM1$_{L}$ PDC1) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2, STT4 (G1766R)*	Host strain: IMX1822
IMX2370	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1:URA3 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_{H}$TPI1$_{P}$ PGK1$_{Q}$ ADH1$_{N}$ PYK1$_{O}$ TDH3$_{A}$ ENO2$_{B}$ HXK2$_{C}$ PGI1$_{D}$ PFK1J PFK2$_{JK}$ GPM1$_{L}$ PDC1) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2, STT4 (F1775I)*	Host strain: IMX1822
IMX2371	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5, hxk1::LEU2, tdh1, tdh2::AB, gpm2::LoxP, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1, pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh, pdc1, eno2, can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ ScADH1$_{BB}$HsHK4$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1), STT4 (G1766R)*	Host strain: IMX1814
IMX2372	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5, hxk1::LEU2, tdh1, tdh2::AB, gpm2::LoxP, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1, pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh, pdc1, eno2, can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ScADH1$_{BB}$ HsHK4$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1), STT4 (F1775I)*	Host strain: IMX1814
IMX2373	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5, hxk1::LEU2, tdh1:URA3sc, tdh2::AB, gpm2::LoxP, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1, pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh, pdc1, eno2, can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ ScADH1$_{BB}$HsHK2*$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1), STT4 (G1766R)*	Host strain: IMX1844
IMX2374	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::HIS5, hxk1::LEU2, tdh1:URA3sc, tdh2::AB, gpm2::LoxP, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1, pyk1 pgi1 tpi1 tdh3 pfk2::(pTEFCAS9 nat) pgk1 gpm1 fba1 hxk2 pfk1 adh, pdc1, eno2, can1::(HsALDOA$_{BA}$HsGPI$_{BI}$HsPGK1$_{BG}$ HsPFKM$_{BH}$ HsPKM1$_{BJ}$ HsGAPDH$_{BE}$HsENO3$_{BF}$ ScADH1$_{BB}$HsHK2*$_{BC}$ScPDC1$_{BD}$HsPGAM2$_{BK}$HsTPI1), STT4 (F1775I)*	Host strain: IMX1844

Table S8E – Control strains and intermediate strains.

Strain	Genotype	Source
CEN.PK113-7D	*MATa URA3 HIS3 LEU2 TRP1 MAL2-8c SUC2*	(*49, 96, 132, 133*)
CEN.PK122	*MATa/Matα*	(*49, 96*)
CEN.PK102-12A	*MATa ura3-52 his3-D1 leu2-3,112 TRP1 MAL2-8c SUC2*	P. Kötter
IMX370	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4*	(*26*)
IMX372	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHIS5, hxk1::KlLEU2, tdh1::URA3, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4*	(*26*)
IMX1076	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT)*	This study
IMX589	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_H$TPI1$_P$PGK1$_Q$ADH1$_N$PYK1$_O$ TDH3$_A$ENO2$_B$HXK2$_C$PGI1$_D$ PFK1$_J$PFK2$_K$AmdSYM$_L$ GPM1$_M$PDC1) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*	(*27*)
IMX1769	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_H$TPI1$_P$ PGK1$_Q$ ADH1$_N$ PYK1$_O$ TDH3$_A$ ENO2$_B$ HXK2$_C$ PGI1$_D$ PFK1J PFK2$_{JK}$ GPM1$_L$ PDC1) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*	This study
IMX1822	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::Sphis5 hxk1::KlLEU2 tdh1::ScURA3 tdh2 gpm2 gpm3 eno1 pyk2 pdc5 pdc6 adh2 adh5 adh4 sga1::(FBA1$_H$TPI1$_P$ PGK1$_Q$ ADH1$_N$ PYK1$_O$ TDH3$_A$ ENO2$_B$ HXK2$_C$ PGI1$_D$ PFK1J PFK2$_{JK}$ GPM1$_L$ PDC1) pyk1 pgi1 tpi1 tdh3 pfk2::(pTEF-cas9-tCYC1 natNT1) pgk1 gpm1 fba1 hxk2 pfk1 adh1 pdc1 eno2*	This study
IMX075	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::LEU2*	This study
IMS0336	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::LoxP*	This study
IMX165	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 hxk1::LoxP hxk2::KanMX*	This study
IMX2014	*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ScPDC1p-HXK2*	This study
IMX2015	*MATa URA3, his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4, sga1::(CAS9, NatNT), ScPDC1p-HXK2*	This study

Table S9A – Primers amplification of yeast promoters and terminators with yeast toolkit flanks.

Name	Sequence
9755 pPDC1 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGCATGCGACTGGGTGAGCATATG
9756 pPDC1 rev Ytk	TTATGCCGTCTCAGGTCTCACATATTTGATTGATTTGACTGTGTTATTTTGCG
10755 pTEF1 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGCGGGATAATTAAGACGACAAGAAG
10756 pTEF1 rv Ytk	TTATGCCGTCTCAGGTCTCACATATTTGTAATTAAAACTTAGATTAGATTGCTATG
9419 pFBA1 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGCAATACCAGCCTTCCAACTTC
9420 pFBA1 rv Ytk	TTATGCCGTCTCAGGTCTCACATATTTGAATATGTATTACTTGGTTATGG
9423 pTPI1 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGACCCAGAGATGTTGTTGTCC
9424 pTPI1 rv Ytk	TTATGCCGTCTCAGGTCTCACATATTTTAGTTTATGTATGTGTTTTTTGTAG
10753 pTDH3 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGCGAATATATACTAGCGTTGAATGTTAG
10754 pTDH3 rv Ytk	TTATGCCGTCTCAGGTCTCACATATTTGTTTGTTTATGTGTGTTTATTCG
9421 pPGK1 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGTATTTTAGATTCCTGACTTCAACTC
9422 pPGK1 rv Ytk	TTATGCCGTCTCAGGTCTCACATATGTTTTATATTTGTTGTAAAAAGTAGATAATTAC
9757 pGPM1 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGGTGATACTTTGACAGGAGC
9758 pGPM1 rv Ytk	TTATGCCGTCTCAGGTCTCACATATATTGTAATATGTGTGTTTGTTTGG
9739 pENO2 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGGGATGATGAAAACACTAAACGAAG
9740 pENO2 rv Ytk	TTATGCCGTCTCAGGTCTCACATATATTATTGTATGTTATAGTATTAGTTGCTTGG
10608 pPYK1 fw Ytk	AAGCATCGTCTCATCGGTCTCAAACGCCCTGGTCAAACTTCAGAAC
10609 pPYK1 rv Ytk	TTATGCCGTCTCAGGTCTCACATATGTGATGATGTTTTATTTGTTTTGATTG
10773 tPDC1 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCGCGATTTAATCTCTAATTATTAGTTAAAG
10774 tPDC1 rv Ytk	TTATGCCGTCTCAGGTCTCACAGCCAGTGTTCCTTAATCAAGGATACC
10884 tTEF2 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCGAGTAATAATTATTGCTTCCATATAATATTTTTATATAC
10885 tTEF2 fw Ytk	TTATGCCGTCTCAGGTCTCACAGCAGGAAACGTAAATTACAAGGTATATAC
10767 tTEF1 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCGGAGATTGATAAGACTTTTCTAGTTG
10768 tTEF1 rv Ytk	TTATGCCGTCTCAGGTCTCACAGCGGTATCACCATAGATTTCGAAAC
10757 tFBA1 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCGTTAATTCAAATTAATTGATATAGTTTTTTAATG
10758 tFBA1 rv Ytk	TTATGCCGTCTCAGGTCTCACAGCCGCGAACTCCAAAATGAGC
10765 tTPI1 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCGATTAATATAATTATATAAAAATATTATCTTCTTTTC
10766 tTPI1 rv Ytk	TTATGCCGTCTCAGGTCTCACAGCCGGTACACTTCTGAGTAAC
10761 tTDH3 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCGTGAATTTACTTTAAATCTTGCATTTAAATAAATTTTC
10762 tTDH3 rv Ytk	TTATGCCGTCTCAGGTCTCACAGCGTAACTTCAGAATCGTTATCCTGG
10763 tPGK1 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCATTGAATTGAATTGAAATCGATAG
10764 tPGK1 rv Ytk	TTATGCCGTCTCAGGTCTCACAGCCGAAATAATATCCTTCTCGAAAG
10759 tGPM1 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCGTCTGAAGAATGAATGATTTGATG
10760 tGPM1 rv Ytk	TTATGCCGTCTCAGGTCTCACAGCCATTAAACTACGATGTAAACATC
10886 tPYK1 fw Ytk	AAGCATCGTCTCATCGGTCTCAATCCAAAAAGAATCATGATTGAATGAAGATATT
10887 tPYK1 rev Ytk	TTATGCCGTCTCAGGTCTCACAGCGTATCCTTTCGCCATCCTG

Table S9B – Diagnostic primers.

Name	Sequence
Primers to confirm integration human genes	
9442 URA3 5' barcode Ytkit	GTAATGTTATCCATGTGGGC
7653_ URA3 upstream	ATTCCAACTAATGAGATGGAATCG
4728_ URA3 downstream	CCAGCCCATATCCAACTTCC
9441 URA3 3' barcode Ytkit	AGAGCACTTGAATCCACTGC
Primers to confirm yeast gene deletion	
2798 ENO2 fw	TGAAGTGTGATACCAAGTCAGC
5237 ENO2 rev	AATAGACAGCACGAGTCTTTG
1152 PYK1 fw	TGGCGTGTGATGTCTGTATCTG
4667 PYK1 rev	CCTTGAGGGAAGATTATCTTGCG
5134 TDH3 fw	CCAAAAATAGCCGAGCAAGCTC
4788 TDH3 rev	AACGCTAAGAGTAACTTCAGAATCG
7414 PGI1 fw	AATGTAGCGACACCACTTCC
5004 PGI1 rev	GTAGATTGCACCATCTGAAGAGGC
10590 PGK1 fw	ACTGTAATTGCTTTTAGTTGTG
4698 PGK1 rev	TACGCTGAACCCGAACATAG
12330 GPM1 fw	GCAGACGACAGATCTAAATGAC
12331 GPM1 rev	GCCACCGTACATTTAATATGTC
11067 FBA1 fw	AACTACACGGAAGCTCTAAAGATG
5024 FBA1 rev	CCCTCTTATTTATTAGCATTGTCTTCCG
3481 HXK2 fw	GCCTAGCGTCTGGGATTTATTC
3070 HXK2 rev	AGTGCTTCCGTTCGTTCCAG
3514 TPI1 fw	CTGACAGGTGGTTTGTTACG
8726 TPI1 rev	TCAGCCATTGAGCAGAGAAC
4925 PFK1 fw	AATTTTTACCCTGATCTAACTAACTTTGG
4924 PFK1 rev	GTAGACCGATGACAATACGACTAC
4777 PFK2 fw	CGTGAGCCTTAACCAATGAG
4776 PFK2 rev	CTCCGTTCTTCGTGATAAGTTC

Table S9C – Primers to amplify glycoblocks.

Fragment	Name	Sequence
HsALDOA$_{can1/BA}$	12952 tFBA1 + can1	GTTTTTAATCTGTCGTCAATCGAAAGTTTATTTCAGAGTTCTTCAG ACTTCTTAACTCCTGTGCATGACAAAAGATGAGCTAGG
	12446 pFBA+ BA	TAAGTCTCTTGACATCTCGGAACATATCCACTCAGCGGTGTATCAT TCTGTGGTCGGCGCCATGCCTCCAACGGCTACTATC
HsGPI$_{BA/BI}$	12447 pTEF2 + BA	GCGCCGACCACAGAATGATACACCGCTGAGTGGATATGTTCCGAGA TGTCAAGAGACTTAAACGTTGATAGGTCAAGATCAATG
	12448 tTEF2 + BI	TCTGTCAGTTGGTTAAGCGCCGCTACGATTACTACACATGCCACAG ACTGATCTACAATGAATTACAAGGTATATACATACGCTGACATG
HsPGK1$_{BI/BG}$	12474 tPGK1 + BI	CATTGTAGATCAGTCTGTGGCATGTGTAGTAATCGTAGCGGCGCTT AACCAACTGACAGATGGCAGCCGAAATAATATCCTTC
	12475 pPGK1 + BG	GAGGCTTCACAGTGCTTTATTAGTATGATTGCCTAGCTGGTATATG TGTTCCTGGAGCGCTTCCTGACTTCAACTCAAGACGC
HsPFKM$_{BG/BH}$	12453 pTEF1 + BG	GCGCTCCAGGAACACATATACCAGCTAGGCAATCATACTAATAAAG CACTGTGAAGCCTCCGGGATAATTAAGACGACAAGAAG

	12454 tTEF1 + BH	AGGATCGCTCGCGTACTCATGCATTCTCCCACATATTGAGGCCCTG ATTCCATGCAATGTGGCAGCGGTATCACCATAG
HsPKM1$_{BH/BJ}$	12455 pPYK1 + BH	ACATTGCATGGAATCAGGGCCTCAATATGTGGGAGAATGCATGAG TACGCGAGCGATCCTCCTGGTCAAACTTCAGAACTAAG
	12456 tPYK1 + BJ	GGCGCACATGGTATATTATGATCGGAGATGCGGCAACATAGCTGG GTGTGATCCTCTCTACGTATCCTTTCGCCATCCTG
HsGAPDH$_{BJ/BE}$	12457 tTDH3 + BJ	TAGAGAGGATCACACCCAGCTATGTTGCCGCATCTCCGATCATAAT ATACCATGTGCGCCCAGAATCGTTATCCTGGCGG
	12458 pTDH3 + BE	TCAATCATTCGTTCTCGCAGATCTACAATCGTCCTGAGCTCTGTGA GTGATGTACGCTCCTACTAGCGTTGAATGTTAGCGTC
HsENO3$_{BE/BF}$	12459 pENO2 + BE	GGAGCGTACATCACTCACAGAGCTCAGGACGATTGTAGATCTGCGA GAACGAATGATTGATGATGAAAACACTAAACGAAGG
	12460 tENO2 + BF	GCGCGACGTGTCTCGTATATTAGTGAAGTTGGATCTGTCCATGAAT CCTCGGCTCTGGTGTATTTTTCAAACTGCAAATTCAAG
ScADH1$_{BF/BB}$	12461 tADH1 + BF	CACCAGAGCCGAGGATTCATGGACAGATCCAACTTCACTAATATAC GAGACACGTCGCGCATGCCGGTAGAGGTGTGGTC
	12462 pADH1 + BB	GCAACGCATTCCATACATGATGCGTTGCTTGGTGTCCACAGCCGTA CTTGAGAAGCTCTGAGTCCAATGCTAGTAGAGAAGGG
HsHK4$_{BB/BC}$	12463 pHXK2 + BB	CAGAGCTTCTCAAGTACGGCTGTGGACACCAAGCAACGCATCATGT ATGGAATGCGTTGCGCTGGTAAAGTACAGCTACATTC
	12464 tHXK2 + BC	CTAGGCTCTGCTGCATGTCAGTGATTTCTATTAGGCAGCGCTTACC CATGATTAGCGCAGACTTGAACAATAAATACGAAATCC
ScPDC1$_{BC/BD}$	12465 tPDC1 + BC	CTGCGCTAATCATGGGTAAGCGCTGCCTAATAGAAATCACTGACAT GCAGCAGAGCCTAGTGTTCCTTAATCAAGGATACCTC
	12466 pPDC1 + BD	AGTCACGCTGAGTCCATGCTGACCATGATTCACACTCAGTGCCGAT AATTCCATAGTCTGCGACTGGGTGAGCATATGTTC
HsPGAM2$_{BD/BK}$	12467 pGPM1 + BD	CAGACTATGGAATTATCGGCACTGAGTGTGAATCATGGTCAGCATG GACTCAGCGTGACTGATACTTTGACAGGAGCTATATC
	12468 tGPM1 + BK	GAGCATACTGTCCTATCATGTCGACTCTTGTCACATCTGACGCCTC TCTGCGATAGGATTTGCTATAACATGTCATGTCACC
HsTPI$_{BK/can1}$	12469 tTPI1 + BK	AATCCTATCGCAGAGAGGCGTCAGATGTGACAAGAGTCGACATGA TAGGACAGTATGCTCTGAGTAACCCATATAGAGATCG
	12470 pTPI + can1	GTGTATGACTTATGAGGGTGAGAATGCGAAATGGCGTGGAAATGT GATCAAAGGTAATAACCAGAGATGTTGTTGTCCTAG
HsHK2	13506 HK2 + pHXK2 flank	CTTTGAAAAGATTGTAGGAATATAATTCTCCACACATAATAAGTA CGTTAATTAAATAAAATGATCGCCTCTCATTTGTTG
	13507 HK2 + tHXK2 flank	GTTCACATAATTAAAAAAAGGGCACCTTCTTGTTGTTCAAACTTAA TTTACAAATTAAGTTCATCTTTGACCAGCTTCTCT
Overexpression		
HsALDOA$_{X2\ flank/BA}$	12446 pFBA1+ BA	TAAGTCTCTTGACATCTCGGAACATATCCACTCAGCGGTGTATCAT TCTGTGGTCGGCGCCATGCCTCCAACGGCTACTATC
	12650 tFBA1 + X2 flank	GCTGAAGATTTATCATACTATTCCTCCGCTCGTTTCTTTTTTCAGT GAGGTGTGTCGTGAGTGCATGACAAAAGATGAGCTAGG
HsHK2 and *HsHK4* $_{BA/BD}$	14540 pPDC1 + BA	GCGCCGACCACAGAATGATACACCGCTGAGTGGATATGTTCCGAGA TGTCAAGAGACTTAGCGACTGGGTGAGCATATG
	14541 tPDC1 + BD	AGTCACGCTGAGTCCATGCTGACCATGATTCACACTCAGTGCCGAT AATTCCATAGTCTGCTCATTGGCAGCCAGTGTTC
HsPGAM2$_{BD/X2\ flank}$	12467 pGPM1 + BD	CAGACTATGGAATTATCGGCACTGAGTGTGAATCATGGTCAGCATG GACTCAGCGTGACTGATACTTTGACAGGAGCTATATC
	14542_tGPM1 + X2 flank	ATTCTCGCCAAGGCATTACCATCCCATGTAAGAACGGAATAAAACA GCATTCGAAGGTTATTGCTATAACATGTCATGTCACC

Table S9D – gRNA oligos and primers for backbone amplification.

Target	Name	Sequence
ScHXK2	10205 HXK2 gRNA fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCGGTAAGTCCGTTGGTATCATGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	10206 HXK2 gRNA rv	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACATGATACCAACGGACTTACCGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScPGI1	10080 PGI1 gRNA fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCCAAAAAATTTATGAATCTCAGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	10081 PGI1 gRNA rv	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACTGAGATTCATAAATTTTTTGGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScPFK1	10207 PFK1 gRNA	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCCATCATCTCTGAAGCAAGCAGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAG
ScPFK2	10208 PFK2 gRNA	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCCCAGGTCATGTACAACAAGGGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAG
ScFBA1	12332 FBA1 gRNA	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCCAAATCTTAAAGAGAAAGACGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAG
ScTPI1	10972 TPI1 gRNA fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCCTTAGACTACTCTGTCTCTTGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	10973 TPI1 gRNA rev	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACAAGAGACAGAGTAGTCTAAGGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScTDH3	10968 TDH3 gRNA fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCTTTGAGTAGCAGTCAAAGAGGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	10969 TDH3 gRNA rev	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACCTCTTTGACTGCTACTCAAAGATCATTTATCTTTC ACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScPGK1	10970 PGK1 gRNA fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCCAGACACGAATTGAGCTCTTGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	10971 PGK1 gRNA rev	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACAAGAGCTCAATTCGTGTCTGGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScGPM1	10976 GPM1 gRNA fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCATTGCCAAGGACTTGTTGAGGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	10977 GPM1 gRNA fw	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACCTCAACAAGTCCTTGGCAATGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScENO2	10076 ENO2 gRNA fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCCAAGGCCAACCTAGATGTTAGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	10077 ENO2 gRNA rv	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACTAACATCTAGGTTGGCCTTGGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScPYK1	10974 PYK1 gRNA fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCTATCAACTTCGGTATTGAAAGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	10975 PYK1 gRNA rev	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACTTTCAATACCGAAGTTGATAGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA

HsHK4	13696 gRNA HsHK4II fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCATGTGTTCTGCTGGTTTGGCGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	13697 gRNA HsHK4II rev	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACGCCAAACCAGCAGAACACATGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScSTT4 (I)	16748 gRNA STT4 (IMS0990) fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCAACATTATGTACGATGATCAGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	16749 gRNA STT4 (MS0990) rev	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACTGATCATCGTACATAATGTTGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
ScSTT4 (II)	16755 gRNA STT4 (IMS0992) fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCATTGTCTACATATCGATTTTGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	16756 gRNA STT4 (IMS0992) rev	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACAAAATCGATATGTAGACAATGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
AmdS	11588 gRNA AmdS fw	TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGAT AAATGATCATCACATCCGAACATAAACAGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
	11589 gRNA AmdS rev	GTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACTGTTTATGTTCGGATGTGATGATCATTTATCTTT CACTGCGGAGAAGTTTCGAACGCCGAAACATGCGCA
pROS13, pMel13 and pMel10 backbone amplification	6005_p426 CRISP rv	GATCATTTATCTTTCACTGCGGAGAAG
	6006_p426 CRISP fw	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC

Table S9E – Repair fragments.

Name	Sequence
5888 HXK2 repair oligo fw	TTTCTAATGCCTTTTCCATCATGTTACTACGAGTTTTCTGAACCTCCTCGCACATT GGTAGCTTAATTTTAAATTTTTTTGGTAGTAAAAGATGCTTATATAAGGATTTCG TATTTATTG
5889 HXK2 repair oligo rv	CAATAAATACGAAATCCTTATATAAGCATCTTTTACTACCAAAAAAATTTAAAAT TAAGCTACCAATGTGCGAGGAGGTTCAGAAAACTCGTAGTAACATGATGGAAAAG GCATTAGAAA
10084 PGI1 repair oligo fw	ATACACCGCTATGTATTTCAGGGCACTACTTCTACACATCAACGGTACTAAACATT TCGCAAAAATTTTAAAAATTAGAGCACCTTGAACTTGCGAAAAAGGTTCTCATCA ACTGTTTAA
10085 PGI1 repair oligo rv	TTAAACAGTTGATGAGAACCTTTTTCGCAAGTTCAAGGTGCTCTAATTTTTAAAA TTTTTGCGAAATGTTTAGTACCGTTGATGTGTAGAAGTAGTGCCCTGAAATACAT AGCGGTGTAT
10209 PFK2 repair oligo fw	CCAGTCCCGCATACCCCCTTTGCAACGTTAACGTTACCGCTAGCGTTTACCATCTC CACGACTTATGTATACTGGAATATGTGATATAGACGATTTAAAAGATAATTCCAA TAAACGTCC
10210 PFK2 repair oligo rv	GGACGTTTATTGGAATTATCTTTTAAATCGTCTATATCACATATTCCAGTATACA TAAGTCGTGGAGATGGTAAACGCTAGCGGTAACGTTAACGTTGCAAAGGGGGTAT GCGGGACTGG
10211 PFK1 repair oligo fw	AATTAATATCTCATTAACAAAGTTATTGTACATAATCCGGTACAATATTCTTCAA TGTACGTTTTAGGGTGTGCTTAATCTGCGTTGACAATGGTTCACGAAGACGACAT CGGCAACTTT
10212 PFK1 repair oligo rv	AAAGTTGCCGATGTCGTCTTCGTGAACCATTGTCAACGCAGATTAAGCACACCCTA AAACGTACATTGAAGAATATTGTACCGGATTATGTACAATAACTTTGTTAATGAG ATATTAATT
12333 FBA1 repair oligo fw	TCTTCTGTTCTTCTTTTTCTTTTGTCATATATAACCATAACCAAGTAATACATATT CAAAGTTAATTCAAATTAATTGATATAGTTTTTTAATGAGTATTGAATCTGTTTA GAAATAATG
12334 FBA1 repair oligo rev	CATTATTTCTAAACAGATTCAATACTCATTAAAAAACTATATCAATTAATTTGAA TTAACTTTGAATATGTATTACTTGGTTATGGTTATATATGACAAAAGAAAAAGAA GAACAGAAGA

10980 TPI1 repair oligo fw	TGTTTGTATTCTTTTCTTGCTTAAATCTATAACTACAAAAAACACATACATAAAC TAAAAGATTAATATAATTATATAAAAATATTATCTTCTTTTCTTTATATCTAGTG TTATGTAAAA
10981 TPI1 repair oligo rev	TTTTACATAACACTAGATATAAAGAAAAGAAGATAATATTTTTATATAATTATAT TAATCTTTTAGTTTATGTATGTGTTTTTTGTAGTTATAGATTTAAGCAAGAAAAG AATACAAACA
10978 TDH3 repair oligo fw	TTTTTTTAGTTTTAAAACACCAAGAACTTAGTTTCGAATAAACACACATAAACAA ACAAAGTGAATTTACTTTAAATCTTGCATTTAAATAAATTTTCTTTTTATAGCTT TATGACTTAG
10979 TDH3 repair oligo rv	CTAAGTCATAAAGCTATAAAAAGAAAATTTATTTAAATGCAAGATTTAAAGTAA ATTCACTTTGTTTGTTTATGTGTGTTTATTCGAAACTAAGTTCTTGGTGTTTTAA AACTAAAAAAA
10986 PGK1 repair oligo fw	AAGTTCGTTCGATCGTACTGTTACTCTCTCTCTTTCAAACAGAATTGTCCGAATCG TGTGATTTATATACGTATATATAGACTATTATTTATCTTTTAATGATTATTAAGA TTTTTATTA
10987 PGK1 repair oligo rev	TAATAAAAATCTTAATAATCATTAAAAGATAAATAATAGTCTATATATACGTATA TAAATCACACGATTCGGACAATTCTGTTTGAAAGAGAGAGAGTAACAGTACGATC GAACGAACTT
10984 GPM1 repair oligo fw	AATTTCAGCTGACAGCGAGTTTCATGATCGTGATGAACAATGGTAACGAGTTGTG GCTGTTTTTTCCCTCCATTTTTCTTACTGAATATATCAATGATATAGACTTGTATA GTTTATTAT
10985 GPM1 repair oligo rev	ATAATAAACTATACAAGTCTATATCATTGATATATTCAGTAAGAAAAATGGAGGG AAAAAACAGCCACAACTCGTTACCATTGTTCATCACGATCATGAAACTCGCTGTCA GCTGAAATT
10086 ENO2 repair oligo fw	TTTTCTTTTCTTAGTTTCTTTCATAACACCAAGCAACTAATACTATAACATACAAT AATATTTAACTAAGAATTATTAGTCTTTTCTGCTTATTTTTTCATCATAGTTTAG AACACTTTA
10087 ENO2 repair oligo rv	TAAAGTGTTCTAAACTATGATGAAAAAATAAGCAGAAAAGACTAATAATTCTTAG TTAAATATTATTGTATGTTATAGTATTAGTTGCTTGGTGTTATGAAAGAAACTAA GAAAAGAAAA
10982 PYK1 repair oligo fw	ATTATTCTCTCTTGTTTCTATTTACAAGACACCAATCAAAACAAATAAAACATCA TCACAAAAAAGAATCATGATTGAATGAAGATATTATTTTTTTGAATTATATTTTT TAAATTTTAT
10983 PYK1 repair oligo rev	ATAAAATTTAAAAAATATAATTCAAAAAAATAATATCTTCATTCAATCATGATTC TTTTTTGTGATGATGTTTTATTTGTTTTGATTGGTGTCTTGTAAATAGAAACAAG AGAGAATAAT
6075 COUNTER SELECT oligo fw	TTTTTCTCATCTCTTGGCTCTGGATCCGTTATCTGTTCTGTTACACAAGAAATCGT ACATACTAGAGCAAGATTTCAAATAAGTAACAGCAGCCATACGTTGAAACTACGG CAAAGGATT
6076 COUNTER SELECT oligo rv	AATCCTTTGCCGTAGTTTCAACGTATGGCTGCTGTTACTTATTTGAAATCTTGCTC TAGTATGTACGATTTCTTGTGTAACAGAACAGATAACGGATCCAGAGCCAAGAGA TGAGAAAAA
16750 STT4 repair IMS990 G1766R fw	CGTAATTTCGTATTTGTTGCAATTCAAGGATAGACATAATGGTAACATTATGTAC GATGATCAA**A**GACATTGTCTACATATCGATTTTGGGTTTATTTTTGATATTGTCC CAGGTGGTAT
16751 STT4 repair IMS990 G1766R rv	ATACCACCTGGGACAATATCAAAAATAAACCCAAAATCGATATGTAGACAATGTC TTTGATCATCGTACATAATGTTACCATTATGTCTATCCTTGAATTGCAACAAATA CGAAATTACG
16757 STT4 repair IMS992 F1775I fw	GATAGACATAATGGTAACATTATGTACGATGATCAAGGACATTGTCTACATATCG ATTTTGG**CA**TTATTTTTGATATTGTCCCAGGTGGTATCAAGTTTGAAGCAGTACC ATTCAAGCTG
16758 STT4 repair IMS992 F1775I rv	CAGCTTGAATGGTACTGCTTCAAACTTGATACCACCTGGGACAATATCAAAAATA ATGCCAAAATCGATATGTAGACAATGTCCTTGATCATCGTACATAATGTTACCAT TATGTCTATC
11590 Repair KL fw	AAGATAGTCGCCGAACTCGCAAGAGTCATTAACACCTCGCAATTGATGGGAAGTC CTCGCATATGACCTGAACCGACGGCAAATGCTCTTCAACTACGGCATACTTGCGGA AGCTACGGC
11591 Repair KL rev	GCCGTAGCTTCCGCAAGTATGCCGTAGTTGAAGAGCATTTGCCGTCGGTTCAGGTC ATATGCGAGGACTTCCCATCAATTGCGAGGTGTTAATGACTCTTGCGAGTTCGGC GACTATCTT

Table S9F – primers for construction of IMX165 and IMX2015.

Fragment	Name	Sequence
HXK2 deletion cassette	2788 HXK2 deletion cassette fw	ATTGTAGGAATATAATTCTCCACACATAATAAGTACGTTAATTAA ATAAACAGCTG AAGCTTCGTACGC
	2789 HXK2 deletion cassette rv	TTAAAAAAAGGGCACCTTCTTGTTGTTCAAACTTAATTTACAAAT TAAGTGCATAGGCCA CTAGTGGATCTG
HXK1 deletion cassette	1710 hxk1 deletion cassette fw	AAACTCACCCAAACAACTCAATTAGAATACTGAAAAAATAAGATG ATGACAAGAGGGTCGAACTCCAGCTGAAGCTTCGTACGC
	1711_hxk1 deletion cassette rv	AGGGAGGGAAAAACACATTTATATTTCATTACATTTTTTTCATTA GCCTAAGTCGTAATTGAGTCGCATAGGCCACTAGTGGATCTG
pPDC1 amplification	14670_pPDC1Sc+HXK2flank fw	TTTCTAATGCCTTTTCCATCATGTTACTACGAGTTTTCTGAACCTC CTCGCACATTGGTAGCGACTGGGTGAGCATATG
	14671_pPDC1Sc+HXK2flank rev	TGGCACATCGGCCATGGAACCCTTTCTGGCTTGTGGTTTTTTTGGA CCTAAATGAACCATTTTGATTGATTTGACTGTGTTATTTTG
Confirmation promoter replacement	3238 HXK2 outside fw	GCCTTTTCCATCATGTTACTAC
	4834 HXK2 rev	ACCCAATGGAATTGGCTCAG

Table S9G – Primers for amplification *URA3, HsPKL* and diagnostic primers.

Fragment	Name	Sequence
ScURA3 fragment	11766_URA3 + TDH1 flank fw	GATATTTCACCAACACACACAAAAAACAGTACTTCACTAAATTTAC ACACAAAACAAAATTGAGTATTTTCAATAAATTTGTAGAGGACT
	11767_URA3 + TDH1 flank rev	CGGTAGTATTTATGTATATTCAAAAAAAAATCATTATCCTCATCA AGATTGCTTTATTTATTGCTTTTGTTCCACTACTTTTTG
Confirmation integration *URA3* in *tdh1*	1989_URA3 outside fw	CCACGTGCAGAACAACATAG
	8306_URA3 rev	TGCTCCTTCCTTCGTTCTTC
	8377_URA3 fw	GGGAATCTCGGTCGTAATG
	2347_URA3 outside rev	GTCACATATTGTGGGTATGTGC
Confirmation removal SinLoG cassette	11898_SeqFW_SGA	CGCGGAAACGGGTATTAGGG
	11899_SeqRV_SGA	CTAGATCCGGTAAGCGACAG
Confirmation of replacement *HK4* with *HK2*	2794 HXK2-FW KO conformation	CACCTTCGCCACTGTCTTATCTAC
	2923 HXK2-RV wca del conf2	GGGCACCTTCTTGTTGTTCAAACT
	1452 HXK2FW1	TTCGCCACTGTCTTATCTAC
	13508 HK2 rev	ATCCTTGATTTGCAACTTGTC
HsPKL gene	10846 HsPKL gene fw	CCATAGGTCTCATATGGAAGGTCCAGCTGGTTATTTGAG
	10847 HsPKL gene rev	GGCCGGTCTCAGGATTCAGGAGATG

Table S10A – Plasmids containing human genes.

Name	Fragment	Source
pGGKp001	*HK1*	GeneArt
pGGKp002	*HK2*	GeneArt
pGGKp003	*HK3*	GeneArt
pGGKp004	*HK4*	GeneArt
pGGKp005	*GPI*	GeneArt
pGGKp006	*PFKM*	GeneArt
pGGKp007	*PFKP*	GeneArt
pGGKp008	*PFKL*	GeneArt
pGGKp009	*ALDOA*	GeneArt
pGGKp010	*ALDOB*	GeneArt
pGGKp011	*ALDOC*	GeneArt
pGGKp012	*TPI*	GeneArt
pGGKp013	*GAPDH*	GeneArt
pGGKp014	*GAPDHS*	GeneArt
pGGKp015	*PGK1*	GeneArt
pGGKp016	*PGK2*	GeneArt
pGGKp017	*PGAM1*	GeneArt
pGGKp018	*PGAM2*	GeneArt
pGGKp019	*ENO1*	GeneArt
pGGKp020	*ENO2*	GeneArt
pGGKp021	*ENO3*	GeneArt
pGGKp022	*PKM1*	GeneArt
pGGKp023	*PKM2*	GeneArt
pGGKp024	*PKR*	GeneArt

Table S10B – Plasmids containing yeast promoter or terminator.

Name	Fragment	Source
pUD565	Entry vector, CamR	GeneArt
pGGKp025	*pPDC1 sc*	This study
pGGKp026	*pGPM1 sc*	This study
pGGKp027	*pFBA1 sc*	This study
pGGKp028	*pENO2 sc*	This study
pGGKp029	*pADH1 sc*	This study
pGGKp030	*pTPI1 sc*	This study
pGGKp031	*pPFK2 sc*	This study
pGGKp032	*pTEF1 sc*	This study
pGGKp033	*pPGI1 sc*	This study
pGGKp034	*pPYK1 sc*	This study
pGGKp035	*pTDH3 sc*	This study
pGGKp036	*pPGK1 sc*	This study
pGGKp037	*tADH1 sc*	This study
pGGKp038	*tTEF2 sc*	This study
pGGKp039	*tTEF1 sc*	This study
pGGKp040	*tPYK1 sc*	This study
pGGKp041	*tTDH3 sc*	This study

Table S10B – Continued

Name	Fragment	Source
pGGKp042	*tTPI1 sc*	This study
pGGKp043	*tPGK1 sc*	This study
pGGKp044	*tPGI1 sc*	This study
pGGKp045	*tPDC1 sc*	This study
pGGKp046	*tFBA1 sc*	This study
pGGKp047	*pACT1 sc*	This study
pGGKp048	*tGPM1 sc*	This study
pGGKp096	*pHXK2 sc*	(*102*)
pGGKp097	*tHXK2 sc*	(*102*)
pYTK055	*tENO2 sc*	(*101*)
pYTK014	*pTEF2 sc*	(*101*)

Table S10C – Plasmids used to construct pGGKd002 and pGGKd003.

Name	Fragment	Source
pYTK002	ConLS	*(101)*
pYTK047	GFP dropout	*(101)*
pYTK067	ConR1	*(101)*
pYTK074	URA3	*(101)*
pYTK086	URA3 3' Homology	*(101)*
pYTK090	KanR-ColE1	*(101)*
pYTK092	URA3 5' Homology	*(101)*
pGGKd002	GFP dropout integration plasmid	This study
pGGKd003	GFP dropout plasmid made from	This study

Table S10D – Integration plasmids containing human transcriptional unit.

Name	Construct	Source
pUDI133	*ScPDC1p-HK1-ScPDC1t*	This study
pUDI134	*ScPDC1p-HK2-ScPDC1t*	This study
pUDI135	*ScPDC1p-HK3-ScPDC1t*	This study
pUDI136	*ScPDC1p-HK4-ScPDC1t*	This study
pUDI206	*ScHXK2-HK2-tHXK2*	This study
pUDI207	*ScHXK2-HK4-tHXK2*	This study
pUDI137	*ScTEF2p-GPI1-ScTEF2t*	This study
pUDI138	*ScTEF1p-PFKM-ScTEF1t*	This study
pUDI139	*ScTEF1p-PFKP-ScTEF1t*	This study
pUDI140	*ScTEF1p-PFKL-ScTEF1t*	This study
pUDI141	*ScFBA1p-ALDOA-ScFBA1t*	This study
pUDI142	*ScFBA1p-ALDOB-ScFBA1t*	This study
pUDI143	*ScFBA1p-ALDOC-ScFBA1t*	This study
pUDI144	*ScTPI1p-TPI-ScTPI1t*	This study
pUDI145	*ScTDH3p-GAPDH-ScTDH3t*	This study
pUDI146	*ScTDH3p-GAPDHS-ScTDH3t*	This study
pUDI147	*ScPGK1p-PGK1-ScPGK1t*	This study
pUDI148	*ScPGK1p-PGK2-ScPGK1t*	This study

Table S10D – Continued

Name	Construct	Source
pUDI149	*ScGPM1p-PGAM1-ScGPM1t*	This study
pUDI150	*ScGPM1p-PGAM2-ScGPM1t*	This study
pUDI151	*ScENO2p-ENO1-ScENO2t*	This study
pUDI152	*ScENO2p-ENO2-ScENO2t*	This study
pUDI153	*ScENO2p-ENO3-ScENO2t*	This study
pUDI154	*ScPYK1p-PKM1-ScPYK1t*	This study
pUDI155	*ScPYK1p-PKM2-ScPYK1t*	This study
pUDI156	*ScPYK1p-PKR-ScPYK1t*	This study
pUDI157	*ScPYK1p-PKL-ScPYK1t*	This study

Table S10E – Plasmids used as PCR template in this study.

Name	Relevant construct	Source
Plasmids used for construction of IMX165		
pSH47	PGAL1-Cre-TCYC1, KlURA3	(*134*)
pUG73	loxP-*KlLEU2*-loxP cassette	(*134*)
pUG6	loxP-*KanMX*-loxP cassette	(*134*)
Plasmids used for construction of IMX1076		
pUG-natNT1	NatNT1	(*134*)
p414-TEF1p-	TEF1p-Cas9-CYC1t	(*135*)
Multicopy plasmid used as PCR template		
pUDE750	ScHXK2p-HK4-ScHXK2t	This

Table S10F – gRNA plasmids for CRISPR Cas9 genome editing.

Name	Construct	Source
pMel13	2µm ampR KanMX gRNA-CAN1.Y	(*104*)
pRos13	2µm ampR KanMX gRNA-CAN1.Y gRNA-ADE2.Y	(*104*)
pMel10	2µm ampR KlURA3 gRNA-CAN1.Y	(*104*)
pUDE342	URA3 SNR52p-gRNA.SGA1-SUP4t RECYCLE SinLoG	(*27*)
pUDE327	URA3 SNR52p-gRNA.*HXK2*-SUP4t	*(28)*
pUDR265	2µm ampR KanMX gRNA-PFK1 gRNA-PFK2	This study
pUDR337	2µm ampR KanMX gRNA-GPM1 gRNA-GPM1	This study
pUDR338	2µm ampR KanMX gRNA-FBA1 gRNA-FBA1	This study
pUDR371	2µm ampR KanMX gRNA-HXK2	This study
pUDR387	2µm ampR KanMX gRNA-HK4	This study
pUDR376	2µm ampR AmdS gRNA-X2	(*136*)
pUDR666	2µm ampR KanMX gRNA-STT4	This study
pUDR667	2µm ampR KanMX gRNA-STT4	This study

References

1. P. M. Nurse, Nobel lecture: cyclin dependent kinases and cell cycle control. *Bioscience reports* **22**, 487-499 (2002).
2. J. L. Woolford, S. J. Baserga, Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* **195**, 643-681 (2013).
3. D. Petranovic, J. Nielsen, Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol.* **26**, 584-590 (2008).
4. J. M. Laurent, J. H. Young, A. H. Kachroo, E. M. Marcotte, Efforts to make and apply humanized yeast. *Briefings in functional genomics* **15**, 155-163 (2016).
5. N. Zhang *et al.*, Using yeast to place human genes in functional categories. *Gene* **303**, 121-129 (2003).
6. A. Hamza *et al.*, Complementation of yeast genes with human genes as an experimental platform for functional testing of human genetic variants. *Genetics* **201**, 1263-1274 (2015).
7. A. Hamza, M. R. Driessen, E. Tammpere, N. J. O'Neil, P. Hieter, Cross-species complementation of nonessential yeast genes establishes platforms for testing inhibitors of human proteins. *Genetics*, (2020).
8. R. K. Garge, J. M. Laurent, A. H. Kachroo, E. M. Marcotte, Systematic humanization of the yeast cytoskeleton discerns functionally replaceable from divergent human genes. *bioRxiv*, (2019).
9. A. H. Kachroo *et al.*, Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **348**, 921-925 (2015).
10. J. M. Laurent *et al.*, Humanization of yeast genes with multiple human orthologs reveals principles of functional divergence between paralogs. *bioRxiv*, 668335 (2019).
11. S. Sun *et al.*, An extended set of yeast-based functional assays accurately identifies human disease mutations. *Genome Res.* **26**, 670-680 (2016).
12. V. E. Prince, F. B. Pickett, Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics* **3**, 827-837 (2002).
13. D. Steinke, S. Hoegg, H. Brinkmann, A. Meyer, Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates. *BMC Biol.* **4**, 16 (2006).
14. J. L. Gordon, K. P. Byrne, K. H. Wolfe, Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet.* **5**, (2009).
15. A. H. Kachroo *et al.*, Systematic bacterialization of yeast genes identifies a near-universally swappable pathway. *eLife* **6**, e25093 (2017).
16. N. Agmon *et al.*, Phylogenetic debugging of a complete human biosynthetic pathway transplanted into yeast. *Nucleic Acids Res.* **48**, 486-499 (2020).
17. S. R. Hamilton, D. Zha, in *Glyco-Engineering*. (Springer, 2015), pp. 73-90.
18. V. M. Labunskyy *et al.*, The insertion Green Monster (iGM) method for expression of multiple exogenous genes in yeast. *G3 (Bethesda)* **4**, 1183-1191 (2014).
19. D. M. Truong, J. D. Boeke, Resetting the yeast epigenome with human nucleosomes. *Cell* **171**, 1508-1519. e1513 (2017).
20. O. Warburg, The metabolism of carcinoma cells. *The Journal of Cancer Research* **9**, 148-163 (1925).
21. I. Mayordomo, P. Sanz, Human pancreatic glucokinase (GlkB) complements the glucose signalling defect of *Saccharomyces cerevisiae hxk2* mutants. *Yeast* **18**, 1309-1316 (2001).
22. M. Lu, D. Ammar, H. Ives, F. Albrecht, S. L. Gluck, Physical interaction between aldolase and vacuolar H+-ATPase is essential for the assembly and activity of the proton pump. *J. Biol. Chem.* **282**, 24495-24503 (2007).
23. J. J. Heinisch, Expression of heterologous phosphofructokinase genes in yeast. *FEBS Lett.* **328**, 35-40 (1993).
24. G. Sriram, J. A. Martinez, E. R. McCabe, J. C. Liao, K. M. Dipple, Single-gene disorders: what role could moonlighting enzymes play? *Am. J. Hum. Genet.* **76**, 911-924 (2005).
25. C. Gancedo, C. L. Flores, Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.* **72**, 197-210 (2008).
26. D. Solis-Escalante *et al.*, A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell* **14**, 804-816 (2015).
27. N. G. Kuijpers *et al.*, Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. USA* **113**, 15060-15065 (2016).

28. J. J. Marsh, H. G. Lebherz, Fructose-bisphosphate aldolases: an evolutionary history. *Trends Biochem Sci* **17**, 110-113 (1992).
29. W. J. Israelsen, M. G. Vander Heiden, Pyruvate kinase: Function, regulation and role in cancer. *Semin Cell Dev Biol* **43**, 43-51 (2015).
30. M. A. Blazquez, R. Lagunas, C. Gancedo, J. M. Gancedo, Trehalose-6-phosphate, a new regulator of yeast glycolysis that inhibits hexokinases. *FEBS Lett.* **329**, 51-54 (1993).
31. R. L. Vicente *et al.*, Trehalose-6-phosphate promotes fermentation and glucose repression in *Saccharomyces cerevisiae*. *Microb Cell* **5**, 444-459 (2018).
32. B. M. Bonini, P. Van Dijck, J. M. Thevelein, Uncoupling of the glucose growth defect and the deregulation of glycolysis in *Saccharomyces cerevisiae Tps1* mutants expressing trehalose-6-phosphate-insensitive hexokinase from *Schizosaccharomyces pombe*. *Biochim Biophys Acta* **1606**, 83-93 (2003).
33. H. Ardehali *et al.*, Functional organization of mammalian hexokinase II. Retention of catalytic and regulatory functions in both the NH2- and COOH-terminal halves. *J Biol Chem* **271**, 1849-1852 (1996).
34. M. H. Nawaz *et al.*, The catalytic inactivation of the N-half of human hexokinase 2 and structural and biochemical characterization of its mitochondrial conformation. *Bioscience reports* **38**, (2018).
35. H. J. Tsai, J. E. Wilson, Functional organization of mammalian hexokinases: both N- and C-terminal halves of the rat type II isozyme possess catalytic sites. *Arch Biochem Biophys* **329**, 17-23 (1996).
36. J. H. van Heerden *et al.*, Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science* **343**, 1245114 (2014).
37. B. Teusink, M. C. Walsh, K. van Dam, H. V. Westerhoff, The danger of metabolic pathways with turbo design. *Trends Biochem. Sci.* **23**, 162-169 (1998).
38. J. D. Dombrauckas, B. D. Santarsiero, A. D. Mesecar, Structural basis for tumor pyruvate kinase M2 allosteric regulation and catalysis. *Biochemistry* **44**, 9417-9429 (2005).
39. G. J. Crowther, M. F. Carey, W. F. Kemper, K. E. Conley, Control of glycolysis in contracting skeletal muscle. I. Turning it on. *Am J Physiol Endocrinol Metab* **282**, E67-73 (2002).
40. G. van Hall, Lactate kinetics in human tissues at rest and during exercise. *Acta physiologica* **199**, 499-508 (2010).
41. G. J. Crowther, W. F. Kemper, M. F. Carey, K. E. Conley, Control of glycolysis in contracting skeletal muscle. II. Turning it off. *Am J Physiol Endocrinol Metab* **282**, E74-79 (2002).
42. V. B. Ritov, D. E. Kelley, Hexokinase isozyme distribution in human skeletal muscle. *Diabetes* **50**, 1253-1262 (2001).
43. G. I. Bell, S. J. Pilkis, I. T. Weber, K. S. Polonsky, Glucokinase mutations, insulin secretion, and diabetes mellitus. *Annu Rev Physiol* **58**, 171-186 (1996).
44. F. M. Matschinsky, Glucokinase as glucose sensor and metabolic signal generator in pancreatic beta-cells and hepatocytes. *Diabetes* **39**, 647-652 (1990).
45. R. Diaz-Ruiz, S. Uribe-Carvajal, A. Devin, M. Rigoulet, Tumor cell energy metabolism and its common features with yeast metabolism. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* **1796**, 252-265 (2009).
46. P. Fernandez-Garcia, R. Pelaez, P. Herrero, F. Moreno, Phosphorylation of yeast hexokinase 2 regulates its nucleocytoplasmic shuttling. *J Biol Chem* **287**, 42151-42164 (2012).
47. P. Herrero, C. Martínez-Campa, F. Moreno, The hexokinase 2 protein participates in regulatory DNA-protein complexes necessary for glucose repression of the *SUC2* gene in *Saccharomyces cerevisiae*. *FEBS Lett.* **434**, 71-76 (1998).
48. K. Elbing *et al.*, Role of hexose transport in control of glycolytic flux in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **70**, 5323-5330 (2004).
49. J. Van Dijken *et al.*, An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* **26**, 706-714 (2000).
50. P. Van Hoek, J. P. Van Dijken, J. T. Pronk, Effect of specific growth rate on fermentative capacity of baker's yeast. *Appl. Environ. Microbiol.* **64**, 4226-4233 (1998).
51. E. Boles *et al.*, Characterization of a glucose-repressed pyruvate kinase (Pyk2p) in *Saccharomyces cerevisiae* that is catalytically insensitive to fructose-1, 6-bisphosphate. *J. Bacteriol.* **179**, 2987-2993 (1997).
52. J. H. van Heerden *et al.*, Lost in transition: startup of glycolysis yields subpopulations of nongrowing cells. *Science*, (2014).
53. M. Lu, Y. Y. Sautin, L. S. Holliday, S. L. Gluck, The glycolytic enzyme aldolase mediates assembly, expression, and activity of vacuolar H+-ATPase. *J. Biol. Chem.* **279**, 8732-8739 (2004).

54. B. L. Decker, W. T. Wickner, Enolase activates homotypic vacuole fusion and protein transport to the vacuole in yeast. *J. Biol. Chem.* **281**, 14523-14528 (2006).
55. J. v. d. Brink *et al.*, Dynamics of glycolytic regulation during adaptation of *Saccharomyces cerevisiae* to fermentative metabolism. *Appl. Environ. Microbiol.* **74**, 5710-5723 (2008).
56. S. L. Tai *et al.*, 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 (2007).
57. M. L. Jansen *et al.*, Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity. *Microbiology* **151**, 1657-1669 (2005).
58. H. Bisswanger, Enzyme assays. *Perspectives in Science* **1**, 41-55 (2014).
59. T. E. Sandberg, M. J. Salazar, L. L. Weng, B. O. Palsson, A. M. Feist, The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab Eng* **56**, 1-16 (2019).
60. P. J. Trotter, W. I. Wu, J. Pedretti, R. Yates, D. R. Voelker, A genetic screen for aminophospholipid transport mutants identifies the phosphatidylinositol 4-kinase, STT4p, as an essential component in phosphatidylserine metabolism. *J Biol Chem* **273**, 13189-13196 (1998).
61. A. Audhya, M. Foti, S. D. Emr, Distinct roles for the yeast phosphatidylinositol 4-kinases, Stt4p and Pik1p, in secretion, cell growth, and organelle membrane dynamics. *Mol Biol Cell* **11**, 2673-2689 (2000).
62. J. L. DeRisi, V. R. Iyer, P. O. Brown, Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686 (1997).
63. R. L. Smith, A. D. Johnson, Turning genes off by Ssn6–Tup1: a conserved system of transcriptional repression in eukaryotes. *Trends Biochem. Sci.* **25**, 325-330 (2000).
64. M. A. Treitel, M. Carlson, Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein. *Proceedings of the National Academy of Sciences* **92**, 3132-3136 (1995).
65. A. S. Deshmukh *et al.*, Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Molecular & Cellular Proteomics* **14**, 841-853 (2015).
66. A. Shimizu, F. Suzuki, K. Kato, Characterization of αα, ββ, γγ and αγ human enolase isozymes, and preparation of hybrid enolases (αγ, βγ and αβ) from homodimeric forms. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **748**, 278-284 (1983).
67. N. Durany, J. Carreras, Distribution of phosphoglycerate mutase isozymes in rat, rabbit and human tissues. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **114**, 217-223 (1996).
68. R. N. Harkins, J. A. Black, M. B. Rittenberg, M2 isozyme of pyruvate kinase from human kidney as the product of a separate gene: its purification and characterization. *Biochemistry* **16**, 3831-3837 (1977).
69. D. A. Skaff, C. S. Kim, H. J. Tsai, R. B. Honzatko, H. J. Fromm, Glucose 6-phosphate release of wild-type and mutant human brain hexokinases from mitochondria. *J Biol Chem* **280**, 38403-38409 (2005).
70. B. Teusink *et al.*, Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.* **267**, 5313-5329 (2000).
71. T. Rapoport, R. Heinrich, Mathematical analysis of multienzyme systems. I. Modelling of the glycolysis of human erythrocytes. *BioSyst.* **7**, 120-129 (1975).
72. J. P. Schmitz, N. A. Van Riel, K. Nicolay, P. A. Hilbers, J. A. Jeneson, Silencing of glycolysis in muscle: experimental observation and numerical analysis. *Experimental physiology* **95**, 380-397 (2010).
73. K.-J. Ahn, J.-S. Kim, M.-J. Yun, J.-H. Park, J.-D. Lee, Enzymatic properties of the N-and C-terminal halves of human hexokinase II. *BMB reports* **42**, 350-355 (2009).
74. J. E. Wilson, Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *J Exp Biol* **206**, 2049-2057 (2003).
75. K. van Eunen, J. A. Kiewiet, H. V. Westerhoff, B. M. Bakker, Testing biochemistry revisited: how *in vivo* metabolism can be understood from *in vitro* enzyme kinetics. *PLoS Comp. Biol.* **8**, (2012).
76. X. Liu, C. S. Kim, F. T. Kurbanov, R. B. Honzatko, H. J. Fromm, Dual mechanisms for glucose 6-phosphate inhibition of human brain hexokinase. *J Biol Chem* **274**, 31155-31159 (1999).
77. L. Grossbard, R. T. Schimke, Multiple Hexokinases of Rat Tissues purification and comparison of soluble forms. *J. Biol. Chem.* **241**, 3546-3560 (1966).
78. S.-M. Fendt, U. Sauer, Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. *BMC systems biology* **4**, 12 (2010).
79. V. Pancholi, Multifunctional alpha-enolase: its role in diseases. *Cell Mol Life Sci* **58**, 902-920 (2001).

80. M. Didiasova, L. Schaefer, M. Wygrecka, When place matters: shuttling of Enolase-1 across cellular compartments. *Front Cell Dev Biol* **7**, 61 (2019).
81. N. Entelis *et al.*, A glycolytic enzyme, enolase, is recruited as a cofactor of tRNA targeting toward mitochondria in *Saccharomyces cerevisiae*. *Genes Dev.* **20**, 1609-1620 (2006).
82. N. S. Entelis, O. A. Kolesnikova, R. P. Martin, I. A. Tarassov, RNA delivery into mitochondria. *Adv Drug Deliv Rev* **49**, 199-215 (2001).
83. O. A. Kolesnikova *et al.*, Nuclear DNA-encoded tRNAs targeted into mitochondria can rescue a mitochondrial DNA mutation associated with the MERRF syndrome in cultured human cells. *Hum Mol Genet* **13**, 2519-2534 (2004).
84. L. Ye, A. L. Kruckeberg, J. A. Berden, K. van Dam, Growth and glucose repression are controlled by glucose transport in *Saccharomyces cerevisiae* cells containing only one glucose transporter. *J. Bacteriol.* **181**, 4673-4675 (1999).
85. H. Ma, L. Bloom, C. Walsh, D. Botstein, The residual enzymatic phosphorylation activity of hexokinase II mutants is correlated with glucose repression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **9**, 5643-5649 (1989).
86. J. M. Gancedo, Yeast carbon catabolite repression. *Microbiol. Mol. Biol. Rev.* **62**, 334-361 (1998).
87. D. Ahuatzi, P. Herrero, T. De La Cera, F. Moreno, The glucose-regulated nuclear localization of hexokinase 2 in *Saccharomyces cerevisiae* is Mig1-dependent. *J. Biol. Chem.* **279**, 14440-14446 (2004).
88. P. Lynedjian, Molecular physiology of mammalian glucokinase. *Cell. Mol. Life Sci.* **66**, 27 (2009).
89. P. Daran-Lapujade *et al.*, The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proceedings of the National Academy of Sciences* **104**, 15753-15758 (2007).
90. M. J. de Groot *et al.*, Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. *Microbiology* **153**, 3864-3878 (2007).
91. K. F. Tolias, L. C. Cantley, Pathways for phosphoinositide synthesis. *Chem Phys Lipids* **98**, 69-77 (1999).
92. H. Hu *et al.*, Phosphoinositide 3-kinase regulates glycolysis through mobilization of aldolase from the actin cytoskeleton. *Cell* **164**, 433-446 (2016).
93. T. Kusakabe, K. Motoki, K. Hori, Mode of interactions of human aldolase isozymes with cytoskeletons. *Arch Biochem Biophys* **344**, 184-193 (1997).
94. V. F. Waingeh *et al.*, Glycolytic enzyme interactions with yeast and skeletal muscle F-actin. *Biophys. J.* **90**, 1371-1384 (2006).
95. M. Wijsman *et al.*, A toolkit for rapid CRISPR-Sp Cas9 assisted construction of hexose-transport-deficient *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.* **19**, foy107 (2019).
96. K.-D. Entian, P. Kötter, 25 Yeast genetic strain and plasmid collections. *Methods in Microbiology* **36**, 629-666 (2007).
97. C. Verduyn, E. Postma, W. A. Scheffers, J. P. Van Dijken, 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 (1992).
98. D. Solis-Escalante *et al.*, *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **13**, 126-139 (2013).
99. G. Bertani, Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* **186**, 595-600 (2004).
100. G. Bertani, Studies on lysogenis I.: The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293-300 (1951).
101. M. E. Lee, W. C. DeLoache, B. Cervantes, J. E. Dueber, A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4**, 975-986 (2015).
102. F. J. Boonekamp *et al.*, The genetic makeup and expression of the glycolytic and fermentative pathways are highly conserved within the *Saccharomyces* genus. *Frontiers in genetics* **9**, 504 (2018).
103. R. D. Gietz, R. A. Woods, Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87-96 (2002).
104. R. Mans *et al.*, CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **15**, (2015).
105. M. D. Mikkelsen *et al.*, Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* **14**, 104-111 (2012).
106. U. Güldener, S. Heck, T. Fiedler, J. Beinhauer, J. H. Hegemann, A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24**, 2519-2524 (1996).

107. B. Santos, M. Snyder, Sbe2p and sbe22p, two homologous Golgi proteins involved in yeast cell wall formation. *Molecular biology of the cell* **11**, 435-452 (2000).
108. E. Postma, C. Verduyn, W. A. Scheffers, J. P. Van Dijken, Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **55**, 468-477 (1989).
109. L. A. Cruz *et al.*, Similar temperature dependencies of glycolytic enzymes: an evolutionary adaptation to temperature dynamics? *BMC systems biology* **6**, 151 (2012).
110. O. H. Lowry, N. J. Rosebrough, A. L. Farr, R. J. Randall, Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265-275 (1951).
111. M. C. Silveira, E. Carvajal, E. P. Bon, Assay for *in vivo* yeast invertase activity using NaF. *Anal. Biochem.* **238**, 26-28 (1996).
112. S. B. Haase, S. I. Reed, Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* **1**, 132-136 (2002).
113. A. M. Almonacid Suarez, Q. Zhou, P. van Rijn, M. C. Harmsen, Directional topography gradients drive optimum alignment and differentiation of human myoblasts. *Journal of tissue engineering and regenerative medicine* **13**, 2234-2245 (2019).
114. R. W. Guynn, D. Veloso, J. R. Lawson, R. L. Veech, The concentration and control of cytoplasmic free inorganic pyrophosphate in rat liver *in vivo*. *Biochem. J.* **140**, 369-375 (1974).
115. P. Rorsman, G. Trube, Glucose dependent K+-channels in pancreaticβ-cells are regulated by intracellular ATP. *Pflügers Archiv* **405**, 305-309 (1985).
116. J. Tschopp, K. Schroder, NLRP3 inflammasome activation: The convergence of multiple signalling pathways on ROS production? *Nature reviews immunology* **10**, 210-215 (2010).
117. L. O. Kristensen, Associations between transports of alanine and cations across cell membrane in rat hepatocytes. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **251**, G575-G584 (1986).
118. K. Conley, M. Blei, T. Richards, M. Kushmerick, S. A. Jubrias, Activation of glycolysis in human muscle *in vivo*. *American Journal of Physiology-Cell Physiology* **273**, C306-C315 (1997).
119. S. D. Lidofsky, M.-H. Xie, A. Sostman, B. F. Scharschmidt, J. G. Fitz, Vasopressin increases cytosolic sodium concentration in hepatocytes and activates calcium influx through cation-selective channels. *J. Biol. Chem.* **268**, 14632-14636 (1993).
120. G. E. Breitwieser, A. A. Altamirano, J. M. Russell, Osmotic stimulation of Na (+)-K (+)-Cl-cotransport in squid giant axon is [Cl-] i dependent. *American Journal of Physiology-Cell Physiology* **258**, C749-C753 (1990).
121. L. Janssen, S. Sims, Acetylcholine activates non-selective cation and chloride conductances in canine and guinea-pig tracheal myocytes. *The Journal of Physiology* **453**, 197-218 (1992).
122. J. S. Ingwall, Phosphorus nuclear magnetic resonance spectroscopy of cardiac and skeletal muscles. *American Journal of Physiology-Heart and Circulatory Physiology* **242**, H729-H744 (1982).
123. E. Murphy, C. Steenbergen, L. A. Levy, B. Raju, R. E. London, Cytosolic free magnesium levels in ischemic rat heart. *J. Biol. Chem.* **264**, 5622-5627 (1989).
124. M. Bárány, *Biochemistry of smooth muscle contraction*. (Elsevier, 1996).
125. P. Bruch, K. D. Schnackerz, R. W. Gracy, Matrix-Bound Phosphoglucose Isomerase: Formation and Properties of Monomers and Hybrids. *Eur. J. Biochem.* **68**, 153-158 (1976).
126. A. K. Groen, R. Vervoorn, R. Van der Meer, J. Tager, Control of gluconeogenesis in rat liver cells. I. Kinetics of the individual enzymes and the effect of glucagon. *J. Biol. Chem.* **258**, 14346-14353 (1983).
127. H. Ishibashi, G. L. Cottam, Glucagon-stimulated phosphorylation of pyruvate kinase in hepatocytes. *J. Biol. Chem.* **253**, 8767-8771 (1978).
128. S. Oudard *et al.*, High glycolysis in gliomas despite low hexokinase transcription and activity correlated to chromosome 10 loss. *Br. J. Cancer* **74**, 839-845 (1996).
129. V. N. Civelek, J. A. Hamilton, K. Tornheim, K. L. Kelly, B. E. Corkey, Intracellular pH in adipocytes: effects of free fatty acid diffusion across the plasma membrane, lipolytic agonists, and insulin. *Proceedings of the National Academy of Sciences* **93**, 10139-10144 (1996).
130. E. Van Schaftingen, H. G. Hers, Formation of fructose 2,6-bisphosphate from fructose 1,6-bisphosphate by intramolecular cyclisation followed by alkaline hydrolysis. *European Journal of Biochemistry* **117**, 319-323 (1981).
131. J. C. Wolters *et al.*, Translational targeted proteomics profiling of mitochondrial energy metabolic pathways in mouse and human samples. *Journal of proteome research* **15**, 3204-3213 (2016).

132. J. F. Nijkamp *et al.*, De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* **11**, 36 (2012).
133. A. N. Salazar *et al.*, Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.* **17**, 10.1093/femsyr/fox1074 (2017).
134. U. Gueldener, J. Heinisch, G. Koehler, D. Voss, J. Hegemann, A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res.* **30**, e23-e23 (2002).
135. J. E. DiCarlo *et al.*, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336-4343 (2013).
136. A. K. Wronska *et al.*, Exploiting the diversity of *Saccharomycotina* yeasts to engineer biotin-independent growth of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, (2020).

Outlook

To develop powerful cell factories, metabolic engineering approaches need to intensively and extensively remodel microbial genomes, naturally optimized for growth and fitness and not product formation. This remodeling is hindered by our limited understanding of biological systems. Synthetic Biology aims at increasing the predictability of engineered biological systems by improving our understanding and delivering powerful molecular tools. The work performed in this thesis addresses these key issues, providing new tools, methodologies and understanding for the industrial and model yeast *S. cerevisiae*.

A persistent challenge for strain construction is the limited availability of strong, constitutive promoters, required for the expression of heterologous proteins often characterized by suboptimal performance. By mining the genome of the *Saccharomyces* genus, Chapter 3 brings a new addition to yeast molecular toolbox, with a series of glycolytic promoters that can be used for strong, context-independent expression in *S. cerevisiae*. Beyond this technical contribution, this first cross-species exploration of glycolysis revealed a remarkable conservation of the set-up and expression of glycolytic genes between *S. cerevisiae*, *S. eubayanus* and *S. kudriavzevii*. Further genome mining within the *Saccharomyces* genus promises to deliver a broader range of strong and constitutive promoters. Furthermore, the differences between the *Saccharomyces* species (e.g. sugar utilization, temperature optima) might offer new sets of valuable context-dependent promoters for *S. cerevisiae*.

Chapter 2 demonstrates that core metabolic processes can be easily and rapidly remodeled using the pathway swapping approach. While glycolysis is used as paradigm, this approach can be extended to any pathway or cellular process and paves the way towards modular genome engineering. An interesting observation is the reduced growth rate of the SwYG strain, a platform strain for pathway swapping. Why SwYG grows slower than an equivalent strain in which the glycolytic genes remain in their native locus after minimization remains unknown. Several mechanisms can be speculated, such as an impact of this co-localization on the glycolytic genes themselves, or on the local genetic structure at the integration site resulting in changes in gene expression or DNA replication for instance. While the precise cause is yet unknown, this work has identified a gap in our understanding of the genetic factors involved in genome engineering. SwYG is perfectly suited to help understanding these factors, thereby bringing new insights to improve the genetic design of engineered pathways and processes.

Platforms such as SwYG can be used to address a broad range of technical, fundamental and applied questions. The ability to rapidly remodel the complete glycolytic pathway was exploited in chapters 4 and 5. In Chapter 4, all native glycolytic genes were

simultaneously swapped with watermarked versions. In-depth comparison of native and watermarked versions of glycolysis enabled to validate a simple and innocuous watermarking method for open reading frames that can be used for any gene. Considering the increasing number of synthetic biology endeavours worldwide, DNA watermarking will most certainly gain in popularity, making this watermarking method a valuable addition to *S. cerevisiae* molecular toolbox.

Chapter 5 reports the very first example of full humanization of an essential pathway in central carbon metabolism. This work explored a new strategy to study the feasibility and impact of pathway transplantation. Combining single complementation, full humanization and laboratory evolution has brought new insight on the functionality of glycolytic enzymes and the regulation of the glycolytic pathway. Remarkably, all human glycolytic genes, with the notable exception of three hexokinases, were able to complement their yeast ortholog. The present results point towards metabolic regulation by small molecular weight effectors as the main cause of the lack of functionality of human hexokinases 2 *in vivo*. Further analysis of these enzymes in their native and mutated versions will bring insights in the biochemical requirements for hexokinase functionality in yeast. Other remarkable findings include the conservation of moonlighting functions across yeast and human enzymes, particularly for enolases, as well as the potential regulation of aldolase activity by *in vivo* interactions with the cytoskeleton. These preliminary results will have to be further explored, but already strengthen the position of *S. cerevisiae* as eukaryotic model.

The physiological role of allosteric regulation of several glycolytic enzymes in *S. cerevisiae* is to date unresolved. So far explored using computational models or by single mutation, pathway swapping enables to simultaneously alter all these regulations, as well as other transcriptional, translational or post-translational regulations. In this first attempt at pathway swapping, alteration of the allosteric regulations at the hexokinase and pyruvate kinase steps did not visibly affect the glycolytic activity during growth on fermentative and respiratory carbon sources or during transition between carbon sources. The glycolysis swapping platform is an excellent tool to explore the regulation of glycolysis, and other variants with different kinetic properties should be tested as well, to understand how metabolic regulation contributes to the flux in the pathway and to *S. cerevisiae* fitness and competitiveness. Computational models have a great role to play in predicting the metabolic requirements for glycolytic activity. As kinetic models have already been developed for both yeast and human glycolytic pathways, the future challenge will be to define the best model to predict the impact of a complete redesign, such as the humanization, of the pathway on yeast physiology.

In conclusion, this thesis paves the way towards extensive, modular remodeling of (essential) pathways and processes for fundamental and applied purposes.

Acknowledgements

This thesis would not have been there without the help and support of a lot of people. Due to the Corona pandemic I spent the last two months of writing on my thesis at home, which made me realize even more how important it is to have nice people around you and how much I already miss everyone at IMB!

First and foremost I would like to thank my two promotors Pascale and Jack. The transition from where I came from to a PhD at IMB was not a very easy one, but I am very grateful that you gave me this opportunity and you always kept trust in me. Pascale, I cannot describe how much I learned during the past 4.5 years! Thank you for guiding me through this project, for explaining me everything about glycolysis, for teaching me how to critically look at data and for always being ambitious. Next to the science I really enjoyed all our nice chats and occasional exchange of pictures of our recent artworks. Also I would like to thank you for your support during the difficult moments of my PhD and sharing your personal experiences with me. Jack, although we didn't discuss so much science anymore after the first year, you definitely contributed to the successful completion of my thesis. Thank you for your contagious enthusiasm and positivity and your contributions to my personal development. It helped me a lot.

Furthermore, I also would like to thank the other PIs of IMB, Jean-Marc, Ton and Robert for your contributions to make IMB a nice working environment. Jean-Marc, thank you for your input in the discussions during the ERC meetings as well.

During my PhD I had the pleasure to be part of the ERC team of which everyone has in one way or another contributed to this thesis. Sofia, if was fun working with you and I will never forget our nightly fermentation adventures! Ewout, I am happy that you became my project buddy for the human and Yarrowia project. Thank you for all our insightful discussions about the project, for always staying (apparently?) calm, and last but not least, for being my paranymph! Eline, I admire your perseverance with the synthetic chromosomes and your critical eye for details. I enjoyed collaborating with you on the watermark paper and I am really happy to have you as paranymph on my side! Michał, I wish our ways would not have split so early in the project. Thank you for your support and for all the fun we had! Melanie, without all your hard work and thorough way of working the human glycolysis project would have moved on a lot slower. Thanks a lot for all your contributions to this thesis! Jordi, thank you for providing me with the last results I needed to finish my thesis! Mark, a great part of what I have learned about fermentations, fluorescent proteins and research in general I have learned from you. Thank you for your willingness to help and for your extremely clear way of explaining things! Furthermore I would like to thank all of you, including Paola and Charlotte as 'extended ERC members' for all the input during the ERC meetings.

IMB would not run so smoothly without four great staff technicians. Marcel thanks a lot for everything you explained me about the fascinating world of sequencing data analysis and all your contributions to the analysis of my data. Pilar, our 'queen bee' of the molecular lab, thank you for all the strains you sequenced for me, for keeping the molecular lab organised and above all, for your friendship. I wish you all the best in Spain! Erik, thank you for everything you explained me about fermenters and for your willingness to always think along and help to find a solution for experimental challenges. Also your thinking along with finding a new job was very much appreciated. Marijke, I lost count how many days we spent in the Mickel's lab measuring glycolysis. Thanks a lot for all your help! I enjoyed our conversations and your encouragement to self-reflect definitely helped me to move in the right direction sometimes. Michelle, as management assistant you were very approachable, which together with your very pro-active attitude definitely helped for a smooth planning of all our meetings.

During the past four years I had the pleasure to supervise six students. In chronological order: Rik, Lycka, Ingeborg, Koen, Agnes and Liset, although a large part of your work did not end up in my thesis you made very valuable contribution to the project and I enjoyed working with you. Hoping that you learned something from me, I definitely learned a lot from you as well.

All my experiments would not have been possible without the ladies from MSD: Jannie, Astrid and Apilena thank you for providing us with everything we needed for the experiments, for making the autoclaves run smoothly and for always being so friendly and flexible!

During my PhD I had the opportunity to collaborate with several people from outside our group. I would like to thank Marcel Vieira Lara and Prof. Barbara Bakker for their input in the human glycolysis project and for the nice days I spent working in Groningen. Also, I would like to thank Prof. Bas Teusink for our inspiring brainstorm meetings. In another project, which did not make it into my thesis, we collaborated with dr. Lisset Flores and I would like to thank her for providing the Yarrowia material and her input in the project.

The past four years included many long days and tough moments, but this was definitely compensated by the very nice working atmosphere at IMB. I really felt part of the group over the years which is really thanks to every single one of you! Thanks for all your support, the nice evenings in Keldertje and the Botanical garden, parties, dinners, game nights, art nights etc.! Over the years I shared my office with a lot of people who contributed to a lot of fun, help, advice and new plants. Thanks Charlotte, Jasmijn, Jordi, Ewout, Jasmine, Arthur and several students for being my office mates. Arthur, your efforts to get everyone involved in the group really contributed to make IMB from the start a very nice place to work for me. Also I want to thank you for your support and for all the nice dinners you cooked for me :)

The Flamingo y Flamenco chicas, Pilar, Susan, Sofia and Melanie, thank you for a great holiday to Spain! I keep fond memories of visiting the Feria, spotting flamingos and drinking orange wine. Pilar, you were a great guide! To stay in holiday atmosphere, Sofia, Susan, Pilar, Arthur, Erik, also our mini holiday to the Irish coast after the Cork Conference was very nice. The dolphin and the puffins made up for one of the most awful boat trips of my life ;)

De afgelopen jaren heb ik niet alleen op wetenschappelijk gebied veel geleerd. Louise van Swaaij en Lucienne Willems bedankt voor alle waardevolle inzichten waar ik dagelijks veel aan heb.

Aan het begin van mijn PhD werd ik lid van Krashna Musica. Ik wil het koor en met name de alten bedanken voor alle fijne woensdagavonden met mooie muziek, fijne gesprekken en gezelligheid, wat enorm hielp om even uit de PhD bubbel te ontsnappen. Niet alleen bij Krashna, maar ook daarbuiten heb ik de afgelopen jaren veel muziek gemaakt. Alle lieve, enthousiaste en inspirerende mensen met wie ik dit heb mogen doen, dank jullie wel!

Verder wil ik graag al mijn vrienden bedanken voor alle gezelligheid, muziek, post, knuffels en steun de afgelopen jaren. Jullie maken mijn leven stuk voor stuk een beetje mooier en gaven mij nieuwe energie om mijn PhD af te maken. Drie mensen wil ik nog even in het bijzonder noemen: Akke, dankjewel voor je oneindige zorgzaamheid, dat betekent veel voor me. Bram, in dit proefschrift kan je eindelijk lezen wat voor een gistjes ik nou eigenlijk al die jaren bij elkaar heb gepipetteerd! Hilje, jij blijft mijn favoriete fluitmaatje!

Marianne, een mooie bijkomstigheid van een PhD in Delft doen was dat we weer dicht bij elkaar kwamen te wonen. Ik heb ervan genoten om zo vaak bij je te komen eten op woensdag in je paleis aan het Mijnbouwplein en elkaar elke week bij Krashna te zien. Verder ben ik heel blij met jouw bijdrage aan dit proefschrift in de vorm van de cover!

Lieve mam en pap, dankjewel voor jullie steun en zorgzaamheid afgelopen jaren. Het was fijn om te weten altijd welkom te zijn in Arnhem om weer even op te laden, van de tuin te genieten en samen prachtige wandelingen op de Veluwe te maken.

Lieve Thomas, zonder jou had ik dit niet gekund. Dankjewel voor alles! Er zijn denk ik weinig niet-biologen die zoveel van microbiologie weten ;). Ik kijk uit naar onze nieuwe avonturen samen! Ik hou van je.

Curriculum Vitae

Francine Judith Boonekamp was born on the 26th of October 1990 in Leidschendam, the Netherlands. She grew up in Arnhem where she finished pre-university education in 2009 at the Stedelijk Gymnasium Arnhem. Subsequently, Francine enrolled in the Bachelor program Biology at Utrecht University where she, besides the curricular study program, actively participated in the biology honours program. She performed her Bachelor thesis in the Plant Ecophysiology group of Prof. R. Voesenek under supervision of dr. Lot Gommers. At the same time, Francine became member of several orchestras and ensembles where she played flute and bassoon and was active in the board and several committees. After obtaining her bachelor degree in 2013, Francine enrolled in the Master program Environmental Biology, also at Utrecht University. The first year of the program was completed in the Plant-Microbe Interactions group of Prof. Corné Pieterse under supervision of dr. Iván Fernández-López, studying the molecular interactions of non-host plants and arbuscular mycorrhizal fungi. The second year of the master program was completed in Wageningen where she followed several biotechnology courses at Wageningen University followed by a six months research internship at the biotechnology company Dyadic. In 2015 Francine obtained her Master degree (with honours) and started a PhD project at TU Delft in the Industrial Microbiology group under supervision of Prof. Pascale Daran-Lapujade and Prof. Jack Pronk, of which the results are described in this thesis. As from August 2020 Francine is working as Scientist upstream process development at Janssen Vaccines in Leiden.

List of publications

- 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., Daran-Lapujade, P. Pathway swapping: Toward modular engineering of essential cellular processes. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 15060-15065 (2016)

- **Boonekamp, F. J**., Dashko, S., van den Broek, M., Gehrmann, T., Daran, J. M., & Daran-Lapujade, P. The genetic makeup and expression of the glycolytic and fermentative pathways are highly conserved within the *Saccharomyces* genus. *Frontiers in genetics*, *9*, 504 (2018)

- **Boonekamp, F. J**., Dashko, S., Duiker, D., Gehrmann, T., van den Broek, den Ridder, M., Pabst, M., Robert, V., Abeel, T., Postma, E.D., Daran, J. M., Daran-Lapujade, P. Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish near-identical DNA and RNA sequences. *ACS Synthetic biology, 9,* 1361–1375 (2020)

- **Boonekamp, F. J**., Knibbe, E., Vieira-Lara, M. A., Wijsman, M., van Eunen, K., Bron, R., Almonacid Suarez, A.M., Harmsen, M.C., van Rijn, P., Wolters, K., Bakker, B., Daran, J. M., Daran-Lapujade, P. (2020). A yeast with muscle doesn't run faster: full humanization of the glycolytic pathway in *Saccharomyces cerevisiae. Manuscript in preparation for submission*