

## Genome analysis and engineering of industrial lager brewing yeasts

Gorter de Vries, Arthur

**DOI**

[10.4233/uuid:7f6985bb-3383-48f6-bcd3-f2758f35e3c2](https://doi.org/10.4233/uuid:7f6985bb-3383-48f6-bcd3-f2758f35e3c2)

**Publication date**

2019

**Document Version**

Final published version

**Citation (APA)**

Gorter de Vries, A. (2019). *Genome analysis and engineering of industrial lager brewing yeasts*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:7f6985bb-3383-48f6-bcd3-f2758f35e3c2>

**Important note**

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

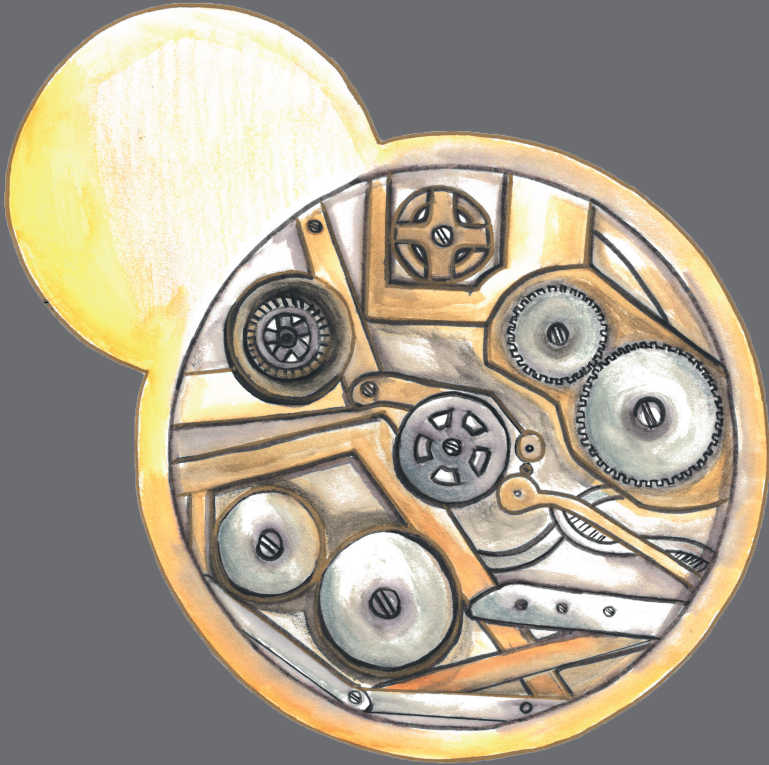
**Copyright**

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.

# Genome analysis and engineering of industrial lager brewing yeasts



Arthur R. Gorter de Vries



# **Genome analysis and engineering of industrial lager brewing yeasts**

## **Dissertation**

for the purpose of obtaining the degree of doctor  
at Delft University of Technology  
by the authority of the Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen;  
Chair of the Board for Doctorates  
to be defended publicly on  
Friday 6 September 2019 at 10:00 o'clock

By

**Arthur Roelof GORTER DE VRIES**

Master of Science in Biotechnology, ETH Zürich, Switzerland  
born in Voorburg, The Netherlands

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof. dr. J.T. Pronk	Delft University of Technology, promotor
dr. ir. J.-M.G. Daran	Delft University of Technology, promotor

Independent members:

Prof. dr. P.A.S. Daran-Lapujade	Delft University of Technology
Prof. dr. ir. D. de Ridder	Wageningen University & Research
Prof. dr. E.J. Louis	University of Leicester
dr. H. Bachmann	Vrije Universiteit Amsterdam
dr. G. Liti	Institute for Research on Cancer and Aging, Nice

Reserve member:

Prof. dr. U. Hanefeld	Delft University of Technology
-----------------------	--------------------------------

The research presented in this thesis was performed at the Industrial Microbiology Section, Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, the Netherlands. The project was financed by the BE-Basic R&D Program (<http://www.be-basic.org/>), which was granted a TKI subsidy from the Dutch Ministry of Economic Affairs, Agriculture and Innovation (EL&I). Research was performed in BE-Basic flagship FS10 in collaboration with HEINEKEN Supply Chain B.V. (Zoeterwoude, the Netherlands).

Cover	Pilar de la Torre Cortés
Layout	Arthur Gorter de Vries
Printed by	ProefschriftMaken     <a href="http://www.proefschriftmaken.nl">www.proefschriftmaken.nl</a>
ISBN	978-94-6380-407-3

© 2019 Arthur R. Gorter de Vries

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

Foreword .....	5
Summary .....	7
Samenvatting.....	11
Chapter 1: Introduction.....	15
Chapter 2: Industrial relevance of chromosomal copy number variation in <i>Saccharomyces</i> yeasts....	35
Chapter 3: Nanopore sequencing enables near-complete <i>de novo</i> assembly of <i>Saccharomyces cerevisiae</i> reference strain CEN.PK113-7D.....	55
Chapter 4: Nanopore sequencing and comparative genome analysis confirm lager-brewing yeasts originated from a single hybridization .....	71
Chapter 5: CRISPR-Cas9 mediated gene deletions in lager yeast <i>Saccharomyces pastorianus</i> .....	97
Chapter 6: Allele-specific genome editing using CRISPR-Cas9 is associated with loss of heterozygosity in diploid yeast .....	123
Chapter 7: Phenotype-independent isolation of interspecies <i>Saccharomyces</i> hybrids by dual-dye fluorescent staining and fluorescence-activated cell sorting.....	139
Chapter 8: Laboratory evolution of a <i>Saccharomyces cerevisiae</i> x <i>S. eubayanus</i> hybrid under simulated lager-brewing conditions.....	155
Chapter 9: <i>In vivo</i> recombination of <i>Saccharomyces eubayanus</i> maltose-transporter genes yields a chimeric transporter that enables maltotriose fermentation .....	179
Chapter 10: Outlook .....	205
Acknowledgments .....	209
Curriculum Vitae.....	215
List of Publications.....	217



## Foreword

This thesis emanates from a deep-rooted desire to understand the world surrounding me, which I believe to have obtained primarily from my grandfather. Therefore, I dedicate this thesis to dr. ir. Hendrik de Vries, a man of exceptional scientific and intellectual curiosity.

Born in 1920 in Amsterdam, he came to Delft to study Physics in 1938 and became a member of the *Delftsch Studenten Corps*. During the Second World War, he refused to sign a declaration of loyalty to the occupying forces, interrupted his studies and sought to escape to England to join the war effort. After several months of travelling and of imprisonment in Spain, he boarded a ship to England from Gibraltar and joined the Royal Air Force as an engineer. After the war, he completed his studies and married Janske Gorter in 1949. Fascinated by science, he pursued a PhD on the optical and rheological properties of nylon and other fibres while working at the company currently called AkzoNobel. He successfully defended his doctoral thesis in Delft in 1953 and was inventor on a patent concerning an optical aspect for phase-contrast microscopy.

I remember my grandfather as a man of exceptionally broad interests, who loved to travel and to spend time in nature: when he visited us in Germany or France, we often went to open-air archaeological sites and to museums depicting tales of more adventurous times. His home was an exciting treasure trove of artefacts and memorabilia of his travels to understand past civilizations and present cultures across the globe. My grandfather was a true engineer: he enjoyed taking apart broken things in his atelier, and could repair anything. I have always felt great interest and affection from my grandparents, and the close bonds of our family remain long after their death.

When my grandfather died in 2009, I felt an overwhelming sadness. All the knowledge and wisdom he had accumulated over the years was gone, and I would never be able to hear all the wonderful stories he still had to tell. In those days, I also noticed the many similarities between my father and him. I became more aware of how my father's desire to understand the world we live in had fuelled my own curiosity, of how my father's passion for his work drew me to study the wonders of life, and of how his patience to explain every concept I didn't understand made me love science. My father is the calm and appeasing source of knowledge and wisdom that my grandfather used to be, and my grandfather continues to live through him; I admire both tremendously.

Without the role model of my grandfather, I would not have been who I am today and I would not have pursued this PhD. While he cannot attend my doctoral defence, he will still be somewhat present: my grandfather and I share the same size, therefore I intend to wear the tail suit that he wore at his doctoral ceremony during the defence of this thesis.





## Summary

Lager beer, also referred to as Pilsner, is the most popular alcoholic beverage in the world, with an annual consumption of almost 200 billion litres per year. To make lager beer, brewer's wort is fermented with the yeast *Saccharomyces pastorianus*. This microorganism converts wort sugars into ethanol and contributes key flavour compounds to the beer. *S. pastorianus* is an interspecific hybrid which likely formed about 500 years ago by spontaneous mating between an ale-brewing *S. cerevisiae* strain and a wild *S. eubayanus* contaminant.

The genome of lager brewing yeast is exceptionally complex: not only does it contain chromosomes from the two parental species, but these have also undergone extensive recombination and are present in varying copy numbers, a situation referred to as aneuploidy. The *S. eubayanus* ancestor was only discovered in 2011, enabling an improved understanding of the complex genome and convoluted evolutionary ancestry of *S. pastorianus*. Furthermore, recent advances in whole-genome sequencing technology and in gene editing tools have simplified the genetic accessibility and amenability of *Saccharomyces* yeast genomes. The aim of this thesis was to leverage these advances to investigate how the genetic complexity of current *S. pastorianus* strains emerged and how it contributes to industrial lager brewing performance, and to develop new methods for strain improvement of brewing yeasts.

**Chapter 1** outlines the state-of-the-art of *S. pastorianus* genetics and strain improvement techniques. After reviewing the emergence of the lager brewing industry, recent insights in the genetics and evolutionary origin of *S. pastorianus* are discussed. Specific attention is given to the genetic differences between Group 1 and 2 lager brewing strains, and the ongoing discussion on how they may have emerged. Moreover, the mechanisms by which parental subgenomes can interact in hybrids are reviewed, as well as the way in which they contribute to the brewing performance of *S. pastorianus*. Recent progress in whole-genome sequencing and gene-editing technologies has increased the genetic accessibility of *S. pastorianus*, which resulted in new possibilities for strain improvement. Research on *Saccharomyces* yeasts has provided a large array of strain improvement methods such as mutagenesis and selection, laboratory evolution and gene-editing, which can be applied to lager brewing yeast. Moreover, recent research has shown the potential of laboratory-made hybrids to expand the limited genetic and phenotypic diversity of lager brewing yeasts.

To further explore aneuploidy and chromosome copy number diversity in *S. pastorianus*, **Chapter 2** reviews the origin, impact and industrial relevance of deviating chromosome copy number in *Saccharomyces* yeasts. Aneuploidy is widespread among industrial yeast strains, particularly in lager brewing yeasts. Moreover, it frequently emerged during laboratory evolution in response to specific selective pressures. Two major effects of aneuploidy can be distinguished: chromosome-independent effects, which cause a general aneuploidy-associated stress response, and chromosome-specific effects of copy number, which cause complex phenotypes that result from the cumulative effect of copy number changes of all affected genes. Aneuploidy and chromosome copy number already contribute to many industrially beneficial traits of *Saccharomyces* yeasts and their modification offers interesting opportunities for industrial strain improvement and diversification programmes.

Whole-genome sequencing of *S. cerevisiae* has been initiated 25 years ago with reference strain S288C. The rapid development of next generation sequencing methods intensified its application

over the last decade, however, whole genome sequencing mostly relied on short-read sequencing. Due to the presence of repetitive sequences, the ability of short-read sequencing to resolve *Saccharomyces* genomes is limited, resulting in incomplete and fragmented *de novo* assemblies. Long sequence reads can span such repetitive regions and thereby enable the reconstruction of complete chromosomes. Recent advances in nanopore sequencing technology have resulted in substantial increases in read length and in improvements of their low sequencing accuracy. In **Chapter 3**, nanopore sequencing was applied to the *S. cerevisiae* strain CEN.PK113-7D, resulting in a near-complete chromosome-level assembly. The resulting assembly contained about 5 % previously unassembled genes, notably in the industrially relevant subtelomeric regions. In addition to providing a high-quality reference genome of one of the two parental species of *S. pastorianus*, availability of the CEN.PK113-7D sequence was of particular value to research on lager brewing hybrids, as it has been used to construct laboratory-made *S. cerevisiae* x *S. eubayanus* strains. In **Chapter 4**, nanopore sequencing was applied to the *S. pastorianus* strain CBS 1483. Added sequences relative to previous genome assemblies were analysed and the ability of nanopore sequencing to resolve complex alloaneuploid genomes was assessed. While the obtained assembly was a consensus sequence, structural and sequence heterogeneity could be recovered by alignment of long- and short-reads, respectively. The chromosome level-assembly was used to study the ancestry of *S. pastorianus* using Alpaca, a newly developed algorithm to analyse non-linear genome similarity across large datasets. Because our results revealed an absence of ancestry differences between Group 1 and Group 2 strains, they support ancestry from a single hybridization event.

The advent of RNA-programmed endonucleases such as Cas9 has revolutionized gene editing in *Saccharomyces cerevisiae*, however, such tools were not immediately transferable to *S. pastorianus*. In **Chapter 5**, a broad-host-range Cas9 gene-editing tool for yeast genomes was developed based on plasmid-based co-expression of the *Streptococcus pyogenes* Cas9 endonuclease and of a guide-RNA molecule (gRNA) that confers target-sequence specificity. To circumvent gRNA expression issues, the gRNA was flanked with Hammerhead and Hepatitis Delta Virus ribozymes and expressed using the RNA polymerase II-dependent *TDH3* promoter. The newly developed editing tool was applied successfully to *S. pastorianus* and enabled simultaneous deletion of all four alleles of the *SeLV6* gene in CBS 1483. Moreover, multiplexed gRNA expression enabled simultaneous deletion of all *SeATF1* and *SeATF2* alleles in *S. pastorianus* strains CBS 1483 and Weihenstephan 34/70. In further applications of the newly developed CRISPR-Cas9 tool, low gene-editing efficiency and unwanted loss of heterozygosity were observed in heterozygous genomes. **Chapter 6** describes how heterozygosity affects double-strand-break mediated gene-editing. Targeting of the *S. cerevisiae*-specific gene *MAL11* in a laboratory-made *S. cerevisiae* x *S. eubayanus* hybrid invariably resulted in loss of the targeted *S. cerevisiae* chromosome arm and in its replacement with an additional copy of the homologous *S. eubayanus* chromosome arm. Cas9 targeting of homozygous and heterozygous targets in a diploid laboratory *S. cerevisiae* strain confirmed that allele-specific introduction of a double-strand break resulted in low gene-editing efficiency and in extensive loss of heterozygosity. Whole-genome sequencing of a highly heterozygous *S. cerevisiae* diploid after allele-specific targeting identified systematic loss of heterozygosity, affecting up to several hundred thousand base pairs. The presence of mosaic heterozygosity indicated that these mutations emerged by homologous recombination. As this mechanism is highly conserved in higher eukaryotes and critical to gene-editing, we propose that allele-specific gene editing could cause extensive loss of

heterozygosity in a broad range of organisms, and therefore warrant caution, particularly in the context of human gene therapy.

Laboratory-made hybrids are of interest for industrial applications due to their wide phenotypic diversity and their superior performance relative to parental species, a phenomenon that is known as hybrid vigour. With the discovery of *S. eubayanus*, research into laboratory-made hybrids for lager brewing has gained momentum as well. However, current methods for interspecific mating rely on selectable phenotypes to recover hybrids. Such phenotypes may occur naturally or may be introduced prior to mating, either via gene-editing or via non-GMO methods. To allow the mating of a wide array of strains without introducing mutations, we developed a method based on fluorescent labelling of parental species and on flow cytometric sorting of dual-stained hybrid cells, described in **Chapter 7**. We applied this method to intraspecific mating of *S. cerevisiae* haploids, to interspecific mating of *S. cerevisiae* haploids and *S. eubayanus* spores and to rare mating between *S. cerevisiae* and *S. eubayanus* strains of various ploidies.

To evaluate the applicability of laboratory-made hybrids for lager brewing applications, and to investigate the evolution of a hypothesized ancestral *S. pastorianus* hybrid, we subjected a haploid *S. cerevisiae* x *S. eubayanus* hybrid to laboratory evolution under simulated lager brewing conditions (**Chapter 8**). After up to 418 generations in repeated batch fermentations on industrial brewer's wort, 55 single-cell colonies were isolated, and characterized phenotypically and genotypically. While some brewing-relevant phenotypes such as flocculation and maltotriose utilization were altered, the isolates did not acquire large-scale aneuploidy, as is present in *S. pastorianus* strains. Instead, isolates were mostly affected by loss of heterozygosity mediated by non-copy-number conservative recombinations between both subgenomes. The genetic changes acquired during lager brewing were consistent with a common origin of naturally-occurring *S. pastorianus* strains, and indicated sufficient stability of laboratory-made hybrids for industrial application.

In addition to their use for the generation of laboratory-made hybrids, recently discovered *S. eubayanus* strains have already been applied directly to industrial-scale lager brewing. However, the absence of maltotriose utilisation resulted in high residual sugar concentrations and in low ethanol titres. In **Chapter 9**, we therefore subjected an industrial *S. eubayanus* strain to UV-mutagenesis and to laboratory evolution in a maltotriose-limited chemostat culture. In the resulting maltotriose-utilising strains, whole-genome sequencing revealed a recombination between three maltose-transporter genes, which resulted in a new chimeric open reading frame. Reverse engineering of the new, recombinant *MALT413* gene confirmed that it encoded a functional maltotriose transporter. The sequence of the *S. pastorianus* maltotriose transporter gene *MTY1* is consistent with a similar evolutionary history, involving recombination between maltose transporter genes from *S. cerevisiae*, *S. eubayanus* and *S. paradoxus*. The emergence of a novel gene function, as a result of random mutagenesis and a short period of laboratory evolution, is a unique illustration of how recombination can facilitate neofunctionalization.



## Samenvatting

Lagerbier, ook wel Pilsener genoemd, is de populairste alcoholische drank ter wereld, met een verbruik van bijna 200 miljard liter per jaar. Om lagerbier te maken, wordt gerstwort gefermenteerd met de gist *Saccharomyces pastorianus*. Dit micro-organisme zet wortsuikers om in ethanol en produceert belangrijke smaakstoffen in het hierbij gevormde bier. *S. pastorianus* is een hybride die ongeveer 500 jaar geleden waarschijnlijk gevormd is door spontane kruising van een *S. cerevisiae*-stam die gebruikt werd om bier te brouwen en een wilde *S. eubayanus*-contaminant.

Het genoom van pilsgist is uitzonderlijk complex: het bevat niet alleen chromosomen van de twee oudersoorten, maar deze zijn ook vaak gerecombineerd en verschillen in het aantal kopieën, een situatie die wordt aangeduid als aneuploidie. De voorouder van *S. eubayanus* werd pas in 2011 ontdekt en maakte een beter begrip van het complexe genoom en de ingewikkelde evolutionaire geschiedenis van *S. pastorianus* mogelijk. Bovendien hebben recente ontwikkelingen het bepalen van de DNA-volgorde van complete genomen en de genetische toegankelijkheid van *Saccharomyces*-genomen vergroot. Het doel van dit proefschrift is om deze ontwikkelingen te gebruiken om te onderzoeken hoe de genetische complexiteit van de huidige *S. pastorianus*-stammen tot stand gekomen is, hoe deze bijdraagt aan de industriële productie van lager bier, en om nieuwe methoden te ontwikkelen voor verbetering van brouwgisten.

**Hoofdstuk 1** schetst de stand van zaken in *S. pastorianus*-genetica en technieken voor stam-verbetering van deze gist. Na een beschrijving van de opkomst van de pilsindustrie worden recente inzichten in de genetica en evolutionaire oorsprong van *S. pastorianus* besproken. Hierbij wordt specifiek aandacht besteed aan de genetische verschillen tussen Groep 1 en Groep 2 pilsstammen, en aan de nog steeds voortdurende discussie over hoe deze kunnen zijn ontstaan. Bovendien worden de mechanismen besproken waarmee de twee voorouderlijke subgenomen kunnen interageren in hybriden, evenals de manier waarop deze bijdragen aan de brouwprestaties van *S. pastorianus*. Recente vooruitgang in technieken voor DNA-sequentiebepaling en genmodificatie heeft de genetische toegankelijkheid van *S. pastorianus* vergroot, hetgeen heeft geresulteerd in nieuwe mogelijkheden voor stamverbetering. Onderzoek aan *Saccharomyces*-gisten heeft een groot aantal verschillende methoden opgeleverd voor het verbeteren van giststammen, zoals mutagenese en selectie, laboratoriumevolutie en genbewerking, die kunnen worden toegepast op pilsgist. Bovendien heeft recent onderzoek aangetoond dat nieuwe, in het laboratorium gemaakte hybriden de beperkte genetische en fenotypische diversiteit van biergisten significant kunnen uitbreiden.

Om aneuploidie en diversiteit in het aantal kopieën van chromosomen in *S. pastorianus* nader te onderzoeken, worden in **Hoofdstuk 2** de oorsprong, impact en industriële relevantie van deze kopie-aantallen in *Saccharomyces*-gisten besproken. Aneuploidie komt in industriële giststammen, en met name in pilsgisten, veelvuldig voor. Bovendien wordt aneuploidie vaak aangetroffen tijdens laboratoriumevolutie, als aanpassing op verschillende vormen van selectiedruk. Er zijn twee belangrijke effecten van aneuploidie te onderscheiden. Chromosoom-onafhankelijke effecten veroorzaken een algemene aneuploidie-geassocieerde stressreactie. Daarnaast leiden chromosoom-specifieke effecten van het aantal kopieën tot complexe fenotypen, die resulteren van het cumulatieve effect van de veranderde kopie-aantallen van alle betrokkenen genen. Aneuploidie en het aantal chromosoom kopieën dragen nu al bij aan veel industrieel gunstige eigenschappen van

*Saccharomyces*-gisten en hun modificatie biedt interessante kansen voor industriële stamverbetering- en diversificatieprogramma's.

Hoewijl DNA-sequentiebepalingstechnologie al meer dan een decennium lang intensief wordt gebruikt in gistgenetica, was dit tot voor kort vooral afhankelijk van *short-read* technologie. Vanwege de aanwezigheid van repetitieve sequenties is het vermogen van *short-read* technologie om *Saccharomyces*-genomen te ontrafelen beperkt, waardoor genomen slechts onvolledig en gefragmenteerd in kaart gebracht worden. Technologieën die langere DNA-fragmenten kunnen aflezen, kunnen dergelijke zich herhalende gebieden wel ontrafelen en daardoor de reconstructie van complete chromosomen mogelijk maken. Recente ontwikkelingen in nanoporie-sequentiebepalingstechnologie hebben geresulteerd in een aanzienlijke toename van de leeslengte en in verbeteringen van de nauwkeurigheid. In **Hoofdstuk 3** wordt nanoporie-technologie toegepast op de *S. cerevisiae* stam CEN.PK113-7D. Dit onderzoek resulteerde in een bijna complete genomkaart die ongeveer 5 % nieuw beschreven genen bevat, met name in de industrieel relevante subtelomerische gebieden. Naast het leveren van een hoogwaardig referentiegenoom van een van de twee ouderlijke soorten van *S. pastorianus* is de beschikbaarheid van de CEN.PK113-7D-sequentie van bijzonder belang voor het onderzoek naar pilsgist, omdat deze stam is gebruikt als ouder van een eerder aan de TU Delft gemaakte laboratoriumhybride van *S. cerevisiae* en *S. eubayanus*. In **Hoofdstuk 4** wordt nanoporie-technologie toegepast op de *S. pastorianus*-stam CBS 1483. Nieuw toegevoegde sequenties ten opzichte van eerdere genomkaarten zijn geanalyseerd en het vermogen van nanoporie-technologie om complexe alloaneuploïde genomen te ontrafelen is bepaald. Hoewel dit onderzoek een consensussequentie heeft gegenereerd, kan de heterogeniteit van structuur en sequentie worden achterhaald door uitlijning van respectievelijk nanoporie- en *short-read* data. De nieuwe genomkaart is gebruikt om de voorgeschiedenis van *S. pastorianus* te bestuderen met behulp van Alpaca, een nieuw algoritme voor het analyseren van niet-lineaire genomovereenkomsten in grote datasets. Omdat onze resultaten een afwezigheid van voorgeschiedenisverschillen tussen Groep 1 en Groep 2 stammen onthullen, ondersteunen ze de hypothese dat alle huidige stammen van *S. pastorianus* zijn voortgekomen uit één enkele hybridisatiegebeurtenis.

De opkomst van het gebruik van sequentie-specifieke endonucleasen, zoals Cas9, heeft een revolutie teweeggebracht in het bewerken van genen in *Saccharomyces cerevisiae*. Dergelijke hulpmiddelen waren echter niet onmiddellijk toepasbaar op *S. pastorianus*. In **Hoofdstuk 5** wordt een gen-modificatiehulpmiddel gebaseerd op Cas9 ontwikkeld, dat geschikt is voor modificatie van diverse gistgenomen. Dit systeem is gebaseerd op een plasmide dat zorgde voor co-expressie van de *Streptococcus pyogenes* cas9 endonuclease en van een gids-RNA molecuul (gRNA) dat doelwit-sequentie-specificiteit verleent. Om problemen met gRNA-expressie te omzeilen, wordt het gRNA geflankeerd met Hammerhead- en Hepatitis Delta-virus-ribozymen en tot expressie gebracht vanaf een RNA-polymerase II-afhankelijke *TDH3*-promotor. Dit nieuwe hulpmiddel is met succes toegepast op *S. pastorianus* om gelijktijdige deletie van alle vier allelen van het *SeLV6*-gen in CBS 1483 te bereiken. Bovendien maakte gelijktijdige expressie van verschillende gRNA's simultane verwijdering mogelijk van alle *SeATF1*- en *SeATF2*-allelen in *S. pastorianus*-stammen CBS 1483 en Weihenstephan 34/70. Bij verdere toepassingen van dit Cas9-hulpmiddel in heterozygote gistgenomen worden een lage efficiëntie van genmodificatie en, bovendien, ongewenst verlies van heterozygositeit waargenomen. **Hoofdstuk 6** beschrijft de invloed van heterozygositeit op genmodificatietechnieken waarbij gericht dubbelstrengs breuken van DNA worden aangebracht.

Pogingen om het *S. cerevisiae*-specifieke gen *MAL11* te modificeren in een laboratoriumhybride van *S. cerevisiae* en *S. eubayanus* resulteerden in verlies van de chromosoomarm waarop dit *S. cerevisiae*-doelwit zich bevond en in de vervanging ervan door een extra kopie van het homologe *S. eubayanus*-chromosoom. Experimenten op andere homozygote en heterozygote doelen in een diploïde laboratoriumstam van *S. cerevisiae* bevestigen dat allelspecifieke introductie van een dubbelstrengs breuk resulteert in lage efficiëntie van genmodificatie en in uitgebreid verlies van heterozygositeit. Genoomsequentiebepaling van een sterk heterozygote *S. cerevisiae*-diploïde na pogingen tot allelspecifieke genmodificatie tonen een systematisch verlies van heterozygositeit aan, die tot honderdduizenden basenparen in het DNA beïnvloedden. De aanwezigheid van mozaïekpatronen in de heterozygositeit van deze gebieden van het genoom laat zien dat deze mutaties ontstaan zijn door homologe recombinatie. Omdat dit mechanisme sterk geconserveerd is in hogere eukaryoten en cruciaal is voor genmodificatie, wordt geconcludeerd dat allelspecifieke genmodificatie hoogstwaarschijnlijk ook uitgebreid verlies van heterozygositeit kan veroorzaken in een breed scala van andere organismen, en dat daarom voorzichtigheid vereist is, in het bijzonder bij menselijke genterapie.

Laboratoriumhybriden zijn van belang voor industriële toepassingen vanwege hun brede fenotypische diversiteit en hun superieure prestaties ten opzichte van de oudersoorten. Met de ontdekking van *S. eubayanus* is ook het onderzoek naar laboratoriumhybriden voor het brouwen van pils toegenomen. Huidige methoden voor kruising van verschillende *Saccharomyces*-stammen en soorten zijn echter afhankelijk van selecteerbare fenotypen om hybriden te kunnen isoleren. Dergelijke fenotypen kunnen van nature bestaan of kunnen voorafgaand aan kruising worden geïntroduceerd via genmodificatie of via niet-GMO-werkwijzen voor het aanbrengen van veranderingen in het DNA. Om kruisingen tussen een breed scala van stammen mogelijk te maken, is in dit promotieonderzoek een methode ontwikkeld op basis van fluorescente kleuring van stammen en flowcytometrische sortering van dubbel gekleurde hybride cellen. Deze nieuwe methode, die wordt beschreven in **Hoofdstuk 7**, is met succes toegepast voor kruising van *S. cerevisiae*-haploïden en van *S. cerevisiae*- en *S. eubayanus*-stammen van verschillende ploïdie.

Om de mogelijke toepasbaarheid van laboratoriumhybriden voor het brouwen van bier te evalueren en om de evolutionaire historie van een veronderstelde voorouderlijke *S. pastorianus*-hybride beter te kunnen interpreteren, is een allohaploïde *S. cerevisiae* x *S. eubayanus*-hybride onderworpen aan laboratoriumevolucie onder gesimuleerde brouwomstandigheden (**Hoofdstuk 8**). Na 418 generaties in herhaalde fermentaties op industrieel gerstwort zijn 55 isolaten fenotypisch en genotypisch gekarakteriseerd. Hoewel enkele voor het bierbrouwen relevante fenotypen zoals flocculatie en maltotriosegebruik veranderd zijn, hebben de isolaten geen grootschalige aneuploidie ontwikkeld, zoals deze wel aanwezig is in *S. pastorianus*-stammen. In plaats hiervan vertonen de isolaten verlies van heterozygositeit door recombinatie tussen de beide subgenomen. De genetische veranderingen die verkregen zijn tijdens dit modelexperiment waren consistent met een gemeenschappelijke oorsprong van natuurlijk voorkomende *S. pastorianus*-stammen en duiden op voldoende stabiliteit van laboratoriumhybriden voor industriële toepassing.

Naast hun gebruik voor het genereren van laboratoriumhybriden, zijn *S. eubayanus*-stammen ook al zelf gebruikt voor het brouwen van pils op industriële schaal. Hun onvermogen om maltotriose, een belangrijke suiker in wort, te vergisten resulteerde echter in hoge suiker en lage ethanol concentratie in het bier. In **Hoofdstuk 9** wordt daarom een industriële *S. eubayanus*-stam onderworpen aan



mutagenese met ultraviolette straling en daarna aan laboratoriumevolutie in een maltotriose-gelimeerde chemostaatcultuur. In de maltotriose-gebruikende stammen die uit dit experiment voortkomen, is een recombinatie tussen drie maltose-transportergenen geïdentificeerd, waardoor een nieuw chimeer gen ontstaan is. Expressie van het nieuwe, recombinante *MALT413*-gen in de ongeëvolueerde stam bevestigde dat dit chimere gen codeert voor een functionele maltotriose-transporter. De DNA-volgorde van het *S. pastorianus* maltotriose-transportergen *MTY1* geeft aan dat het waarschijnlijk een vergelijkbare evolutionaire geschiedenis heeft, waarbij een recombinatie tussen maltosetransportergenen van *S. cerevisiae*, *S. eubayanus* en *S. paradoxus* betrokken is geweest. De opkomst van een nieuwe genfunctie na willekeurige mutagenese en een korte periode van laboratoriumevolutie, is een unieke illustratie van hoe recombinatie de evolutie van genen kan vergemakkelijken en versnellen.

# Chapter 1: Introduction

## The emergence and industrialisation of lager brewing

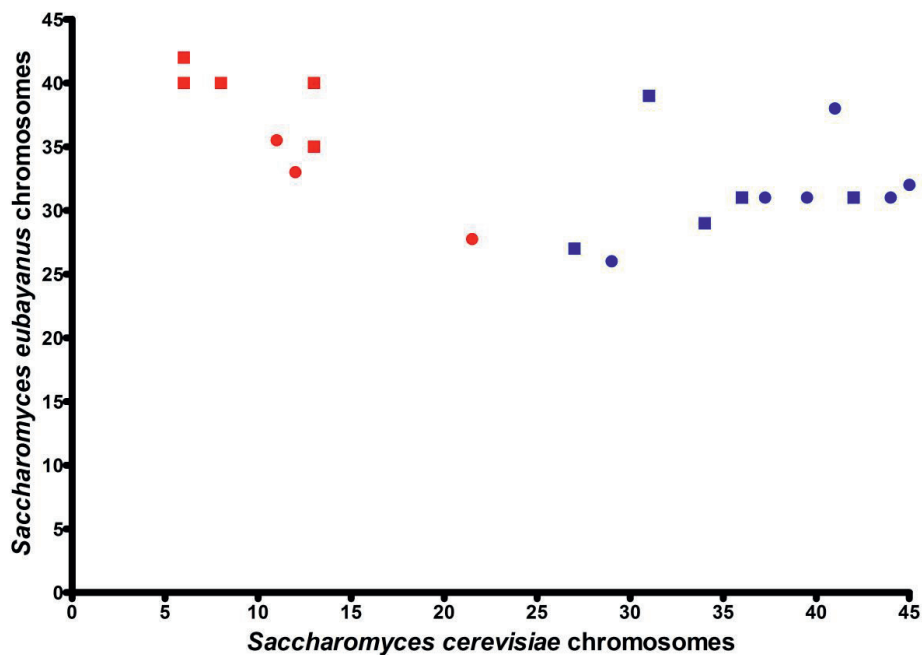
Beer brewing is tightly intertwined with human culture. Archaeological remains from the 12<sup>th</sup> millennium BC indicate that microbial fermentation of cereals may even predate the agricultural revolution (1). Chemical archaeology and pictographic evidence show that beer brewing was customary as early as the 4<sup>th</sup> millennium BC (2, 3). Lager-style beer only emerged in 16<sup>th</sup> century Bavaria under the influence of novel regulations to standardize the brewing process and to improve quality. For example, the well-known *Reinheitsgebot* of 1516 restricted the ingredients used for brewing to water, barley and hops (4). When, in 1553, beer brewing was legally restricted to winter months, bottom-fermenting yeast emerged as a consequence of the lower fermentation temperatures (5). In contrast to the top-fermenting yeasts used at higher temperatures for brewing ale-type beers, bottom-fermenting yeast form flocs, which sediment at the end of the fermentation (6). Bottom-fermenting yeasts were initially used to brew a dark brown beer, which was stored to enable consumption during the summer months. This beer was designated as lager, in reference to the German *lagern*, meaning to store (7). In 1842, the Bavarian brew master Josef Groll, working in the bohemian city Pilsen, started brewing a pale style of lager beer with fruity Saaz-type hops, which became known as Pilsner beer (7). The emergence of Pilsner coincided with rapid technological advances that enabled industrialization of beer brewing. Pasteur's 1876 discovery that yeast is responsible for fermentation (8) and the isolation of pure lager brewing strains by Hansen in 1883 (9,10), enabled inoculation of beer fermentations with pure cultures, resulting in more consistent quality. Moreover, the steam engine simplified production and transportation methods (11), the invention of ammonia refrigeration by Linde alleviated the need for natural ice to achieve low temperatures (12), the invention of iron moulds enabled industrial glass bottle production (13), the invention of beer filtration improved product stability after bottling (14) and the invention of crown corks enabled automated bottling (15). As a result of these innovations, global beer production soared to 17,7 billion litres in 1899 (16), and further increased to 193 billion litres in 2015, of which 89 % was lager-type beer (17).

## The lager brewing yeast *Saccharomyces pastorianus*

Lager beers are fermented with *Saccharomyces pastorianus* strains. These hybrids of *S. cerevisiae* and *S. eubayanus* have only been encountered in brewing-related contexts (18). *S. cerevisiae* has a long history of use in bakery, wine fermentation and brewing of ale-type beers and has been intensively studied for well over a century (19). In contrast, *S. eubayanus* was only discovered in 2011 (18). First isolated in South America, *S. eubayanus* was subsequently also found in North America, Asia and Oceania (20-22). While *S. eubayanus* sequences were detected in European samples by ITS sequencing (23), these may originate from *S. eubayanus* hybrids such as *S. pastorianus* or *S. bayanus*, rather than from a pure *S. eubayanus* isolate. Therefore, despite the European origin of lager brewing, presence of wild *S. eubayanus* strains has so far not been demonstrated in Europe. Currently, Tibetan *S. eubayanus* isolates have the highest degree of genetic identity to the *S. eubayanus*-derived genome sequences of *S. pastorianus* strains (21). Trade along the Silk Road has therefore been hypothesized to have enabled migration of *S. eubayanus* from Asia to the European birthplace of lager brewing (21). Alternatively, a now extinct or as yet undiscovered European *S. eubayanus* wild stock may be the ancestor of current *S. pastorianus* strains. *S. cerevisiae* x *S. eubayanus* hybrids recreated in the laboratory outcompeted their parental strains in

lager-brewing related environments by combining the fermentative vigour of *S. cerevisiae* with the low temperature optimum of *S. eubayanus* (24-26). These observations are consistent with the emergence of *S. pastorianus* by spontaneous hybridization between an ale-brewing *S. cerevisiae* strain and a wild *S. eubayanus* contaminant and with its subsequent dominance in lager-beer production.

In contrast to the genomes of laboratory-made hybrids, those of *S. pastorianus* strains are extensively aneuploid, with 45 to 79 chromosomes instead of an allodiploid complement of 32 chromosomes (Figure 1) (27-31). Two *S. pastorianus* subgroups were identified based on genetic differences: Group 1 strains ('Saaz') and Group 2 strains ('Frohberg'), that show marked differences in chromosome copy numbers (27, 32). While both have an approximately diploid *S. eubayanus* chromosome complement, the *S. cerevisiae* chromosome complement is approximately haploid in Group 1 strains and diploid or higher in Group 2 strains (Figure 1) (27-29). In addition, genome sequence comparison of the two groups revealed group-specific genes, substantial differences in subtelomeric regions and different frequencies of synonymous nucleotide mutations (32-34). While Group 1 strains display superior growth kinetics at low temperatures, their generally limited maltotriose utilization results in an overall inferior brewing performance relative to Group 2 strains (35).



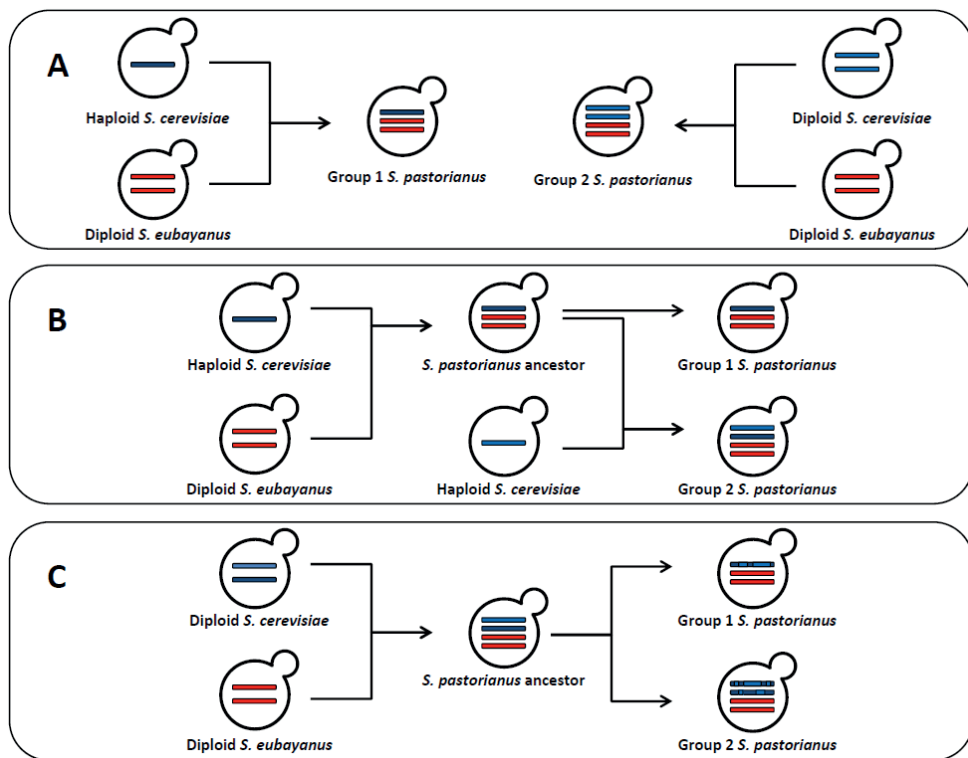
**Figure 1: Estimated chromosome copy numbers in *S. pastorianus* strains as determined by whole-genome sequencing.** Chromosome copy number estimates of various Group 1 (red) and Group 2 (blue) strains were estimated from short-read sequencing data published by Van den Broek *et al.*, 2015 (circles) and Okuno *et al.*, 2016 (squares) (28, 29). For each strain, the estimated total number of chromosomes derived from *S. eubayanus* is plotted against the estimated total number of chromosomes derived from *S. cerevisiae*. Due to copy number differences within individual chromosomes, copy number estimates should be interpreted as indicative.

### Evolutionary history of *S. pastorianus*: multiple hybridization events or man-made population bottlenecks?

Based on their phenotypic and genotypic differences, Group 1 and 2 strains were initially hypothesized to have emerged from two independent hybridizations (Figure 2A) (27, 36). Indeed, distinct haploid and diploid *S. cerevisiae* ancestors could explain the ploidy of Group 1 and 2 strains, respectively (37). However, identical recombinations between *S. cerevisiae* and *S. eubayanus* chromosomes were found at the *ZUO1*, *MAT*, *HSP82* and *XRN1/KEM1* loci in all Group 1 and 2 strains (29, 31, 38). When evolved under lager-brewing conditions, *S. cerevisiae* x *S. eubayanus* hybrids acquired a diverse range of interchromosomal recombinations, but these did not include those present in *S. pastorianus* strains, thus indicating a common hybrid ancestry of all current *S. pastorianus* strains (39).

Two theories have been forwarded to reconcile the evidence for a common ancestry of Group 1 and Group 2 strains with their genetic differences (Figure 2B): (i) Group 1 and 2 strains shared an initial hybridization event, with Group 2 strains resulting from a subsequent hybridization between the initial hybrid and a distinct *S. cerevisiae* strain, or (ii) Group 1 and 2 strains resulted from the same hybridization event involving a heterozygous *S. cerevisiae* ancestor, after which different paths of loss of heterozygosity and loss of genetic material caused the two Groups to diverge (29). Long-read nanopore sequencing and comparative genome analysis indicated that the *S. cerevisiae* genetic material is highly similar in both groups, thereby reducing the likelihood of multiple hybridization events (40).

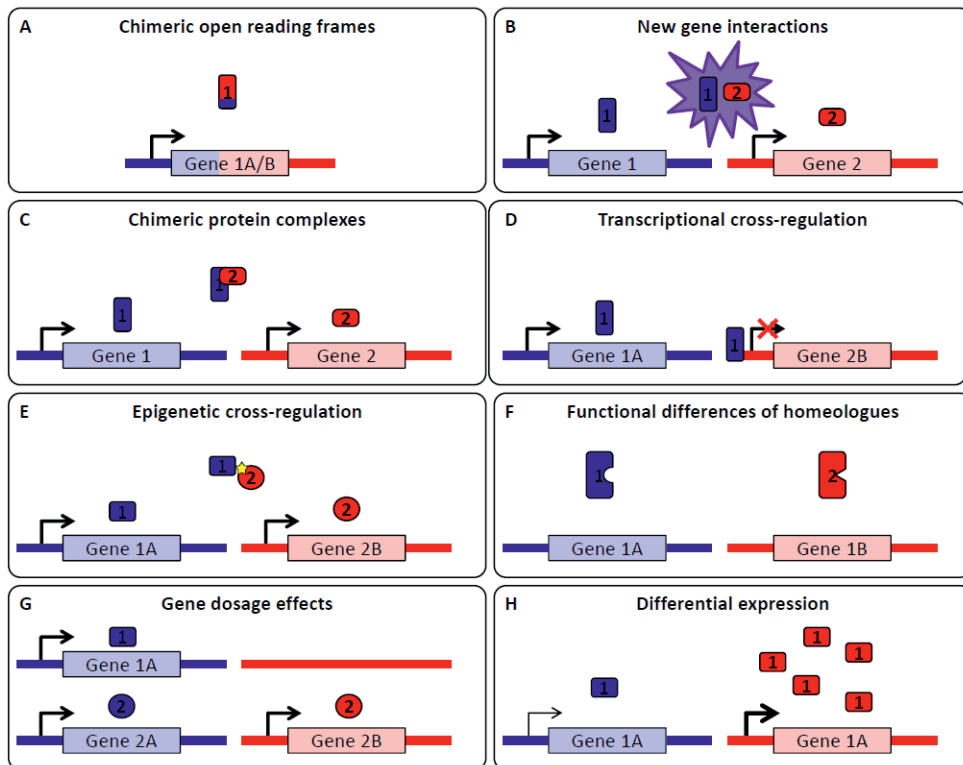
In widely different genetic contexts, domestication has been shown to stimulate rapid genetic adaptation and diversification (19, 41-45). In hybrids such as *S. pastorianus*, genetic plasticity is exacerbated by an increased incidence of (segmental) aneuploidy and loss of heterozygosity (39, 46-48). Therefore, rather than reflecting different origins, the separation of Group 1 and 2 strains may reflect genetic divergence during domestication, followed by severe population bottlenecks (Figure 2C). The industry practice of replacing locally evolved brewing strains by strains from successful breweries, as illustrated by the Bavarian origin of the Carlsberg strain isolated by Hansen (7), is likely to already have reduced genetic diversity among *S. pastorianus* strains. Even narrower bottlenecks may have occurred when Hansen isolated the first pure Group 1 strain at Carlsberg in 1883 and Elion isolated a pure Group 2 strain at Heineken in 1886 (10, 49). These isolates likely spread to other European breweries as they increasingly implemented pure-culture brewing, thereby replacing previously used yeast cultures. Furthermore, in the 19<sup>th</sup> and early 20<sup>th</sup> centuries, small breweries commonly used yeast starter cultures sold by large breweries such as Carlsberg and Heineken, thereby further reducing the diversity of industrial strains (50). In this 'population bottleneck hypothesis', the limited genetic diversity within the two distinct *S. pastorianus* groups may be explained by the independent isolation of a limited set of distinct strains, from which all currently available brewing strains are derived.



**Figure 2: Theories formulated about the emergence of Group 1 and 2 *S. pastorianus* strains.** (A) Emergence by two independent hybridizations, as hypothesized by Dunn *et al* (27). While both groups shared a similar *S. eubayanus* ancestor, Group 1 emerged from hybridization with a haploid *S. cerevisiae*, while Group 2 emerged from a diploid *S. cerevisiae*. (B) Emergence by two successive hybridizations as hypothesized by Okuno *et al* (29). *S. pastorianus* emerged from an initial hybridization between a haploid *S. cerevisiae* and a diploid *S. eubayanus*. Group 1 strains evolved directly from this ancestor, while Group 2 strains emerged from a subsequent hybridization between the *S. pastorianus* ancestor and a haploid *S. cerevisiae* strain of different origin. (C) Emergence by a single hybridization followed by different evolutionary trajectories, as hypothesized by Salazar *et al* (40). *S. pastorianus* emerged from the hybridization between a heterozygous diploid *S. cerevisiae* and a mostly homozygous diploid *S. eubayanus*. Group 1 and 2 strains both evolved from this ancestor. However, Group 1 and Group 2 strains were affected differently by loss of heterozygosity and by loss of *S. cerevisiae* genome content. As a result, Group 2 strains are more heterozygous than Group 1 strains and their *S. cerevisiae* subgenomes differ despite common ancestry.

### Complexity of *S. pastorianus* genomes

*S. pastorianus* genomes are alloaneuploid, with varying, strain-dependent numbers of homologous and homeologous chromosome copies. The chromosome copy number variation referred to as aneuploidy affects the phenotype due to two general mechanisms: (i) the general aneuploidy-associated stress response, encompassing growth defects, genetic instability and low sporulation efficiency, and (ii) chromosome-specific copy-number effects, resulting from the cumulative impact of copy number effects of genes harboured by the affected chromosome. For an overview of the relevance of aneuploidy for industrial application, we refer to our recent review (51). In *S. pastorianus*, the genetic differences between the *S. cerevisiae* and *S. eubayanus* subgenomes present an additional degree of complexity (Figure 3). During genome evolution, recombinations between both subgenomes can create new genetic complexity, for example by creating novel, hybrid



**Figure 3: Mechanisms of subgenome interactions in hybrid organisms that can contribute to synergies between heterozygous genetic material, a phenomenon referred to as heterosis.** Components of the two subgenomes are shown in red and blue. (A) Generation of chimeric proteins due to recombinations within reading frames of (homeologous) genes from different subgenomes. (B) Interactions resulting from the simultaneous expression of subgenome-specific genes which were not expressed together in either parental genome. (C) Formation of chimeric protein complexes due to the assembly of subunits from different subgenomes. (D) Effects on transcription of genes from one subgenome by regulatory proteins from the other subgenome due to non-specificity of regulation. (E) Effects on the activity of proteins from one subgenome by regulatory proteins from the other subgenome due to non-specificity of regulation. (F) Functional differences between the homeologous genes of each subgenome, which can lead to subfunctionalization. (G) Effects due to differences in the relative copy number of different homeologous genes due to differences in gene composition of subgenomes. (H) Differences in transcription of homeologous genes, resulting in different contributions of each subgenome to the resulting phenotype.

open-reading frames (Figure 3A) (38, 52, 53). Since gene complements of the two subgenomes differ (54, 55), interactions of genes and gene products which are not present together in either parental genome can result in novel, difficult to predict phenotypes when simultaneously present in hybrids (Figure 3B). For example, protein subunits encoded by different subgenomes can assemble into novel, chimeric protein complexes (Figure 3C) (56), while non-specificity of regulatory elements can cause regulatory cross-talk of transcription (Figure 3D) and of protein modification (Figure 3E) (57, 58). Moreover, functional differences between homeologous genes (Figure 3F) (59, 60), as well as gene dosage-effects (Figure 3G) (61, 62), can result in complex interactions. Expression levels of homeologous genes generally differ, resulting in stronger expression of one of the two versions (Figure 3H) (63, 64). Overall, understanding the complex interactions between subgenomes is critical, as they underlie the synergistic phenomenon of heterosis (65-67), which enables hybrids such as

*S. pastorianus* to outcompete their parental species (24, 37, 68). The importance of subgenome interactions is consistent with the frequent loss of heterozygosity during evolution of *Saccharomyces* hybrids, since it facilitates elimination of non-beneficial genome content from the least adapted parental species (39, 69-71). The exclusive presence of mitochondrial DNA descending from *S. eubayanus* in *S. pastorianus* strains is a further indication of the importance of loss of heterozygosity (29, 34, 72). Indeed, the loss of *S. cerevisiae* mtDNA was likely instrumental in the lager brewing domestication process, as its replacement by *S. eubayanus* mtDNA enables improved growth at low temperatures (73).

Elucidation of the genetic complexity of *S. pastorianus* strains was initially limited by the accuracy of available genome assemblies (17). The first *S. pastorianus* genome was published in 2009 and consisted of 25 Mbp split over 3184 contigs (30). While many more strains were sequenced since, short-read sequencing invariably yielded incomplete and fragmented genome assemblies with hundreds of contigs at best (28, 29, 31). Indeed, short-read sequencing cannot resolve repetitive sequences, such as TY-transposons and paralogous genes within each subgenome, or homeologous gene pairs (74, 75). As a result, subtelomeric regions, which are known hotspots of genetic plasticity and inter-strain diversity (32, 33, 76-78) and harbour many industrially-relevant genes (79-82), were poorly assembled.

Recent fast developments in long-read sequencing enabled the generation of chromosome-level *S. pastorianus* genome assemblies that include most telomeres (40). *Saccharomyces* genome assemblies based on long-read sequencing typically capture up to 5 % more genes than high-quality short-read assemblies (40, 54, 55, 83-85). Such added genes were of particular interest due to their role in brewing-relevant traits; such as *FLO* genes involved in the calcium-dependent flocculation process that causes bottom fermentation of *S. pastorianus*, *MAL* genes encoding maltose and maltotriose transporters and hydrolases, and *HXT* genes encoding the uptake of glucose and other hexose sugars (40). Despite the near-complete assembly of all chromosomes, the first long-read *S. pastorianus* genome assembly captured only 23 Mbp of the 46 Mbp genome of strain CBS 1483 because assembled chromosomes were consensus sequences of all chromosomal copies, and intra-chromosomal variation of multi-copy chromosomes was not captured. Nevertheless, alignment of short-read and long-read sequences allowed retrieval of sequence and structural heterozygosity (29, 40, 54).

### **Genome-editing techniques in *S. pastorianus***

Compared to the plethora of genome-editing research in *S. cerevisiae* (86-89), there are only very few accounts of targeted genome editing using cassette integration in *S. pastorianus* (59, 90-93), supposedly due to limited homologous recombination efficiency (92). Even the complete deletion of a gene was, until recently, complicated by the presence of several gene copies, which required repeated rounds of cassette insertion and marker removal. Instead, functional characterization often relied on expressing *S. pastorianus* genes in *S. cerevisiae* strains (59, 94-97). While introduction of a double-strand break can drastically increase gene editing efficiency (98), Cas9 gene editing tools developed for *S. cerevisiae* were not directly applicable in *S. pastorianus* strains (87, 92, 99). However, polymerase-II-based expression of gRNAs flanked by self-cleaving ribozymes was successful in *S. pastorianus*, in laboratory-made *S. cerevisiae* x *S. eubayanus* hybrids and in both parental species (39, 55, 92). While application of genetic modification (GM) to generate industrial strains is limited by customer acceptance issues (100), efficient gene-editing also simplifies non-GM strain

improvement. The single-step deletion of a total of 5 copies of the *ATF1* and *ATF2* genes in *S. pastorianus* illustrated the potential of Cas9 to facilitate functional characterisation by enabling fast and complete gene deletion (92). Furthermore, gene editing can be used to evaluate the desirability of mutations prior to the use of laborious non-GM techniques, as illustrated by the deletion of *FDC1* and *PAD1* genes in *S. eubayanus* prior to mutagenesis to obtain non-GM strains with low phenolic off-flavours (101). In addition, when a phenotypic improvement is achieved through non-GM strain improvement methods such as laboratory evolution or mutagenesis, Cas9 can facilitate the elucidation of the underlying genetic mutation that caused the improvement by enabling reverse engineering (39).

### Improvement strategies for lager brewing strains

Industrial strain improvement relies on five pillars: exploration of existing diversity, strain crossing, laboratory evolution, mutagenesis and selection, and genome editing (102, 103). The complex genetics of *S. pastorianus* and, in particular, the lack of customer acceptance of genetic modification have restricted options for strain improvement of brewing yeasts (104). Compared to ale brewing *S. cerevisiae* strains, the genetic and phenotypic diversity of *S. pastorianus* is limited (19, 27, 29, 35, 40, 105). While diversity has been successfully expanded by crossing spores of an *S. pastorianus* strain with *S. cerevisiae* (106, 107), mating strategies are constrained by the low sporulation efficiency of alloaneuploid *S. pastorianus* strains (108-111). As illustrated by the mating of an unsporulated allopolyploid *S. bayanus* strain with beer-brewing *S. cerevisiae* strains (112), low sporulation efficiencies could be circumvented by using rare mating based on spontaneous or induced mating-type switching (113, 114). Although labour- and time-intensive, non-sexual crossing methods such as spheroplast fusion can also be applied (115).

The low mating efficiency of existing *S. pastorianus* strains was circumvented by mating different *Saccharomyces* species in the laboratory to obtain novel *S. pastorianus*-like lager brewing strains (24, 25). In addition to sharing the hybrid vigour of *S. pastorianus*, laboratory-made *S. cerevisiae* x *S. eubayanus* hybrids displayed phenotypic diversity depending on their ploidy and on the genetic background of parental strains (26, 37). Moreover, hybrids of *S. cerevisiae* with other cold-tolerant *Saccharomyces* species such as *S. arboricola*, *S. mikatae* and *S. uvarum* displayed similar fermentation performance at low temperature as *S. pastorianus* (116, 117). Laboratory hybrids are typically made by crossing strains with complementary selectable phenotypes and selecting hybrid cells which combined both phenotypes. In some cases, natural traits of the parental strains, such as growth at low temperature or the ability to utilise melibiose, can be used as selectable phenotypes (112). In the absence of pre-existing selectable phenotypes, selectable mutants can be generated prior to mating. For example, uracil auxotrophy can be selected by growth in the presence of 5-fluoroorotic acid, lysine auxotrophy can be selected by growth in the presence of  $\alpha$ -amino adipate and respiratory-deficient strains can be obtained by growth in the presence of ethidium analogues (118-120). By crossing strains with different auxotrophies or deficiencies, crossed cells can be isolated by selecting fully prototrophic strains (37, 121, 122). Alternatively, selectable phenotypes may be introduced using genome editing, for example by introducing genes conferring antibiotic resistance (123-125). By combining an uncommon auxotrophy and an introduced antibiotic resistance gene in one parental strain, it can be crossed with a large array of other strains without requiring any additional pre-existing or introduced selectable phenotypes (24), however, GM status of such strains precludes application in the food and beverages industry.



**Table 1: Non-GM mutagenesis, selection and/or laboratory evolution methods which resulted in lager-brewing-relevant phenotypic changes in *Saccharomyces* strains.** For each method, the used *Saccharomyces* species, applied mutagenesis methods, applied selection and/or laboratory methods, and the selected phenotype are indicated. For mutagenesis methods, ultraviolet radiation (UV), ethyl methanesulfonate (EMS), methyl benzimidazole-2-yl-carbamate (MBC), N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) are distinguished. RBS denotes the use of a repeated batch set-up.

Strain	Mutagenesis	Selection and/or laboratory evolution	Selected phenotype	Reference
<i>S. pastorianus</i>	UV mutagenesis	RBS cultivations on high-gravity wort	High gravity fermentation	(130)
<i>S. pastorianus</i>	EMS mutagenesis	Fed-batch cultivation on high-gravity wort	High gravity fermentation	(131)
<i>S. pastorianus</i>	UV and EMS mutagenesis	Growth on solid medium with high ethanol concentrations	High gravity fermentation	(132)
<i>S. cerevisiae</i> x <i>S. eubayanus</i>	-	RBS cultivation with high ethanol concentrations	High gravity fermentation	(121)
<i>S. cerevisiae</i>	-	Batch cultivation in high gravity medium in the presence of ethanol	High gravity fermentation	(133)
<i>S. cerevisiae</i>	-	Turbidostat cultivation with increasing ethanol concentrations	Increased ethanol tolerance	(134)
<i>S. cerevisiae</i>	-	RBS cultivations with increasing ethanol concentrations	Increased ethanol tolerance	(135)
<i>S. cerevisiae</i>	EMS mutagenesis	Turbidostat cultivation with increasing ethanol concentrations	Increased ethanol tolerance	(136)
<i>S. uvarum</i>	EMS mutagenesis	Turbidostat cultivation with increasing ethanol concentrations	Increased ethanol tolerance	(137)
<i>S. cerevisiae</i>	-	Batch cultivations with intermittent exposure to 0.3-4.4M of H <sub>2</sub> O <sub>2</sub> , 52 °C, 20-55 % ethanol and freeze/thawing cycles	Increased tolerance to oxidative-, temperature-, ethanol- and freezing-thawing stress	(138)
<i>S. pastorianus</i>	EMS mutagenesis	Repeated heat shocks at 55 °C	Increased heat shock tolerance	(139)
<i>S. cerevisiae</i>	UV mutagenesis	Subjection to 200 freeze-thaw cycles	Increased freeze tolerance	(140)
<i>S. cerevisiae</i>	MNNG mutagenesis	Differential staining with triphenyltetrazolium chloride	Crabtree-negative mutants	(141)
<i>S. cerevisiae</i>	-	Growth on solid medium with 2-deoxyglucose	Loss of glucose repression	(142)
<i>S. cerevisiae</i>	-	Growth on solid medium with glucosamine	Loss of glucose repression	(143)
<i>S. eubayanus</i>	-	RBS cultivations on synthetic medium with maltose and traces of glucose	Maltose utilization	(144)
<i>S. pastorianus</i>	-	Chemostat cultivation on maltotriose enriched mock-wort	Maltotriose utilization	(17)
<i>S. eubayanus</i>	-	RBS cultivations on synthetic medium with maltotriose and traces of glucose	Maltotriose utilization	(144)
<i>S. eubayanus</i>	UV mutagenesis	RBS cultivations on synthetic medium with maltotriose and chemostat cultivation on maltotriose-enriched wort	Maltotriose utilization	(53)
<i>S. cerevisiae</i> x <i>S. uvarum</i>	-	Chemostat cultivation under ammonium limitation	Increased fitness under nitrogen limitation	(52)
<i>S. cerevisiae</i>	-	Chemostat cultivation under nitrogen limitation	Increased fitness under nitrogen limitation	(145)
<i>S. cerevisiae</i> x <i>S. uvarum</i>	-	Chemostat cultivation under carbon-, phosphate- and sulphate limitation	Increased fitness under nutrient limitation	(69)
<i>S. cerevisiae</i>	-	Chemostat cultivation under carbon-, phosphate- and sulphate limitation	Increased fitness under nutrient limitation	(146)
<i>S. cerevisiae</i> x <i>S. eubayanus</i>	-	RBS cultivations on wort	Increased flocculation	(39)
<i>S. cerevisiae</i>	-	Chemostat cultivation	Increased flocculation	(147)
<i>S. cerevisiae</i>	MNNG mutagenesis	RBS cultivation enriching for slow-sedimenting cells	Loss of flocculation	(148)
<i>S. pastorianus</i>	-	Batch cultivation in the presence of Ethylum Bromide	Loss of respiratory capacity	(148)
<i>S. cerevisiae</i> x <i>S. uvarum</i>	-	Chemostat cultivation at 15 °C	Increased growth at low temperatures	(71)
<i>S. cerevisiae</i>	-	RBS cultivation with sulphate	Increased glycerol production	(149)
<i>S. cerevisiae</i>	-	RBS cultivation with S-methyl-L-cysteine	Increased thiol production	(150)
<i>S. cerevisiae</i>	UV mutagenesis	Screening for lack of coloration on lead plates	Decreased H <sub>2</sub> S production, increased SO <sub>2</sub> production	(151)
<i>S. pastorianus</i>	-	Growth on solid medium with ethionine, screening for coloration on lead plates	Increased SO <sub>2</sub> production	(152)
<i>S. cerevisiae</i>	UV mutagenesis	Growth on solid medium with cadmium	Increased glutathione production	(151)
<i>S. pastorianus</i>	UV mutagenesis	Growth on solid medium with disulfiram	Decreased acetaldehyde production	(153)
<i>S. cerevisiae</i>	UV mutagenesis	Growth on solid medium with cerulenin	Increased fatty-acid synthesis	(154)

**Table 1 (continued): Non-GM mutagenesis, selection and/or laboratory evolution methods which resulted in lager-brewing-relevant phenotypic changes in *Saccharomyces* strains.** For each method, the used *Saccharomyces* species, applied mutagenesis methods, applied selection and/or laboratory methods, and the selected phenotype are indicated. For mutagenesis methods, ultraviolet radiation (UV), ethyl methanesulfonate (EMS), methyl benzenesulfonate (MBC), N-methyl-N'-nitro-N-nitroso-guanidine (MNNG) are distinguished. RBS denotes the use of a repeated batch set-up.

Strain	Mutagenesis	Selection and/or laboratory evolution	Selected phenotype	Reference
<i>S. pastorianus</i>	-	Growth on solid medium with 5,5,5-trifluoro-DL-leucine	Increased isoamyl alcohol and isoamyl acetate production	(155)
<i>S. cerevisiae</i>	EMS mutagenesis	Growth on solid medium with isoamyl monochloroacetate	Increased isoamyl acetate production	(156)
<i>S. cerevisiae</i>	-	RBS cultivation in the presence of 1-farnesylpyridinium	Increased isoamyl acetate production	(157)
<i>S. cerevisiae</i>	EMS mutagenesis	Growth on solid medium with econazole	Increased isoamyl acetate production	(158)
<i>S. cerevisiae</i>	-	Batch cultivation in the presence of 8 mM Cu <sup>2+</sup>	Increased isoamyl acetate production	(159)
<i>S. cerevisiae</i>	EMS mutagenesis	Growth on solid medium with isoamyl monofluoroacetate	Increased isoamyl acetate production	(160)
<i>S. uvarum</i>	-	Growth on solid medium with 5,5,5-trifluoro-DL-leucine and fluoro-DL-phenylalanine	Increased isoamyl acetate and phenylethyl acetate production	(161)
<i>S. cerevisiae</i>	-	Growth on solid medium with <i>p</i> -Fluoro-DL-phenylalanine	Increased $\beta$ -phenethyl alcohol and $\beta$ -phenethyl acetate production	(162, 163)
<i>S. pastorianus</i>	MNNG mutagenesis	Growth on solid medium with thialsoleucine	Increased 2-methyl-1-butanol production	(164)
<i>S. pastorianus</i>	EMS mutagenesis	RBS cultivation in the presence of chlorosulfuron	Decreased diacetyl production	(165)
<i>S. eubayanus</i>	UV mutagenesis	Screening for insensitivity to cinnamic acid	Decreased 4-vinyl guaiacol production	(101)

The requirement for genetic markers can be completely circumvented by staining parental strains with fluorescent dyes prior to mating and, subsequently, sorting double-stained cells using fluorescence-assisted cell sorting. Indeed, hybrids could be obtained with this new method without the use of any selectable phenotype (126). Regardless of the method used to generate new *Saccharomyces* hybrids laboratory, subsequent strain analysis demonstrated that they were sufficiently stable for successive brewing batches (39), while laboratory evolution under high-ethanol conditions demonstrated their increased adaptive ability during strain improvement (121). For an analysis of the potential of laboratory-made hybrids for brewing, we refer to a recent review (127).

Both *S. pastorianus* strains and laboratory-made lager-brewing hybrids can be further improved by laboratory evolution and/or mutagenesis and selection (Table 1). Generation of novel phenotypes can occur by spontaneous acquisition of mutations during growth. Alternatively, the mutation frequency can be increased by mutagenesis using irradiation (such as ultraviolet light) or by exposure to mutagenic compounds (such as ethyl methanesulfonate (EMS), methyl benzimidazole-2-yl-carbamate (MBC), N-methyl-N'-nitro-N-nitroso-guanidine (MNNG)). Mutants of interest can be isolated by screening for desirable phenotypes, or by growth under conditions that confer a selective benefit to mutants with a desirable phenotype.

When growth under conditions favouring desired phenotypes is applied not only to select mutants, but also to generate new mutants in the process, it is designated as laboratory evolution. This strategy has been successfully applied to select for lager brewing-relevant phenotypes of *Saccharomyces* strains, including superior fermentation in 'high gravity' processes, increased ethanol tolerance, improved sugar utilisation, increased performance under nutrient limitation, altered flocculation behaviour and altered flavour profiles (Table 1). For an overview of relevant taste compounds in beer brewing and of relevant phenotypic properties of brewing yeast, we refer to recent reviews (128, 129). This brief overview of methods should not be considered as complete. In particular, commercial breweries may not share proprietary methods for non-GM yeast strain improvement.

### **Potential of genetic engineering for improved lager brewing performance**

The lager brewing industry does not currently use GM yeast for lager beer brewing. Many countries and trade blocks, including important beer markets such as the EU and the USA, tightly regulate use of GM technology in the food and beverages industry (166). Historically, regulation was technology based: methods to modify genomes by non-targeted methods such as UV mutagenesis and chemical mutagenesis were not regulated, while any mutation introduced by targeted genetic engineering was subject to specific legislation (167). Recently, regulation appears to be moving towards product- and risk-based evaluation, in which the type of mutation introduced determines regulatory status (166, 168). For example, Japan regulates genetic engineering less strictly when no foreign DNA is introduced ('self-cloning'). Similarly, in the USA, GM foods which only harbour simple single nucleotide polymorphisms (SNPs) that might also have arisen after non-targeted mutagenesis, have been introduced into the market (169-171). However, similar developments towards product- and risk-based regulation were recently blocked by legislative courts in the European Union. Consequently, updating the aging GM regulations in the EU will now require a considerable political process (172).

**Table 2: Genetic engineering strategies which were successfully applied in *Saccharomyces* yeasts with potential application for the lager brewing industry.**

Application	Modification	Phenotype	Organism	Reference
Substrate utilization	AGT1 overexpression	Increased maltose and maltotriose utilisation	<i>S. pastorianus</i>	(93)
	Heterologous gene expression	Increased $\beta$ -glucan degradation	<i>S. pastorianus</i>	(178)
Industrial performance	Heterologous gene expression	Increased dextrin utilisation	<i>S. pastorianus</i>	(179-181)
	PUT4 overexpression	Increased proline assimilation	<i>S. pastorianus</i>	(182)
	GPD1 overexpression	Increased glycerol production, decreased ethanol production	<i>S. pastorianus</i>	(183)
	FLO1, FLO5, or FLO11 overexpression	Increased flocculation	<i>S. cerevisiae</i>	(184)
	Stationary-phase FLO1 overexpression	Stationary-phase flocculation	<i>S. pastorianus</i>	(185)
	PEP4 disruption	Improved foam stability	<i>S. cerevisiae</i>	(186)
	LEU1 overexpression	Improved high gravity fermentation	<i>S. pastorianus</i>	(130)
	FKS1 disruption	Reduced yeast autolysis	<i>S. pastorianus</i>	(187)
	MET10 disruption	Increased SO <sub>2</sub> production	<i>S. pastorianus</i>	(188)
	MET14 and SSU1 overexpression	Increased SO <sub>2</sub> production	<i>S. cerevisiae</i>	(189)
Off-flavour reduction	HOM3 overexpression, SKP2 disruption	Increased SO <sub>2</sub> and decreased H <sub>2</sub> S production	<i>S. pastorianus</i>	(152)
	CYS4 overexpression	Decreased H <sub>2</sub> S production	<i>S. cerevisiae</i>	(190)
	NH55 overexpression	Decreased H <sub>2</sub> S production	<i>S. pastorianus</i>	(190)
	MXR1 disruption	Decreased dimethylsulfide production	<i>S. cerevisiae</i>	(191)
	ILV5 overexpression	Decreased diacetyl production	<i>S. cerevisiae</i>	(192)
	ILV6 disruption	Decreased diacetyl production	<i>S. pastorianus</i>	(90)
	Heterologous gene expression	Increased diacetyl degradation	<i>S. pastorianus</i>	(193-197)
	FDC1 disruption	Decreased 4-vinyl guaiacol production	<i>S. pastorianus</i>	(177)
	LEU4 overexpression	Increased isoamyl acetate production	<i>S. cerevisiae</i>	(198)
	ATF1 and ATF2 overexpression	Increased acetate ester production	<i>S. pastorianus</i>	(199)
Flavour modulation	ATF1 and ATF2 disruption	Decreased acetate ester production	<i>S. pastorianus</i>	(199)
	ALD3 disruption, ARO9 and ARO10 overexpression	Increased 2-phenylethanol production	<i>S. cerevisiae</i>	(200)
Introduction of new flavours	Heterologous gene expression	Increased ethyl hexanoate production	<i>S. cerevisiae</i>	(201)
	Heterologous gene expression	Hop monoterpene production	<i>S. cerevisiae</i>	(202)
	Heterologous gene expression	Hop lupulone production	<i>S. cerevisiae</i>	(203)
	Heterologous gene expression	$\beta$ -ionone production	<i>S. cerevisiae</i>	(204)
	Heterologous gene expression	Vanillin production	<i>S. cerevisiae</i>	(205, 206)
	Heterologous gene expression	Valencene production	<i>S. cerevisiae</i>	(207)
	Heterologous gene expression	Nootkatone production	<i>S. cerevisiae</i>	(208)
	Heterologous gene expression	Raspberry ketone production	<i>S. cerevisiae</i>	(209, 210)

Since, in the EU, food products only need to be labelled and regulated as GM if they contain more than 0.9 % GM biomass, removal of GM yeast by filtration could, in principle, obviate the need for labelling the resulting beer as a GM product (173). Moreover, already in 1990, a lager brewing strain engineered for dextrin utilization was approved and used to brew a low-caloric beer in the UK (100, 174). As illustrated by the commercial failure of this GMO beer, the application of GM yeasts for beer brewing is precluded primarily due to customer acceptance rather than insurmountable regulatory hurdles (175). Recent regulatory developments have resulted in increased commercialisation foods based on targeted genetic modification, particularly on the US market (171, 176). For example, Lallemand (Montreal, Canada) is currently concluding trials with a brewing yeast engineered to produce lactic acid, called *Sourvisiae*.

Despite the current absence of large-scale industrial application, many possible genetic engineering strategies for lager brewing yeasts are available, based to insights gained from laboratory studies and from analysis of strains obtained by classical strain improvement. Such strategies could rapidly and efficiently improve a vast array of yeast characteristics, including substrate utilization, general brewing performance, energy requirements for cooling, off-flavour and flavour profiles and, moreover, enable the introduction of novel flavours (Table 2). The relatively permissive legislation and relatively high consumer acceptance in countries such as Brazil, USA, Japan and Argentina may enable industrial application of GM yeast for lager beer brewing in the near future (177).

### Scope of this thesis

The general scope of this thesis was to study the complex genomes of *S. pastorianus* strains and to facilitate strain improvement strategies for generating lager brewing yeasts with novel or superior brewing properties for industrial application. In response to the limited genetic accessibility of *S. pastorianus* strains, recent advances in genome mapping and gene editing were applied to generate new tools applicable to alloaneuploid *S. pastorianus* strains, to its parental species *S. cerevisiae* and *S. eubayanus*, and to laboratory-made *S. cerevisiae* x *S. eubayanus* strains. These tools were applied to study the evolution of *S. pastorianus* hybrids, to study heterosis due to subgenome interactions and to generate novel lager brewing strains.

The degree of aneuploidy of *S. pastorianus* genomes is unique within the *Saccharomyces* genus. Moreover, chromosome copy number differs vastly between individual isolates. Therefore, in **Chapter 2**, the industrial relevance and the impact of aneuploidy and of chromosome copy number variation were reviewed.

To understand the ancestry and genomic complexity of *S. pastorianus* strains, it is critical to understand the genomes of their parental species. Therefore, in **Chapter 3**, recent advances in single-molecule nanopore sequencing technologies were applied to generate a more complete and less fragmented genome assembly of the *S. cerevisiae* reference strain CEN.PK113-7D.

In **Chapter 4**, nanopore sequencing technology was applied directly to *S. pastorianus* strain CBS 1483 and the possibility to resolve sequence and structural heterogeneity between different chromosome copies was evaluated. In combination with the newly-developed Alpaca algorithm, which computes sequencing read similarities between different data sets, the resulting chromosome-level genome assembly was used to investigate the ancestry of Group 1 and Group 2 *S. pastorianus* strains.

The limited efficiency of homologous recombination and alloaneuploidy of *S. pastorianus* strains have limited their genetic accessibility. The application of the bacterial adaptive immune system CRISPR-Cas9 to introduce targeted double-strand breaks in DNA has revolutionized gene editing in *S. cerevisiae*. As the gene editing methods developed for *S. cerevisiae* did not work in *S. pastorianus*, a broad-host-range, plasmid-based Cas9 gene editing tool was developed in **Chapter 5** and was shown to enable efficient editing in *S. pastorianus*.

**Chapter 6** describes the observation that, when applied to target heterozygous sequences, the introduction of a double-strand break by Cas9 does not result in efficient gene editing but, instead, causes frequent loss of heterozygosity. This loss of heterozygosity was shown to affect up to several hundred thousand base pairs. Since loss of heterozygosity was linked to the highly conserved mechanism of homologous recombination, potential implications for gene editing in higher eukaryotes, notably humans, were discussed.

The generation of novel, laboratory-made *Saccharomyces* hybrids has been used to obtain strains with new or superior properties for a broad range of applications, including lager beer brewing. However, such interspecific mating currently relies on previously-available or introduced selectable phenotypes in the parental strains. In **Chapter 7**, a method to generate hybrids using fluorescent labelling of parental strains and sorting of double-stained hybrids using fluorescence-activated cell sorting was developed, which does not rely on any selectable phenotype and can be applied to mate a broad range of parental strains.

The allopolyploid genomes of laboratory-made hybrids have been shown to be relatively unstable during evolution under industrial conditions such as wine-making and 2<sup>nd</sup> generation bioethanol production. To evaluate their applicability in lager brewing, **Chapter 8** describes the genetic and phenotypic changes of laboratory-made *S. cerevisiae* x *S. eubayanus* hybrids after subjecting them to prolonged laboratory evolution under conditions mimicking industrial lager brewing.

In addition to the generation of laboratory-made *S. cerevisiae* x *S. eubayanus* hybrids, *S. eubayanus* strains have also been applied directly for lager brewing. However, the inability of many *S. eubayanus* strains to utilise maltotriose limits their applicability, mainly due to lower ethanol yields and to residual sweetness. To address this issue and improve the industrial applicability of *S. eubayanus*, **Chapter 9** describes a strain improvement approach to obtain a maltotriose-utilizing strain and a molecular analysis of the mechanism by which it acquired this trait.

## References

1. Liu L, *et al.* (2018) Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. *J Archaeol Sci* 21:783-793.
2. Michel RH, McGovern PE, & Badler VR (1992) Chemical evidence for ancient beer. *Nature* 360(6399):24.
3. Sicard D & Legras J-L (2011) Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C R Biol* 334(3):229-236.
4. Hornsey IS (2003) *A history of beer and brewing* (Royal Society of Chemistry, Cambridge).
5. Unger RW (2004) *Beer in the Middle Ages and the Renaissance* (University of Pennsylvania Press, Philadelphia).
6. Oliver G & Colicchio T (2011) *The Oxford companion to beer* (Oxford University Press, New York).
7. Meusdoerffer FG (2009) *A comprehensive history of beer brewing* (Wiley-VCH, Eßlinger) pp 1-42.
8. Pasteur L (1876) *Études sur la bière: ses maladies, causes qui les provoquent, procédé pour la rendre inaltérable* (Gauthier-Villars, Paris).
9. Moritz ER & Morris GH (1891) *A Text-book of the Science of Brewing* (Spon, London).
10. Hansen EC (1883) Recherches sur la physiologie et la morphologie des ferments alcooliques. V. Methodes pour obtenir des cultures pures de *Saccharomyces* et de microorganismes analogues. *Compt Rend Trav Lab Carlsberg* 2:92-105.

11. Poelmans E & Swinnen JF (2011) From monasteries to multinationals (and back): A historical review of the beer economy. *J Wine Econ* 6(2):196-216.
12. Appel SK (1990) Artificial Refrigeration and the Architecture of 19th-century American Breweries. *IA. J Soc Indust Archaeol*:21-38.
13. Lockhart B (2007) The origins and life of the export beer bottle. *Origins* 49.
14. Kunze W (2004) *Brewing & Malting* (VLB, Berlin) pp 18-152.
15. Painter W (1892) OFFICE. USP 468226.
16. Michel C (1899) *Geschichte des Bieres von der ältesten Zeit bis zum Jahre 1899* (Verlagsbuchhandlung Gebrüder Reichel, Augsburg).
17. Brickwedde A, et al. (2017) Evolutionary engineering in chemostat cultures for improved maltotriose fermentation kinetics in *Saccharomyces pastorianus* lager brewing yeast. *Front Microbiol* 8:1690.
18. Libkind D, et al. (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA*:201105430.
19. Gallone B, et al. (2016) Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166(6):1397-1410. e1316.
20. Peris D, et al. (2014) Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Mol Ecol* 23(8):2031-2045.
21. Bing J, Han P-J, Liu W-Q, Wang Q-M, & Bai F-Y (2014) Evidence for a Far East Asian origin of lager beer yeast. *Curr Biol* 24(10):R380-R381.
22. Gayevskiy V & Goddard MR (2016) *Saccharomyces eubayanus* and *Saccharomyces arboricola* reside in North Island native New Zealand forests. *Environ Microbiol Rep* 18(4):1137-1147.
23. Alsammar HF, et al. (2019) Targeted metagenomics approach to capture the biodiversity of *Saccharomyces* genus in wild environments. *Environ Microbiol Rep*.
24. Heblly M, et al. (2015) *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res* 15(3).
25. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2015) New lager yeast strains generated by interspecific hybridization. *J Ind Microbiol Biotechnol* 42(5):769-778.
26. Mertens S, et al. (2015) A large set of newly created interspecific yeast hybrids increases aromatic diversity in lager beers. *Appl Environ Microbiol*:AEM. 02464-02415.
27. Dunn B & Sherlock G (2008) Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res* 18(10):1610-1623.
28. Van den Broek M, et al. (2015) Chromosomal copy number variation in *Saccharomyces pastorianus* evidence for extensive genome dynamics in industrial lager brewing strains. *Appl Environ Microbiol*:AEM. 01263-01215.
29. Okuno M, et al. (2016) Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res* 23(1):67-80.
30. Nakao Y, et al. (2009) Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Res* 16(2):115-129.
31. Walther A, Hesselbart A, & Wendland J (2014) Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. *G3 (Bethesda)*:g3. 113.010090.
32. Liti G, Peruffo A, James SA, Roberts IN, & Louis EJ (2005) Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. *Yeast* 22(3):177-192.
33. Monerawela C, James TC, Wolfe KH, & Bond U (2015) Loss of lager specific genes and subtelomeric regions define two different *Saccharomyces cerevisiae* lineages for *Saccharomyces pastorianus* Group I and II strains. *FEMS Yeast Res* 15(2):fou008.
34. Baker E, et al. (2015) The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol Biol Evol* 32(11):2818-2831.
35. Gibson BR, Storgårds E, Krogerus K, & Vidgren V (2013) Comparative physiology and fermentation performance of Saaz and Froberg lager yeast strains and the parental species *Saccharomyces eubayanus*. *Yeast* 30(7):255-266.
36. Rainieri S, et al. (2006) Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome. *Appl Environ Microbiol* 72(6):3968-3974.
37. Krogerus K, et al. (2016) Ploidy influences the functional attributes of *de novo* lager yeast hybrids. *Appl Microbiol Biotechnol* 100(16):7203-7222.
38. Hewitt SK, Donaldson IJ, Lovell SC, & Delneri D (2014) Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. *PLoS One* 9(3):e92203.
39. Gorter de Vries AR, et al. (2019) Laboratory evolution of a *Saccharomyces cerevisiae* × *S. eubayanus* hybrid under simulated lager-brewing conditions, *Frontiers in Genetics* 10:242.
40. Salazar AN, et al. (2019) Nanopore sequencing and comparative genome analysis confirm lager-brewing yeasts originated from a single hybridization. *bioRxiv*:603480.
41. Arnold ML (2004) Natural hybridization and the evolution of domesticated, pest and disease organisms. *Mol Ecol* 13(5):997-1007.

42. Bachmann H, Starrenburg MJ, Molenaar D, Kleerebezem M, & van Hylckama Vlieg JE (2012) Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. *Genome Res* 22(1):115-124.
43. Gibbons JG & Rinker DC (2015) The genomics of microbial domestication in the fermented food environment. *Curr Opin Genet Dev* 35:1-8.
44. Gibbons JG, et al. (2012) The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. *Curr Biol* 22(15):1403-1409.
45. Peter J, et al. (2018) Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556(7701):339.
46. Delneri D, et al. (2003) Engineering evolution to study speciation in yeasts. *Nature* 422(6927):68.
47. Peris D, et al. (2017) Hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production. *Biotechnol Biofuels* 10(1):78.
48. Pérez Través L, Lopes CA, Barrio E, & Querol A (2014) Study of the stabilization process in *Saccharomyces* intra- and interspecific hybrids in fermentation conditions. *Int Microbiol* 17(4):213-224.
49. Struyk AP (1928) Onderzoekingen over de alcoholische gisting. (University of Technology Delft, Delft).
50. Mendlik F (1937) Some aspects of the scientific development of brewing in Holland. *J I Brewing* 43(4):294-300.
51. Gorter de Vries AR, Pronk JT, & Daran J-MG (2017) Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts. *Appl Environ Microbiol*:AEM. 03206-03216.
52. Dunn B, et al. (2013) Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet* 9(3):e1003366.
53. Brouwers N, et al. (2019) In vivo recombination of *Saccharomyces eubayanus* maltose-transporter genes yields a chimeric transporter that enables maltotriose fermentation. *PLoS genetics* 15(4):e1007853.
54. Salazar AN, et al. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res* 17(7).
55. Brickwedde A, et al. (2018) Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357T. *Front Microbiol* 9:1786.
56. Piatkowska EM, Naseeb S, Knight D, & Delneri D (2013) Chimeric protein complexes in hybrid species generate novel phenotypes. *PLoS Genet* 9(10):e1003836.
57. Vidgren V & Gibson B (2018) Trans-Regulation and localization of orthologous maltose transporters in the interspecies lager yeast hybrid. *FEMS Yeast Res*.
58. Tirosh I, Reikhav S, Levy AA, & Barkai N (2009) A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324(5927):659-662.
59. Bolat I, Romagnoli G, Zhu F, Pronk JT, & Daran J-MG (2013) Functional analysis and transcriptional regulation of two orthologs of *ARO10*, encoding broad-substrate-specificity 2-oxo-acid decarboxylases, in the brewing yeast *Saccharomyces pastorianus* CBS1483. *FEMS Yeast Res* 13(6):505-517.
60. Yamagishi H, Ohnuki S, Nogami S, Ogata T, & Ohya Y (2010) Role of bottom-fermenting brewer's yeast *KEX2* in high temperature resistance and poor proliferation at low temperatures. *J Gen Appl Microbiol* 56(4):297-312.
61. Yao H, Gray AD, Auger DL, & Birchler JA (2013) Genomic dosage effects on heterosis in triploid maize. *Proc Natl Acad Sci USA* 110(7):2665-2669.
62. Ogata T, Kobayashi M, & Gibson BR (2013) Pilot-scale brewing using self-cloning bottom-fermenting yeast with high *SSU1* expression. *J I Brewing* 119(1-2):17-22.
63. Gibson B, Londesborough J, Rautio J, Mattinen L, & Vidgren V (2013) Transcription of  $\alpha$ -glucoside transport and metabolism genes in the hybrid brewing yeast *Saccharomyces pastorianus* with respect to gene provenance and fermentation temperature. *J I Brewing* 119(1-2):23-31.
64. He Y, et al. (2014) Monitoring of the production of flavour compounds by analysis of the gene transcription involved in higher alcohol and ester formation by the brewer's yeast *Saccharomyces pastorianus* using a multiplex RT-qPCR assay. *J I Brewing* 120(2):119-126.
65. Shapira R, Levy T, Shaked S, Fridman E, & David L (2014) Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models. *Heredity* 113(4):316.
66. Lippman ZB & Zamir D (2007) Heterosis: revisiting the magic. *Trends Genet* 23(2):60-66.
67. Chen ZJ (2013) Genomic and epigenetic insights into the molecular bases of heterosis. *Nat Rev Genet* 14(7):471.
68. Belloch C, Orlic S, Barrio E, & Querol A (2008) Fermentative stress adaptation of hybrids within the *Saccharomyces sensu stricto* complex. *Int J Food Microbiol* 122(1):188-195.
69. Smukowski Heil CS, et al. (2017) Loss of heterozygosity drives adaptation in hybrid yeast. *Mol Biol Evol* 34(7):1596-1612.
70. Lancaster SM, Payen C, Heil CS, & Dunham MJ (2018) Fitness benefits of loss of heterozygosity in *Saccharomyces* hybrids. *bioRxiv*:452748.
71. Heil CS, Large CR, Patterson K, & Dunham MJ (2018) Temperature preference biases parental genome retention during hybrid evolution. *bioRxiv*:429803.
72. Rainieri S, Kodama Y, Nakao Y, Pulvirenti A, & Giudici P (2008) The inheritance of mtDNA in lager brewing strains. *FEMS Yeast Res* 8(4):586-596.
73. Baker EP, et al. (2019) Mitochondrial DNA and temperature tolerance in lager yeasts. *Sci Adv* 5(1):eaav1869.



74. Kim JM, Vanguri S, Boeke JD, Gabriel A, & Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 8(5):464-478.
75. Matheson K, Parsons L, & Gammie A (2017) Whole-genome sequence and variant analysis of W303, a widely-used strain of *Saccharomyces cerevisiae*. *G3 (Bethesda)*:g3. 117.040022.
76. Pryde FE, Huckle TC, & Louis EJ (1995) Sequence analysis of the right end of chromosome XV in *Saccharomyces cerevisiae*: an insight into the structural and functional significance of sub-telomeric repeat sequences. *Yeast* 11(4):371-382.
77. Bergström A, et al. (2014) A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* 31(4):872-888.
78. Brown CA, Murray AW, & Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol* 20(10):895-903.
79. Jordan P, Choe J-Y, Boles E, & Oreb M (2016) Hxt13, Hxt15, Hxt16 and Hxt17 from *Saccharomyces cerevisiae* represent a novel type of polyol transporters. *Sci Rep* 6:23502.
80. Teste M-A, François JM, & Parrou J-L (2010) Characterization of a new multigene family encoding isomaltases in the yeast *Saccharomyces cerevisiae*: the IMA family. *J Biol Chem*:jbc. M110. 145946.
81. Denayrolles M, de Villechenon EP, Lonvaud-Funel A, & Aigle M (1997) Incidence of *SUC-RTM* telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Curr Genet* 31(6):457-461.
82. Teunissen A & Steensma H (1995) The dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* 11(11):1001-1013.
83. Goodwin S, et al. (2015) Oxford Nanopore sequencing, hybrid error correction, and *de novo* assembly of a eukaryotic genome. *Genome Res* 25(11):1750-1756.
84. Giordano F, et al. (2017) *De novo* yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci Rep* 7(1):3935.
85. Istace B, et al. (2017) *de novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6(2):1-13.
86. Nielsen J & Keasling JD (2016) Engineering cellular metabolism. *Cell* 164(6):1185-1197.
87. DiCarlo JE, et al. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 41(7):4336-4343.
88. Jakočiūnas T, Jensen MK, & Keasling JD (2016) CRISPR/Cas9 advances engineering of microbial cell factories. *Metab Eng* 34:44-59.
89. Nielsen J, Larsson C, van Maris AJA, & Pronk J (2013) Metabolic engineering of yeast for production of fuels and chemicals. *Curr Opin Biotechnol* 24(3):398-404.
90. Duong C, et al. (2011) Identification of *Sc*-type *ILV6* as a target to reduce diacetyl formation in lager brewers' yeast. *Metab Eng* 13(6):638-647.
91. Murakami N, et al. (2012) Construction of a *URA3* deletion strain from the allotetraploid bottom-fermenting yeast *Saccharomyces pastorianus*. *Yeast* 29(5):155-165.
92. Gorter de Vries AR, de Groot PA, Broek M, & Daran J-MG (2017) CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb Cell Fact* 16(1):222.
93. Vidgren V, Huuskonen A, Virtanen H, Ruohonen L, & Londesborough J (2009) Improved fermentation performance of a lager yeast after repair of its *AGT1* maltose and maltotriose transporter genes. *Appl Environ Microbiol* 75(8):2333-2345.
94. Yoshimoto H, et al. (1998) Characterization of the *ATF1* and *Lg-ATF1* genes encoding alcohol acetyltransferases in the bottom fermenting yeast *Saccharomyces pastorianus*. *J Biosci Bioeng* 86(1):15-20.
95. Kodama Y, Omura F, & Ashikari T (2001) Isolation and characterization of a gene specific to lager brewing yeast that encodes a branched-chain amino acid permease. *Appl Environ Microbiol* 67(8):3455-3462.
96. Kobayashi O, Hayashi N, Kuroki R, & Sone H (1998) Region of Flo1 proteins responsible for sugar recognition. *J Bacteriol* 180(24):6503-6510.
97. Salema-Oom M, Pinto VV, Gonçalves P, & Spencer-Martins I (2005) Maltotriose utilization by industrial *Saccharomyces* strains: characterization of a new member of the  $\alpha$ -glucoside transporter family. *Appl Environ Microbiol* 71(9):5044-5049.
98. Pâques F & Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63(2):349-404.
99. Mans R, et al. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 15(2):fov004.
100. Akada R (2002) Genetically modified industrial yeast ready for application. *J Biosci Bioeng* 94(6):536-544.
101. Diderich JA, Weening SM, van den Broek M, Pronk JT, & Daran J-MG (2018) Selection of *Pof Saccharomyces eubayanus* variants for the construction of *S. cerevisiae*  $\times$  *S. eubayanus* hybrids with reduced 4-vinyl guaiacol formation. *Front Microbiol* 9:1640.
102. Steensels J, et al. (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol Rev* 38(5):947-995.
103. Patnaik R (2008) Engineering complex phenotypes in industrial strains. *Biotechnol Prog* 24(1):38-47.

104. Gibson B, *et al.* (2017) New yeasts—new brews: modern approaches to brewing yeast design and development. *FEMS Yeast Res* 17(4).
105. Steensels J, Meersman E, Snoek T, Saels V, & Verstrepen KJ (2014) Large-scale selection and breeding to generate industrial yeasts with superior aroma production. *Appl Environ Microbiol*:AEM. 02235-02214.
106. Sanchez RG, Solodovnikova N, & Wendland J (2012) Breeding of lager yeast with *Saccharomyces cerevisiae* improves stress resistance and fermentation performance. *Yeast* 29(8):343-355.
107. Bilinski CA & Casey GP (1989) Developments in sporulation and breeding of brewer's yeast. *Yeast* 5(6):429-438.
108. Liti G, Barton DB, & Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174(2):839-850.
109. Ogata T, Shikata-Miyoshi M, Tadami H, & Nakazawa N (2011) Isolation of meiotic segregants from a bottom fermenting yeast. *J I Brewing* 117(2):199-205.
110. Santaguida S & Amon A (2015) Short-and long-term effects of chromosome mis-segregation and aneuploidy. *Nat Rev Mol Cell Biol* 16(8):473.
111. Gjermansen C & Sigsgaard P (1981) Construction of a hybrid brewing strain of *Saccharomyces carlsbergensis* by mating of meiotic segregants. *Carlsberg Res Commun* 46(1-2):1.
112. Sato M, Kishimoto M, Watari J, & Takashio M (2002) Breeding of brewer's yeast by hybridization between a top-fermenting yeast *Saccharomyces cerevisiae* and a cryophilic yeast *Saccharomyces bayanus*. *J Biosci Bioeng* 93(5):509-511.
113. Alexander WG, *et al.* (2016) Efficient engineering of marker-free synthetic allotetraploids of *Saccharomyces*. *Fungal Genet Biol* 89:10-17.
114. Gunge N & Nakatomi Y (1972) Genetic mechanisms of rare matings of the yeast *Saccharomyces cerevisiae* heterozygous for mating type. *Genetics* 70(1):41-58.
115. Barney M, Jansen G, & Helbert J (1980) Use of spheroplast fusion and genetic transformation to introduce dextrin utilization into *Saccharomyces uvarum*. *J Am Soc Brew Chem* 38(1):1-5.
116. Nikulin J, Krogerus K, & Gibson B (2018) Alternative *Saccharomyces* interspecies hybrid combinations and their potential for low-temperature wort fermentation. *Yeast* 35(1):113-127.
117. Gonçalves P, Valério E, Correia C, de Almeida JM, & Sampaio JP (2011) Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One* 6(6):e20739.
118. Boeke JD, Trueheart J, Natsoulis G, & Fink GR (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol* 154:164-175.
119. Chattoo BB, *et al.* (1979) Selection of *lys2* mutants of the yeast *Saccharomyces cerevisiae* by the utilization of  $\alpha$ -aminoadipate. *Genetics* 93(1):51-65.
120. Fukunaga M, Yielding LW, Firth III WJ, & Yielding KL (1980) Petite induction in *Saccharomyces cerevisiae* by ethidium analogs: Distinction between resting and growing cells. *Mutat Res* 78(2):151-157.
121. Krogerus K, Holmström S, & Gibson B (2018) Enhanced wort fermentation with *de novo* lager hybrids adapted to high-ethanol environments. *Appl Environ Microbiol* 84(4):e02302-02317.
122. Magalhães F, Krogerus K, Vidgren V, Sandell M, & Gibson B (2017) Improved cider fermentation performance and quality with newly generated *Saccharomyces cerevisiae*  $\times$  *Saccharomyces eubayanus* hybrids. *J Ind Microbiol Biotechnol* 44(8):1203-1213.
123. Goldstein AL & McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* 15(14):1541-1553.
124. Jimenez A & Davies J (1980) Expression of a transposable antibiotic resistance element in *Saccharomyces*. *Nature* 287(5785):869.
125. Gritz L & Davies J (1983) Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* 25(2):179-188.
126. Gorter de Vries AR, *et al.* (2019) Phenotype-independent isolation of interspecies *Saccharomyces* hybrids by dual-dye fluorescent staining and fluorescence-activated cell sorting. *Frontiers in Microbiology* 10:871.
127. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2017) Novel brewing yeast hybrids: creation and application. *Appl Microbiol Biotechnol* 101(1):65-78.
128. Holt S, Miks MH, de Carvalho BT, Foulquié-Moreno MR, & Thevelein JM (2018) The molecular biology of fruity and floral aromas in beer and other alcoholic beverages. *FEMS Microbiol Rev*.
129. Lodolo EJ, Kock JL, Axcell BC, & Brooks M (2008) The yeast *Saccharomyces cerevisiae*—the main character in beer brewing. *FEMS Yeast Res* 8(7):1018-1036.
130. Blicek L, *et al.* (2007) Isolation and characterization of brewer's yeast variants with improved fermentation performance under high-gravity conditions. *Appl Environ Microbiol* 73(3):815-824.
131. Huuskonen A, *et al.* (2010) Selection from industrial lager yeast strains of variants with improved fermentation performance in very-high-gravity worts. *Appl Environ Microbiol* 76(5):1563-1573.
132. Yu Z, *et al.* (2012) Selection of *Saccharomyces pastorianus* variants with improved fermentation performance under very high gravity wort conditions. *Biotechnol Lett* 34(2):365-370.
133. Zheng D-Q, *et al.* (2014) Genomic structural variations contribute to trait improvement during whole-genome shuffling of yeast. *Appl Microbiol Biotechnol* 98(7):3059-3070.
134. Voordeckers K, *et al.* (2015) Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genet* 11(11):e1005635.

135. Dinh TN, Nagahisa K, Hirasawa T, Furusawa C, & Shimizu H (2008) Adaptation of *Saccharomyces cerevisiae* cells to high ethanol concentration and changes in fatty acid composition of membrane and cell size. *PLoS One* 3(7):e2623.
136. Stanley D, Fraser S, Chambers PJ, Rogers P, & Stanley GA (2010) Generation and characterisation of stable ethanol-tolerant mutants of *Saccharomyces cerevisiae*. *J Ind Microbiol Biotechnol* 37(2):139-149.
137. Brown S & Oliver S (1982) Isolation of ethanol-tolerant mutants of yeast by continuous selection. *Appl. Microbiol. Biotechnol.* 16(2-3):119-122.
138. Çakar ZP, Seker UO, Tamerler C, Sonderegger M, & Sauer U (2005) Evolutionary engineering of multiple-stress resistant *Saccharomyces cerevisiae*. *FEMS Yeast Res* 5(6-7):569-578.
139. James TC, Usher J, Campbell S, & Bond U (2008) Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. *Curr Genet* 53(3):139-152.
140. Teunissen A, *et al.* (2002) Isolation and characterization of a freeze-tolerant diploid derivative of an industrial baker's yeast strain and its use in frozen doughs. *Appl Environ Microbiol* 68(10):4780-4787.
141. Böker-Schmitt E, Francisci S, & Schweyen R (1982) Mutations releasing mitochondrial biogenesis from glucose repression in *Saccharomyces cerevisiae*. *J Bacteriol* 151(1):303-310.
142. Jones R, Russell I, & Stewart G (1986) The use of catabolite derepression as a means of improving the fermentation rate of brewing yeast strains. *J Am Soc Brew Chem* 44(4):161-166.
143. Hockney RC & Freeman RF (1980) Gratuitous catabolite repression by glucosamine of maltose utilization in *Saccharomyces cerevisiae*. *Microbiology* 121(2):479-482.
144. Baker EP & Hittinger CT (2018) Evolution of a novel chimeric maltotriose transporter in *Saccharomyces eubayanus* from parent proteins unable to perform this function, *PLoS genetics* 15(4):e1007786.
145. Hong J & Gresham D (2014) Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet* 10(1):e1004041.
146. Gresham D, *et al.* (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* 4(12):e1000303.
147. Hope EA, *et al.* (2017) Experimental evolution reveals favored adaptive routes to cell aggregation in yeast. *Genetics:genetics.* 116.198895.
148. Holmberg S & Kielland-Brandt MC (1978) A mutant of *Saccharomyces cerevisiae* temperature sensitive for flocculation. Influence of oxygen and respiratory deficiency on flocculence. *Carlsberg Res Commun* 43(1):37-47.
149. Kutyna DR, *et al.* (2012) Adaptive evolution of *Saccharomyces cerevisiae* to generate strains with enhanced glycerol production. *Appl Microbiol Biotechnol* 93(3):1175-1184.
150. Belda I, Ruiz J, Navascués E, Marquina D, & Santos A (2016) Improvement of aromatic thiol release through the selection of yeasts with increased  $\beta$ -lyase activity. *Int J Food Microbiol* 225:1-8.
151. Chen Y, *et al.* (2012) Development of *Saccharomyces cerevisiae* producing higher levels of sulfur dioxide and glutathione to improve beer flavor stability. *Appl Biochem Biotechnol* 166(2):402-413.
152. Yoshida S, *et al.* (2008) Development of bottom-fermenting *Saccharomyces* strains that produce high SO<sub>2</sub> levels, using integrated metabolome and transcriptome analysis. *Appl Environ Microbiol* 74(9):2787-2796.
153. Shen N, Wang J, Liu C, Li Y, & Li Q (2014) Domesticating brewing yeast for decreasing acetaldehyde production and improving beer flavor stability. *Eur Food Res Technol* 238(3):347-355.
154. de Araújo Vicente M, *et al.* (2006) Isolation of *Saccharomyces cerevisiae* strains producing higher levels of flavoring compounds for production of "cachaça" the Brazilian sugarcane spirit. *Int J Food Microbiol* 108(1):51-59.
155. Strejcek J, Siříšková L, Karabín M, Almeida e Silva JB, & Brányik T (2013) Production of alcohol-free beer with elevated amounts of flavouring compounds using lager yeast mutants. *J I Brewing* 119(3):149-155.
156. Watanabe M, Nagai H, & Kondo K (1995) Properties of sake yeast mutants resistant to isoamyl monochloroacetate. *J Biosci Bioeng* 80(3):291-293.
157. Hirooka K, Yamamoto Y, Tsutsui N, & Tanaka T (2005) Improved production of isoamyl acetate by a sake yeast mutant resistant to an isoprenoid analog and its dependence on alcohol acetyltransferase activity, but not on isoamyl alcohol production. *J Biosci Bioeng* 99(2):125-129.
158. Asano T, Inoue T, Kurose N, Hiraoka N, & Kawakita S (1999) Improvement of isoamyl acetate productivity in sake yeast by isolating mutants resistant to econazole. *J Biosci Bioeng* 87(5):697-699.
159. Hirooka K, Ogita A, Fujita KI, Yamamoto Y, & Tanaka T (2010) Isolation of a Copper-resistant Sake Yeast Mutant with Improved Flavour Compound Production. *J I Brewing* 116(3):261-264.
160. Watanabe M, Tanaka N, Mishima H, & Takemura S (1993) Isolation of sake yeast mutants resistant to isoamyl monofluoroacetate to improve isoamyl acetate productivity. *J Biosci Bioeng* 76(3):229-231.
161. Lee S, Villa K, & Patino H (1995) Yeast strain development for enhanced production of desirable alcohols/esters in beer. *J Am Soc Brew Chem* 53(4):153-156.
162. Fukuda K, Watanabe M, Asano K, Ouchi K, & Takasawa S (1991) Isolation and genetic study of p-fluoro-dl-phenylalanine-resistant mutants overproducing  $\beta$ -phenethyl-alcohol in *Saccharomyces cerevisiae*. *Curr Genet* 20(6):449-452.
163. Akita O, Ida T, Obata T, & Hara S (1990) Mutants of *Saccharomyces cerevisiae* producing a large quantity of  $\beta$ -phenethyl alcohol and  $\beta$ -phenethyl acetate. *J Biosci Bioeng* 69(2):125-128.
164. Kielland-Brandt MC, Petersen JGL, & Mikkelsen JD (1979) Mutants in the biosynthesis of isoleucine in a non-mating, non-sporulating brewing strain of *Saccharomyces carlsbergensis*. *Carlsberg Res Commun* 44(1):27-36.

165. Gibson B, Vidgren V, Peddinti G, & Krogerus K (2018) Diacetyl control during brewery fermentation via adaptive laboratory engineering of the lager yeast *Saccharomyces pastorianus*. *J Ind Microbiol Biotechnol* 45(12):1103-1112.
166. Sprink T, Eriksson D, Schiemann J, & Hartung F (2016) Regulatory hurdles for genome editing: process-vs. product-based approaches in different regulatory contexts. *Plant Cell Rep* 35(7):1493-1506.
167. Nevoigt E (2008) Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 72(3):379-412.
168. Conko G, Kershen DL, Miller H, & Parrott WA (2016) A risk-based approach to the regulation of genetically engineered organisms. *Nat Biotechnol* 34(5):493.
169. Hino A (2002) Safety assessment and public concerns for genetically modified food products: the Japanese experience. *Toxicol Pathol* 30(1):126-128.
170. Ledford H (2016) Gene-editing surges as US rethinks regulations. *Nature* 532(7598):158.
171. Waltz E (2016) Gene-edited CRISPR mushroom escapes US regulation. *Nature* 532(7599):293.
172. Eriksson D, et al. (2018) A Welcome Proposal to Amend the GMO Legislation of the EU. *Trends Biotechnol* 36(11):1100-1103.
173. Pérez-Torrado R, Querol A, & Guillamón JM (2015) Genetic improvement of non-GMO wine yeasts: Strategies, advantages and safety. *Trends Food Sci Technol* 45(1):1-11.
174. Hammond JR (1995) Genetically-modified brewing yeasts for the 21st century. Progress to date. *Yeast* 11(16):1613-1627.
175. Ishii T & Araki M (2016) Consumer acceptance of food crops developed by genome editing. *Plant Cell Rep* 35(7):1507-1518.
176. Ishii T & Araki M (2017) A future scenario of the global regulatory landscape regarding genome-edited crops. *GM Crops Food* 8(1):44-56.
177. Mertens S, et al. (2019) Reducing phenolic off-flavors through CRISPR-based gene editing of the *FDC1* gene in *Saccharomyces cerevisiae* x *Saccharomyces eubayanus* hybrid lager beer yeasts. *PLoS One* 14(1):e0209124.
178. Penttilä M, Suihko M, Lehtinen U, Nikkola M, & Knowles J (1987) Construction of brewer's yeasts secreting fungal endo- $\beta$ -glucanase. *Curr Genet* 12(6):413-420.
179. Sakai K, Fukui S, Yabuuchi S, Aoyagi S, & Tsumura Y (1989) Expression of the *Saccharomyces diastaticus* *STA1* gene in brewing yeasts. *J Am Soc Brew Chem* 47(4):87-91.
180. Perry C & Meaden P (1988) Properties of a genetically-engineered dextrin-fermenting strain of brewer's yeast. *J I Brewing* 94(2):64-67.
181. Cole GE, et al. (1988) Stable Expression of *Aspergillus awamori* Glucoamylase in Distiller's Yeast. *Nat Biotechnol* 6(4):417.
182. Omura F, Fujita A, Miyajima K, & Fukui N (2005) Engineering of yeast Put4 permease and its application to lager yeast for efficient proline assimilation. *Biosci Biotechnol Biochem* 69(6):1162-1171.
183. Nevoigt E, et al. (2002) Genetic engineering of brewing yeast to reduce the content of ethanol in beer. *FEMS Yeast Res* 2(2):225-232.
184. Govender P, Domingo JL, Bester MC, Pretorius IS, & Bauer FF (2008) Controlled expression of the dominant flocculation genes *FLO1*, *FLO5*, and *FLO11* in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 74(19):6041-6052.
185. Verstrepen K, et al. (2001) Late fermentation expression of *FLO1* in *Saccharomyces cerevisiae*. *J Am Soc Brew Chem* 59(2):69-76.
186. Liu XF, et al. (2009) Expression of *GAI* gene and disruption of *PEP4* gene in an industrial brewer's yeast strain. *Lett Appl Microbiol* 49(1):117-123.
187. Wang J-j, Xu W-n, Li J, & Li Q (2014) Absence of *fks1p* in lager brewing yeast results in aberrant cell wall composition and improved beer flavor stability. *World J Microbiol Biotechnol* 30(6):1901-1908.
188. Hansen J & Kiehlbrandt MC (1996) Inactivation of *MET10* in brewer's yeast specifically increases SO<sub>2</sub> formation during beer production. *Nat Biotechnol* 14(11):1587.
189. Donalies UE & Stahl U (2002) Increasing sulphite formation in *Saccharomyces cerevisiae* by overexpression of *MET14* and *SSU1*. *Yeast* 19(6):475-484.
190. Tezuka H, Mori T, Okumura Y, Kitabatake K, & Tsumura Y (1992) Cloning of a gene suppressing hydrogen sulfide production by *Saccharomyces cerevisiae* and its expression in a brewing yeast. *J Am Soc Brew Chem* 50(4):130-133.
191. Hansen J (1999) Inactivation of *MXR1* abolishes formation of dimethyl sulfide from dimethyl sulfoxide in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 65(9):3915-3919.
192. Omura F (2008) Targeting of mitochondrial *Saccharomyces cerevisiae* *Ilv5p* to the cytosol and its effect on vicinal diketone formation in brewing. *Appl Microbiol Biotechnol* 78(3):503-513.
193. Yamano S, Kondo K, Tanaka J, & Inoue T (1994) Construction of a brewer's yeast having  $\alpha$ -acetolactate decarboxylase gene from *Acetobacter aceti* ssp. *xylinum* integrated in the genome. *J Biotechnol* 32(2):173-178.
194. Sone H, et al. (1988) Nucleotide sequence and expression of the *Enterobacter aerogenes* alpha-acetolactate decarboxylase gene in brewer's yeast. *Appl Environ Microbiol* 54(1):38-42.
195. Fujii T, et al. (1990) Application of a ribosomal DNA integration vector in the construction of a brewer's yeast having alpha-acetolactate decarboxylase activity. *Appl Environ Microbiol* 56(4):997-1003.

196. Blomqvist K, Suihko M-L, Knowles J, & Penttilä M (1991) Chromosomal integration and expression of two bacterial  $\alpha$ -acetolactate decarboxylase genes in brewer's yeast. *Appl Environ Microbiol* 57(10):2796-2803.
197. Yamano S, Tanaka J, & Inoue T (1994) Cloning and expression of the gene encoding  $\alpha$ -acetolactate decarboxylase from *Acetobacter aceti* ssp. *xylinum* in brewer's yeast. *J Biotechnol* 32(2):165-171.
198. Hirata D, Aoki S, Watanabe K-i, Tsukioka M, & Suzuki T (1992) Stable overproduction of isoamyl alcohol by *Saccharomyces cerevisiae* with chromosome-integrated multicopy *LEU4* genes. *Biosci Biotechnol Biochem* 56(10):1682-1683.
199. Verstrepen KJ, *et al.* (2003) Expression levels of the yeast alcohol acetyltransferase genes *ATF1*, *Lg-ATF1*, and *ATF2* control the formation of a broad range of volatile esters. *Appl Environ Microbiol* 69(9):5228-5237.
200. Kim B, Cho BR, & Hahn JS (2014) Metabolic engineering of *Saccharomyces cerevisiae* for the production of 2-phenylethanol via Ehrlich pathway. *Biotechnol Bioeng* 111(1):115-124.
201. Han S-Y, *et al.* (2009) Highly efficient synthesis of ethyl hexanoate catalyzed by CALB-displaying *Saccharomyces cerevisiae* whole-cells in non-aqueous phase. *J Mol Catal B Enzym* 59(1-3):168-172.
202. Denby CM, *et al.* (2018) Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. *Nat Commun* 9(1):965.
203. Guo X, *et al.* (2019) Enabling Heterologous Synthesis of Lupulones in the Yeast *Saccharomyces cerevisiae*. *Appl Biochem Biotechnol*:1-11.
204. Beekwilder J, *et al.* (2014) Polycistronic expression of a  $\beta$ -carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to  $\beta$ -ionone production. *J Biotechnol* 192:383-392.
205. Brochado AR, *et al.* (2010) Improved vanillin production in baker's yeast through *in silico* design. *Microb Cell Fact* 9(1):84.
206. Hansen EH, *et al.* (2009) *De novo* biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl Environ Microbiol* 75(9):2765-2774.
207. Asadollahi MA, *et al.* (2008) Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: effect of *ERG9* repression on sesquiterpene biosynthesis. *Biotechnol Bioeng* 99(3):666-677.
208. Gavira C, *et al.* (2013) Challenges and pitfalls of P450-dependent (+)-valencene bioconversion by *Saccharomyces cerevisiae*. *Metab Eng* 18:25-35.
209. Lee D, Lloyd ND, Pretorius IS, & Borneman AR (2016) Heterologous production of raspberry ketone in the wine yeast *Saccharomyces cerevisiae* via pathway engineering and synthetic enzyme fusion. *Microb Cell Fact* 15(1):49.
210. Beekwilder J, *et al.* (2007) Microbial production of natural raspberry ketone. *Biotechnol J* 2(10):1270-1279.

## Chapter 2: Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts

Arthur R. Gorter de Vries, Jack T. Pronk and Jean-Marc G. Daran

Chromosomal copy number variation (CCNV) plays a key role in evolution and health of eukaryotes. The unicellular yeast *Saccharomyces cerevisiae* is an important model for studying the generation, physiological impact and evolutionary significance of CCNV. Fundamental studies on this yeast have contributed to an extensive set of methods for analyzing and introducing CCNV. Moreover, these studies provided insight into the balance between negative and positive impacts of CCNV in evolutionary contexts. A growing body of evidence indicates that CCNV not only frequently occurs in industrial strains of *Saccharomyces* yeasts but is also a key contributor to the diversity of industrially relevant traits. This notion is further supported by the frequent involvement of CCNV in industrially relevant traits acquired during evolutionary engineering. This review describes recent developments in genome-sequencing and genome-editing techniques and discusses how these offer opportunities to unravel contributions of CCNV in industrial *Saccharomyces* strains, as well as to rationally engineer yeast chromosomal copy numbers and karyotypes.

Essentially as published in *Applied and Environmental Microbiology* 2017;83(11):e03206-16

Supplementary materials available online

<https://doi.org/10.1128/AEM.03206-16>

## Introduction

*Saccharomyces* yeasts are applied in a large and expanding number of industrial processes (1), ranging from traditional applications such as dough leavening (2), beer (3) and wine fermentation (4) to modern processes such as the production of first and second generation fuel ethanol (5, 6), other low-molecular-weight compounds (7) and heterologous proteins (8). Selection and improvement of yeast strains remain essential to meet the complex, diverse and continually changing performance criteria for industrial applications of *Saccharomyces* yeasts (9). Improving and extending yeast strain applications can be pursued by exploration of biodiversity, mating, interspecies hybridization, random mutagenesis and selection, evolutionary engineering, targeted genetic modification or a combination of these approaches (10).

Understanding the genetic basis for industrial performance is invaluable for focusing and accelerating microbial strain improvement. In prokaryotes, genetic variation among related strains and species predominantly encompasses the presence or absence of protein-encoding and regulatory sequences, as well as mutations in these sequences. In eukaryotes, including the *Saccharomyces* yeasts, differences in ploidy, *i.e.* variations in copy number of chromosomes, provides an important additional source of genetic diversity (11).

While most eukaryotic cells are euploid, *i.e.* their chromosomes all have the same copy number, aneuploidy is encountered in nature as well as in man-made contexts. In aneuploid cells, the copy number of one or more chromosomes differs from that of the remainder of the genome. Existence of stable aneuploidy cells implies that chromosomal copy number variation (CCNV) contributes to genetic and physiological diversity within eukaryotic species and, in multicellular eukaryotes, within organisms. The biological significance of CCNV is powerfully illustrated by its impacts on human health. Effects of CCNV of human X and Y chromosomes range from infertility (XXY) to mental retardation (XXXXY), while trisomies of other chromosomes can cause decreased lifespan, mental retardation and premature fetal death (12, 13). Spectacular CCNV occurs in most human cancer cell lines, leading to chromosome numbers of up to 90, and has been linked to the cancer hallmark of increased genome instability (14). Targeting of aneuploid cells is therefore considered a potential strategy for cancer therapy (15). Use of polyploid plants and animals in agriculture is related to their increased size and infertility (16, 17), while allopolyploid plants additionally combine industrially relevant traits from two parental genomes (18, 19). As will be discussed in this paper, CCNV is also an important phenomenon in industrial strains of *Saccharomyces* yeasts, whose history often involves prolonged domestication and/or industrial strain improvement.

*S. cerevisiae* is an important model for studying how aneuploidy arises during mitotic and meiotic cell division, how it affects growth and how it influences evolution of eukaryotes. These research fields are discussed in recent specialized review papers (20-22). The present paper specifically aims to review current knowledge on the analysis, occurrence and significance of CCNV in *Saccharomyces* yeasts in industrial contexts. To this end, we review methods for analyzing CCNV in yeast strains, the mechanisms by which CCNV can arise spontaneously or be induced in the laboratory and the mechanisms by which CCNV can negatively affect fitness of yeast cells. Subsequently, we discuss the occurrence and significance of CCNV for domestication and development of industrial strains of *Saccharomyces* yeasts and its relevance in evolutionary engineering.

## Methods for CCNV analysis in yeasts

Analysis of chromosomal copy numbers in yeasts predominantly relies on five, largely complementary methods (Fig. 1). Flow cytometry analysis of cells stained with fluorescent DNA-intercalating dyes, using reference strains for calibration, enables absolute quantification of cellular DNA content and overall ploidy (23). The choice of fluorescent dyes should consider excitation/emission spectra, RNA/DNA specificity, mutagenicity, effects on viability and the required accuracy (24). When the fluorescent dye does not compromise viability, fluorescence-assisted cell sorting (FACS) can be used to select cells with a deviating DNA content. FACS-based selection has enabled selection of mutants whose DNA content differed by less than 2 % from that of the parent population (25). While this FACS approach cannot select cells with specific chromosome amplifications or deletions, it can pre-select cells with a deviating overall DNA content.

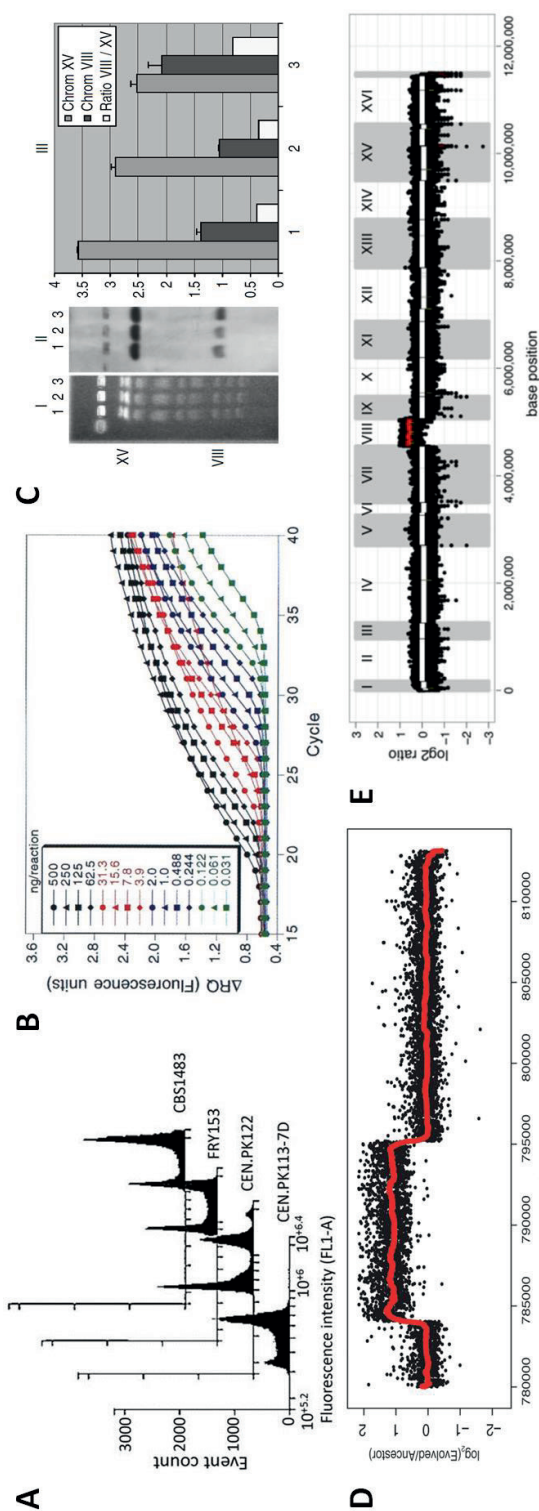
Contour-clamped homogeneous electric field (CHEF) electrophoresis separates yeast chromosomes on agarose gels and is used to analyze chromosome complements (karyotypes) of yeast strains (26, 27). Southern hybridization of CHEF gels can reveal copy numbers of individual chromosomes by comparison of hybridization intensity with reference strains (Fig. 1C) (28). However, the accuracy of CCNV estimates obtained by this method is limited.

Copy numbers of individual yeast chromosomes can be analyzed by quantitative real-time PCR (qPCR, Fig. 1B), using primers that amplify chromosome-specific genomic sequences (29). Accuracy of PCR-based copy-number estimates can be boosted by digital droplet PCR (ddPCR), which uses microfluidics to generate thousands of replicate PCR reactions in water-in-oil emulsions (30, 31). Since qPCR analysis only estimates copy numbers of the amplified region(s), additional methods are required to assess whether these reflect copy number variations of entire chromosomes or of specific chromosomal regions (segmental aneuploidy).

Array comparative genomic hybridization (aCGH) compares local copy number differences by hybridizing genomic DNA from related yeast strains to oligonucleotide arrays (Fig. 1D) (32). Depending on oligonucleotide size and genome coverage of the arrays, copy number variations can be analyzed across entire genomes at resolutions down to 20 bp (33).

High-resolution, accurate analysis of CCNV in yeast increasingly depends on 'next generation' sequencing (NGS) of entire yeast genomes (34). NGS enables ploidy estimation from allele frequency in the whole genome and in specific regions (35). Moreover, when sequence bias in DNA isolation and/or sequencing (36) is prevented, the number of reads generated for any particular sequence (*i.e.*, its read depth) directly reflects its copy number relative to the remainder of the genome (Fig. 1E) (37). Computational tools assist CCNV identification via read depth, either by mapping of NGS reads to a pre-assembled genome sequence or via *de novo* genome assembly (38). With both approaches, accuracy of copy-number estimates increases with increasing sequencing coverage. When many copies of a chromosome are present in a yeast strain, (dis)appearance of a single copy only causes a small relative change. Accurate analysis of aneuploid yeast genomes with large variations in chromosomal copy numbers therefore requires high sequencing coverage. Short-read-length NGS methods currently provide the most cost-effective access to high sequencing depth (> 100 x coverage at read lengths from 75 to 400 bp can be obtained routinely with, for example Illumina and Ion Torrent platforms). Sequencing reads can be mapped to a pre-assembled, accurate reference genome similar to that of the sequenced strain, yielding accurate CCNV estimates.





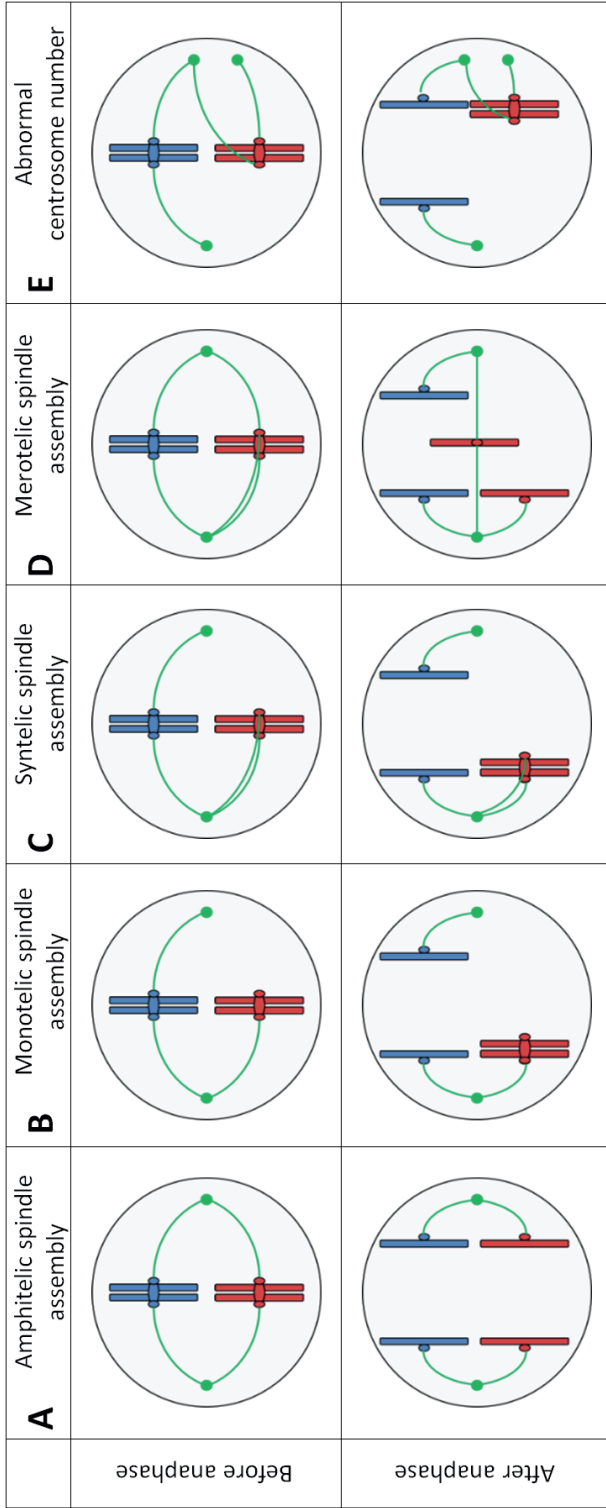
**Figure 1: Methods to analyze chromosome copy number and DNA content in yeast cells.** (A) Absolute quantification of the DNA content of strain CBS 1483 by flow cytometry using the DNA intercalating dye Sytox Green and calibration with three strains of known ploidy. Figure adapted from Van den Broek *et al.* (23). (B) qPCR fluorescence profiles for different initial concentrations of a template DNA sequence can be used to infer the amount of initial template in a reaction and to calculate relative copy numbers of different parts of the template DNA. Figure from Heid *et al.* (164). (C) Chromosome copy number determination of *S. cerevisiae* variants using contour-clamped homogeneous electric field electrophoresis and southern blotting. I, stained CHEF gel, II southern blot hybridization and III, quantification of the hybridization bands. Lane 1 and 2 show two disomic knock-out strains that only have a single copy of chromosome VIII, while lane 3 shows a diploid control strain. Figure adapted from Waghmare and Bruschi (28). (D) Copy number estimation of chromosome II by array comparative genomic hybridization of an evolved strain relative to its unevolved parental strain. Deviating copy number can be detected by significant deviations of the measured signal and has been accentuated by a red line. Figure from Gresham *et al.* (108). (E) Copy number estimation of the genome of the wine production strain VL3, based on whole-genome sequencing and read depth analysis. A marked increase of the read depth for chromosome VIII indicates a gain of copy of that chromosome. Figure adapted from Borneman *et al.* (117).

If no such reference genome is available, *de novo* assembly of the genome and subsequent copy number analysis can provide unbiased and more accurate results (23). However, short-read-length NGS does not allow assembly of repetitive regions whose length exceeds the read length, such as TY-elements, subtelomeres and rDNA sequences in *Saccharomyces* genomes. *De novo* genome assembly is strongly facilitated by long-read-length sequencing platforms (*e.g.* Pacific Biosystems, Oxford Nanopore Technologies), either alone or combined with short-read-length data. Moreover, when genes are present in multiple non-identical copies, it can be difficult to perform full reconstruction of duplicated alleles ('phasing') (39). Indeed, when two SNPs occur only in one copy of a gene, nucleotides can only be assigned to a specific allele if individual reads are available that cover both variable positions. Allelic reconstruction, and by extension reconstruction of (parts of) chromosome copies, is enhanced by the use of long-read or mate-pair sequencing data (39). Long-read sequencing technologies still have higher error rates than short-read platforms. Fast developments in real-time, single-molecule methods for replication (Pacific Biosystems) or nanopore (Oxford Nanopore Technologies) sequencing enable generation of extremely long reads with increasing accuracy (40-43), and are likely to transform whole-genome resequencing (44). The potential of long-read sequencing to capture entire chromosome arms or even entire chromosomes within a single read offers unique possibilities to unravel chromosome structure, translocation breakpoints and allelic variation among duplicate chromosomes and chromosomal fragments (41).

### **Induction of chromosome mis-segregation**

The anaphase of the eukaryotic cell cycle has evolved to conserve chromosomal copy number during cell division. Its crucial steps include chromatid cohesion, centrosome formation at opposite cell poles, kinetochore-microtubule attachment and quality control at the spindle assembly checkpoint (45). Imperfections in any of these steps can cause chromosome mis-segregation and, thereby, CCNV in eukaryotic cell populations, tissues and tumours (45-47). Even in cell lines without predisposing defects, chromosome mis-segregation occurs, albeit at very low frequencies (21, 48). In yeast, chromosome mis-segregation can occur during mitosis (48) and, with a higher incidence, during the meiotic process of sporulation (49). Fig. 2 provides a schematic overview of mechanisms by which mis-segregation of chromosomes can occur.

A wide range of chemical and physical stress factors increase the incidence of chromosome mis-segregation in growing cultures. Stimuli that increase occurrence of CCNV in mitotic yeast cultures include nutrient-limited growth (50), heat shock (51), UV- or X-ray-irradiation (52) and chemical stress. Chemical compounds such as nocodazole, fumaronitrile and methyl benzimidazole-2-yl-carbamate induce a high incidence of chromosome mis-segregation in *S. cerevisiae* (53-55). Polar aprotic solvents, including ethanol esters, are other known inducers of CCNV (56) and, high concentrations of ethanol itself have also been reported to enhance chromosome mis-segregation in fungal cells (57). Exposure to high ethanol concentrations may therefore contribute to the frequent occurrence of CCNV in industrial yeast strains used for production of alcoholic beverages and fuel ethanol (see below).



**Figure 2: Schematic representation of chromosome segregation and of the common mechanisms leading to chromosome mis-segregation.** Two chromatids of two different chromosomes are shown in red and blue, with their centromeres and kinetochores. In green, the centrosomes are shown with the assembled microtubule attached to the kinetochores of the chromatids. For each case, the microtubule-kinetochore assembly is shown before and after the anaphase. **(A)** Correct chromosome segregation is achieved by amphitelic spindle assembly, where microtubules connect each chromatid to a different centrosome, resulting in separation to opposite cellular poles during anaphase and maintaining a stable karyotype in the daughter cells (45). **(B and C)** If only one of the chromatids is attached to a centromere or both chromatids are attached to the same kinetochore, referred to as monotelic and syntelic respectively, proceeding to anaphase would result in the mis-segregation of both chromatids to that centromere. However, monotelic and syntelic spindle assemblies are detected at the spindle assembly checkpoint and therefore rarely cause chromosome mis-segregation. **(D)** In the case of a merotelic spindle assembly, a chromatid is attached to both centrosomes and, as a result, cannot migrate to a cellular pole. The resulting random segregation of the lagging chromosome can cause mis-segregation, damage and micronucleus formation (165). **(E)** When more than two centrosomes are formed, random attachment of chromatids can result in chromosome mis-segregation due to chromosome lagging or unequal chromosome segregation (166).

Chromosome mis-segregation can also be stimulated by genetic factors. Increased ploidy strongly enhances chromosome mis-segregation (58), in particular when uneven numbers of chromosome sets preclude equal distribution of chromosomes during meiosis (59). Strongly increased chromosome mis-segregation rates have also been observed in allopolyploid *Saccharomyces* yeasts, which carry chromosomes from different parental species and show a high incidence of aneuploidy (60). Since aneuploidy itself, including segmental aneuploidy, also stimulates chromosome mis-segregation, aneuploid cells are more prone to acquire further CCNV (61).

In contrast to chemical, physical and genetic stresses, which affect segregation of all chromosomes, targeted molecular genetic approaches enable elimination or amplification of specific chromosomes. In *S. cerevisiae*, copy gain and loss of specific chromosomes has been achieved by cloning a strong inducible promoter upstream of the centromere of the targeted chromosome (62, 63). When induced, transcription from the promoter interferes with centromere function, thus causing mis-segregation during mitosis. Aneuploid daughter cells that have lost or gained a copy of the targeted chromosome can then be isolated from the resulting culture. Alternatively, by crossing with *kar1* null mutants, mating is prematurely aborted but chromosome transfer between nuclei can still occur, yielding aneuploid cells. Aneuploidy of specific chromosomes can be easily selected for when the carrying marker sequences (64).

### **Negative impacts of CCNV on fitness**

Aneuploid yeasts typically show a reduced fitness relative to congeneric euploid strains (64). The molecular basis of generic transcriptional responses to aneuploidy remain to be fully elucidated. Reported transcriptional responses in aneuploid strains include downregulation of genes involved in cell growth and proliferation and upregulation of genes involved in the environmental stress response (ESR) (64, 65). Studies on the impact of gain or loss of chromosomes in otherwise euploid yeast strains showed that the aneuploidy-associated stress response (AASR) includes increased genome instability, low sporulation efficiency, reduced growth rate, increased nutrient uptake rates and reduced replicative lifespan (21, 66, 67). Phenotypic consequences of chromosome gain and loss are similar, suggesting that the responsible cellular mechanisms overlap (68). AASR intensity is positively correlated with the length of the affected chromosome(s) and with the number of affected genes (20, 64, 69). A much less pronounced AASR in polyploid strains has been attributed to a smaller relative impact on chromosome number (64, 70). Absence of AASR-related phenotypes upon introduction of a yeast artificial chromosome harbouring non-transcribed mammalian genes indicates that AASR is not due to increased DNA content *per se* (64).

Genome instability of aneuploid yeasts has been linked to the mis-segregation events that cause aneuploidy and, in particular, to 'lagging' (Figure 2D) of chromosomes during anaphase. DNA damage and imperfect repair of lagging chromosomes cause mutations, deletions and translocations (71, 72). Additionally, formation of transient micronuclei by lagging chromosomes increases mutation rate during subsequent mitosis (73, 74). At a longer time scale, aneuploidy promotes generation of CCNV by enhancing chromosome mis-segregation and mitotic recombination as well as by impairing DNA repair (61, 75, 76). Impaired sporulation of aneuploid strains has been linked to disruption of homologous chromosome pairing during meiosis (77). AASR-related cell-cycle defects involve slow accumulation of G1 cyclins, causing an abnormal delay in the G1 phase (78, 79).

CCNV-associated changes in gene dosage can directly affect expression levels of the affected genes. Typically, gain or loss of a chromosome coincides with an increased or decreased expression level, respectively, of the large majority of expressed genes that it carries (80). Correct subunit folding and assembly of multi-protein complexes (29, 64, 81, 82), which strongly depend on subunit stoichiometry (83), can be disturbed when one or more subunits are encoded by aneuploid chromosomes. A resulting 'overload' of the cellular protein folding machinery can cause accumulation of un- and misfolded proteins and proteotoxic stress (67, 70). Indeed, some aneuploid strains show increased sensitivity to inhibitors of protein folding and degradation (84) and impaired functionality of the proteasome, the chaperone Hsp90 or endocytosis-mediated protein degradation (66, 70). Energy costs of protein misfolding and protein overproduction have been implicated in the increased nutrient consumption and slow growth of aneuploid yeast strains (85). The correlation between protein level and gene copy number is not always straightforward (29, 64) and situations have even been described in which the transcript level of individual genes decreased with increasing copy number (86-88). Signaling cascades and transcriptional regulation are among the core cellular systems that can be affected by aneuploidy (89). The impact of gene-dosage related changes in gene expression on AASR (29) can be further intensified or attenuated by mutations in genes on non-aneuploid chromosomes (90). Such *in trans* effects can, for example, be related to stoichiometric imbalances in protein complexes or pathways, unspecific protein interactions, protein folding and degradation (81).

Sensitivity to AASR is yeast strain dependent (91, 92). In tolerant strains, mutations were identified that attenuate AASR, such as a loss-of-function mutation in the deubiquitinating enzyme Ubp6p (82). While not all mutations involved in AASR tolerance are known, its relevance is amply demonstrated by the frequent occurrence of aneuploidy in wild, clinical and industrial isolates of *Saccharomyces* yeasts (35, 91, 93).

### **CCNV in evolutionary engineering**

In addition to negative impacts on cellular fitness, chromosome-specific effects of CCNV can also confer fitness benefits in specific environmental or genetic contexts. Indeed, CCNV offers a fast way to modify gene copy number during natural evolution of eukaryotes and to increase evolvability by allowing neofunctionalization of amplified essential genes (51, 94-96). Under selective conditions, mutants with CCNV will outgrow the parental population whenever positive effects of CCNV on fitness outweigh any negative impacts of AASR, while further mutations that enhance positive effects or decrease AASR can further increase the initial fitness benefit. CCNV is therefore seen as a significant contributor to evolutionary adaptation in eukaryotes (51, 97).

Technically, adaptive laboratory evolution (ALE) encompasses prolonged cultivation of microorganisms under a defined conditions, combined with an analysis of the phenotypic and and/or genotypic changes that occur during evolutionary adaptation (98). ALE approaches that are specifically designed to select for industrially relevant traits are often referred to as evolutionary engineering (99, 100). Resequencing of evolved strains can provide insight into the genetic basis for industrially relevant traits and enable its reverse engineering into naïve, non-evolved strains (101). Evolutionary engineering is particularly attractive for food and beverage applications, since it does not involve recombinant-DNA techniques and associated consumer-acceptance and regulatory issues (102).

While, on the time scales involved in natural evolution and speciation, CCNV is considered to be a transient adaptation mechanism that is usually replaced by more elegant and efficient mutations (103, 104), most ALE experiments with yeasts cover only 50 to 500 generations of selective growth. It is therefore not surprising that CCNV are frequently encountered during ALE of *Saccharomyces* yeasts, for example for the selection of suppressor mutants (Table 1). Numerous evolutionary engineering studies have linked CCNV to industrially relevant traits, ranging from tolerance to products or inhibitors to improved kinetics of sugar fermentation or sedimentation behavior of yeast cultures (Table 1). In some cases, ALE even resulted in complete duplication of the genome of haploid *S. cerevisiae* strains, for instance after selection for glucose-limited growth, high ethanol tolerance and increased sedimentation (105-107). In the latter case, increased ploidy played a major role in shaping an evolved, multicellular phenotype.

In addition to whole-chromosome copy number variations, ALE frequently involves segmental aneuploidies (108-111). While both can be identified by analysis of high-coverage, short-read NGS data, precise definition of duplication and/or translocation events and karyotypes involved in segmental aneuploidy generally requires additional analysis by long-read sequencing or diagnostic PCR (110, 111).

Several methods can be applied to test if segmental or whole-chromosome aneuploidies do indeed contribute to phenotypes acquired in an ALE experiment. In some cases, hypothesis-based deletion or amplification of one or more genes on (an) affected chromosome(s) can directly confirm the relevance of a CCNV. For example, an increased copy number of chromosome III in *jen1Δ* mutants evolved for restoration of lactate transport could be rapidly linked to the *ADY2* monocarboxylate-transporter gene on this chromosome (112). Overexpression or deletion studies were also successfully used to identify 17 genes that contributed to the benefit of a copy gain of chromosome III in a *S. cerevisiae* strain evolved for heat tolerance (103). Alternatively, the relevance of a CCNV in an evolved strain can be tested by introducing the deviating chromosome copy number in a euploid strain, *e.g.* via transcriptional interference with centromere function (103, 113). Similarly, the chromosome copy number variation can be reverted to wild type, *e.g.* by sporulation and analysis of segregants with wild-type karyotypes (103, 113). Although not routinely applied, specific chromosomal regions that contribute to an acquired phenotype can be identified by targeted introduction of segmental aneuploidy of sets of tiled chromosomal regions (114). Two recently described PCR-based methods enable duplication or deletion of chromosome-segment copies by introduction of telomere seed sequences and of an additional centromere to generate an additional autonomously replicating chromosome fragment. By introducing a centromere and telomere seed sequences pointing outward of the region of interest, this region will be duplicated on an additional, independently replicating chromosome (115). Conversely, by introducing a centromere and telomere seed sequences pointing into the region of interest, the targeted chromosome is split into two autonomously replicating chromosomes that no longer contain the targeted region (116). This approach enables a non-biased, systematic analysis of the positive and negative contributions of chromosomal regions and/or individual genes.

**Table 1: Examples of whole-chromosome copy number variations acquired during laboratory evolution experiments with *Saccharomyces cerevisiae* strains.** In the examples listed in the Table, the acquired CCNV was hypothesized to contribute to the selected phenotype. ‘Confirmed causality’ indicates that a causal link between CCNV and the phenotype acquired during laboratory evolution was experimentally confirmed. In cases where the impact of a CCNV on phenotype was linked to specific genes, this is also indicated. Segmental aneuploidies observed in the cited studies are not included in the Table.

<b>Selected phenotype</b>	<b>Aneuploid chromosomes</b>	<b>Confirmed causality</b>	<b>Contributing genes</b>	<b>Reference</b>
Biomass sedimentation	whole genome duplication	Yes	<i>ACE2</i>	(105)
Glucose-limited growth	whole genome duplication	Yes	-	(107)
High temperature tolerance	III (+1)	Yes	17 individual genes	(103)
High pH tolerance	V (+1)	Yes	-	(103)
Glucose-limited growth	I (+1), III (+1) and V (+1)	No	-	(108)
Phosphate-limited growth	IV (+1), VI (+1), X (+1), XIII (+2) and XVI (+1)	No	-	(108)
Lactate utilization by <i>jen1Δ</i> strain	III (+1)	Yes	<i>ADY2</i>	(112)
Xylose utilization	I (-1)	No	-	(157)
<i>p</i> -coumaric and ferulic acid tolerance	XIV (+1)	No	-	(157)
Copper tolerance	II (+1) and VIII (+1)	No	<i>CUP1</i> , <i>SCO1</i> and <i>SCO2</i>	(104)
Galactose tolerance	VIII (+1)	Yes	<i>GAL80</i>	(158)
Ethanol tolerance	III (+1) and VIII (+1)	No	-	(106)
Radical resistance	XV (+1)	Yes	<i>STI1</i> and <i>PDR5</i>	(113)
Fluconazole resistance	VIII (+1)	No	<i>ERG11</i>	(113)
Tunicamycin resistance	XVI (-1)	Yes	-	(113)
Benomyl resistance	XII (-1)	No	-	(113)
Suppressors of <i>MEC1</i> deficiency	IV (+1)	Yes	<i>RNR1</i>	(159)
Suppressors of <i>MYO1</i> deletion	XIII (+1) and XVI (+1)	Yes	<i>HSP82</i> , <i>HSC82</i> , <i>RLM1</i> and <i>MKK2</i>	(94)
Suppressors of <i>RPS24A</i> and <i>RNR1</i> deletion	IX (+1)	No	<i>RPS24B</i> and <i>RNR3</i>	(160)
Suppressors of telomerase insufficiency	VIII (-1)	No	<i>PRP8</i> , <i>UTP9</i> , <i>KOG1</i> and <i>SCH9</i>	(161)

### CCNV in industrial *Saccharomyces* yeasts

Aneuploidy has been observed in *Saccharomyces* strains used in diverse industrial applications, including dough leavening, bioethanol production, beer brewing, spirits production, wine fermentation and production of cacao and coffee (Table 2). In industrial strains, CCNV may have occurred during centuries-long domestication processes and/or during strain improvement programs that involved CCNV-inducing mutagenesis procedures such as UV irradiation (52).

Currently available information suggests that aneuploidy is not prevalent among *S. cerevisiae* strains used in dough leavening, bioethanol production, ale-type-beer fermentation and distilled-beverage production. In these strains, aneuploidy typically involves small deviations in copy number of one or a few chromosomes (117-119). Since accurate information is available for only few of the many hundreds of such strains stored in culture collections, the incidence of CCNV may well be underestimated. Indeed, a recent whole-genome sequencing study revealed extensive CCNV among several beer-related *S. cerevisiae* strains that were previously assumed to be mostly euploid (93).

There is ample evidence that copy numbers of individual genes or loci affect industrially relevant traits of *S. cerevisiae* strains. For example, rates of sucrose, maltose and melibiose fermentation correlate with copy numbers of *SUC*, *MAL* and *MEL* loci, respectively (120-122), while proline utilization rates correlate with the copy number of the *PUT1* proline oxidase gene (123). So far, the industrial significance of CCNV in industrial *S. cerevisiae* strains has not been systematically explored. *S. cerevisiae* ZTW1, a strain isolated from corn mash used in a Chinese bioethanol factory, provides an interesting exception. In this strain, chromosomal and segmental aneuploidy were shown to directly contribute to industrially relevant traits, including copper tolerance and ethanol yield (124).

Consistent with the increased rate of chromosome mis-segregation in allopolyploid cells, aneuploidy is highly prevalent among wine and lager-type beer yeasts originating from domestication of natural hybrids of different *Saccharomyces* species. Despite its frequent occurrence, the impacts of aneuploidy in these genetic contexts has not been explored in depth and it is unclear how AASR and chromosome-specific copy number effects compare to those observed in otherwise euploid *S. cerevisiae* strains. In general, these allopolyploid genomes tolerate aneuploidy well, with massive diversity in chromosome copy numbers across strains (23, 125, 126). Some aneuploid lager brewing yeasts even sporulate, albeit at low efficiency, by anomalous cell division (79). Wine yeasts include *S. cerevisiae* x *S. kudriavzevii*, *S. cerevisiae* x *S. uvarum* and *S. cerevisiae* x *S. kudriavzevii* x *S. uvarum* hybrids (127, 128), many of which are alloaneuploids, with a large diversity in chromosome copy numbers (129). Aneuploidy has a strong impact on performance of 'flor' wine yeast. An increased copy number of chromosome VII, which carries the alcohol dehydrogenase genes *ADH2* and *ADH3*, correlated with increased ethanol oxidation capacity of the characteristic vellum formed by these yeasts during sherry wine fermentation (130).



**Table 2: Examples of chromosome copy number variation (CCNV) in industrial *Saccharomyces* strains.** The overall ploidy of the strains and identified aneuploid chromosomes are indicated. For strains in which the copy number deviation from euploidy has been determined, this is reported between parentheses. Extensive aneuploidy refers to strains with more than 10 aneuploid chromosomes. Segmental aneuploidies that occur in many of these strains are not indicated in the Table.

Strain	Species	Industrial product	Approximate overall ploidy	Aneuploid chromosomes	Reference
BR001	<i>S. cerevisiae</i>	Bread	4n	IX (+1)	(93)
BR004	<i>S. cerevisiae</i>	Bread	4n	IX (+1)	(93)
E-IM3	<i>S. cerevisiae</i>	Cacao	3n	VII	(162)
AY529517	<i>S. cerevisiae</i>	Cacao	2n	IV, XII	(162)
YE 2-2	<i>S. cerevisiae</i>	Coffee	3n	I, XV, XVI	(162)
JV2	<i>S. cerevisiae</i>	Coffee	4n	Extensive aneuploidy	(162)
Y-393	<i>S. cerevisiae</i>	Kefyr	3n	I, III, IX	(162)
YJM1356	<i>S. cerevisiae</i>	Cider	2n	I (+2)	(147)
YJM1439	<i>S. cerevisiae</i>	Ginger beer	2n	VIII (+2)	(147)
FostersO	<i>S. cerevisiae</i>	Ale beer	>2n	III (+1), XIV (-1)	(117)
FostersB	<i>S. cerevisiae</i>	Ale beer	>2n	III (+1), V (+1), XV (+1)	(117)
CBS 1483	<i>S. cerevisiae</i> x <i>S. eubayanus</i>	Lager beer	>2n	Extensive aneuploidy	(23)
CBS 1270	<i>S. cerevisiae</i> x <i>S. eubayanus</i>	Lager beer	>2n	Extensive aneuploidy	(23)
AWRI796	<i>S. cerevisiae</i>	Wine	2n	I (+1)	(117)
VL3	<i>S. cerevisiae</i>	Wine	2n	VIII (+1)	(117)
F-12	<i>S. cerevisiae</i>	Flor Wine	2n	VII (+1), XIII (+2)	(130)
SA001	<i>S. cerevisiae</i>	Sake	2-3n	V (+1)	(93)
SA003	<i>S. cerevisiae</i>	Sake	2-3n	I (+1)	(93)
SP011	<i>S. cerevisiae</i>	Spirits	2n	I (-1), III (-1), VI (-1), IX (-1), XII (-1)	(93)
SP001	<i>S. cerevisiae</i>	Spirits	2n	I (-1), VI (-1)	(93)
Y-999	<i>S. cerevisiae</i>	Bioethanol from starch	3n	III	(162)
CBS 7960	<i>S. cerevisiae</i>	Bioethanol from sugarcane	2n	VIII	(162)
ZTW1	<i>S. cerevisiae</i>	Bioethanol from corn mash	3n	IX (+1)	(163)

*S. pastorianus* lager beer brewing strains have long been assumed to originate from a hybridization event involving *S. cerevisiae* and another *Saccharomyces* species (131). The genome of the cold-tolerant species *S. eubayanus*, first isolated in Patagonia in 2011 (132) and later also found in North-America, Asia and New Zealand (132-135) was shown to exhibit a 99.56 % identity with the non-*cerevisiae* part of *S. pastorianus* genomes (136). It is postulated that, after one or more spontaneous hybridization events, centuries of domestication and selection of the resulting *S. cerevisiae* x *S. eubayanus* hybrid(s) in brewing environments, generated the current diversity of lager brewing strains (137, 138). *S. cerevisiae* x *S. eubayanus* hybrids made in the laboratory combine at least two important brewing related characteristics of their parents. The *S. cerevisiae* subgenome contributes the ability to ferment maltotriose, a major fermentable sugar in wort, while low-temperature performance, essential for the lager brewing process, is conferred by the *S. eubayanus* subgenome (139, 140).

Historically and mainly based on geographical origin, two groups of *S. pastorianus* strains were distinguished. Group I ('Saaz type') strains tend to ferment well at low temperatures but generally show poor maltotriose fermentation. Conversely, Group II strains ('Frohberg type') tend to have higher optimal growth temperatures and ferment maltotriose well (141). These phenotypic differences correlate with ploidy and with the contribution of genetic material from the two subgenomes. Consistent with their better performance at low temperature, Group I strains contain more *S. eubayanus* DNA, while some *S. cerevisiae* chromosomes can even be absent (e.g. *S. cerevisiae* chromosome III is absent in all group I strains sequenced so far) (23, 32, 141-143); Group II strains generally have a more balanced genome composition, with (multiple) chromosomes from both *S. eubayanus* and *S. cerevisiae* (23, 32, 141-143). These differences have been proposed to reflect different hybridization histories of the two groups (144). In this model, Group I derives from an original hybridization event involving a haploid *S. cerevisiae* strain and a haploid or diploid *S. eubayanus* strain, while Group II strains arose from hybridization of a diploid *S. cerevisiae* strain with a haploid *S. eubayanus* strain (23) or from two subsequent hybridization events (141). Different hybridization histories appear to be contradicted by conserved chromosome rearrangement breakpoints in Group I and Group II strains (32, 143). However, these might also have evolved independently due to fragility of the breakpoint and/or a by conferring a selective advantage (145). The latter hypothesis is consistent with ALE studies with a *S. uvarum* X *S. cerevisiae* hybrid in nitrogen-limited cultures, which selected for recombination between allopolyploid chromosomes in the *MEP2* ammonium permease gene (146).

Two key brewing-related properties of *S. pastorianus* strains have been correlated with CCNV. Production of diacetyl, an important off flavour in lager beers that needs to be removed at the end of fermentation ('Ruh' phase), correlated with copy number of chromosomes III, VIII, X, XII and XIV (23). These chromosomes harbor genes involved in the valine biosynthesis pathway, which generates  $\alpha$ -acetolactate, the precursor for diacetyl production. Similarly,  $\text{Ca}^{2+}$ -dependent flocculation, which is essential for yeast sedimentation during brewing, positively correlated with copy numbers of chromosomes I, VI, XI and XII, all of which harbor flocculin genes (23).

### **Outlook: understanding and engineering CCNV in industrial contexts**

Whole-genome sequences of environmental and industrial isolates of *Saccharomyces* species, which are becoming available at a rapid and still accelerating pace (35, 93, 147), confirm the relevance of CCNV for the natural diversity, domestication and industrial strain improvement of these yeasts.

Experimental hybridization of strains from different *Saccharomyces* species rapidly gains popularity as a strategy for strain improvement and product diversification of wine and beer yeasts (139, 148, 149). Traits that have been improved by hybridization include fermentative vigor over wide temperature ranges and concentrations of minor fermentation products (150), flocculation capacity (151) and sugar uptake kinetics (152). Moreover, ploidy of laboratory-made hybrid strains correlates with fermentation rates, ethanol yield and concentrations of aromatic esters (148). In view of the higher tendency of allopolyploid and allopolyploid genomes to develop aneuploidy, CCNV is likely to be a key factor in the stability and further diversification of the resulting strains.

Targeted introduction of CCNV, *e.g.* by using drugs that interfere with chromosome segregation, is rarely applied in industrial strain improvement (10). Use of the mitotic inhibitor methyl benzimidazole-2-yl-carbamate (MBC) to mutagenize the aneuploid bioethanol strain ZTW1 demonstrates the potential of this approach (153). Treatment of strain ZTW1 with MBC yielded strains with an improved fermentative capacity under industrial high-gravity conditions (119), enhanced viability after drying (154) and higher final ethanol titer (124). These observations and the frequent appearance of CCNV in ALE suggest that such interference with chromosome segregation may deserve reconsideration in industrial yeast strain improvement.

The relatively small number of cases in which molecular mechanisms by which CCNV contributes to industrial performance of *Saccharomyces* yeasts have been investigated in detail, often identified gene-dosage effects as a key contributor. Allelic variation of amplified genes can be an additional, as yet underexplored, source of industrially relevant diversity within strains that carry CCNV, especially in allopolyploid strains with a long history of domestication and/or strain improvement. Novel long-read DNA-sequencing approaches (*e.g.* nanopore MinION sequencing (41)) should enable a much faster identification of such allelic variations and of their correlation with industrially relevant traits, including subtle differences in flavor and aroma production. Recent developments in genome editing, including the advent of CRISPR-based techniques (155, 156) and methods for experimentally introducing defined, segmental aneuploidies (115, 116) will accelerate the functional analysis of CCNV. Moreover, these techniques will enable rapid introduction of relevant mutations into strains that do not contain CCNV, without the potential disadvantages of AASR. The combination of these developments will enable a more thorough investigation of the importance of CCNV for the performance of industrial strains and is likely to open the way to using CCNV induction as a tool for strain improvement, either by direct generation of improved strains or by identification of chromosome fragments or genes whose copy number affects industrial performance.

### **Acknowledgments**

We thank Nick Brouwers, Alex Salazar, Jasper Diderich, Niels Kuijpers (HEINEKEN Supply Chain B.V.) and Jan-Maarten Geertman (HEINEKEN Supply Chain B.V.) for critically reading the manuscript. The authors acknowledge permission by publishers to reuse already published graphical information in Fig. 1. For Fig. 1C, permission to reuse the figure was granted by publisher John Wiley and Sons under license number 3931880399375.

## References

1. Mattanovich D, Sauer M, & Gasser B (2014) Yeast biotechnology: teaching the old dog new tricks. *Microb Cell Fact* 13:34.
2. Bell PJL, Higgins VJ, & Attfield PV (2001) Comparison of fermentative capacities of industrial baking and wild-type yeasts of the species *Saccharomyces cerevisiae* in different sugar media. *Lett Appl Microbiol* 32(4):224-229.
3. Lodolo EJ, Kock JLF, Axcell BC, & Brooks M (2008) The yeast *Saccharomyces cerevisiae* - the main character in beer brewing. *Fems Yeast Res* 8(7):1018-1036.
4. Pretorius IS (2000) Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast* 16(8):675-729.
5. Argueso JL, et al. (2009) Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res* 19(12):2258-2270.
6. Moysés DN, Reis VCB, de Almeida JRM, de Moraes LMP, & Torres FAG (2016) Xylose Fermentation by *Saccharomyces cerevisiae*: challenges and prospects. *Int J Mol Sci* 17(3):207.
7. Zeng A-P & Biebl H (2002) Bulk chemicals from biotechnology: the case of 1,3-propanediol production and the new trends. *Tools and Applications of Biochemical Engineering Science*, (Springer), pp 239-259.
8. Mendoza-Vega O, Sabatie J, & Brown SW (1994) Industrial production of heterologous proteins by fed-batch cultures of the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol Rev* 15(4):369-410.
9. Pretorius IS, du Toit M, & van Rensburg P (2003) Designer Yeasts for the Fermentation Industry of the 21<sup>st</sup> Century. *Food Technol Biotech* 41(1):3-10.
10. Steensels J, et al. (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol Rev* 38(5):947-995.
11. Comai L (2005) The advantages and disadvantages of being polyploid. *Nat Rev Genet* 6(11):836-846.
12. Visootsak J & Graham Jr JM (2006) Klinefelter syndrome and other sex chromosomal aneuploidies. *Orphanet J Rare Dis* 1(42):1-5.
13. Hassold T & Hunt P (2001) To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet* 2(4):280-291.
14. Duesberg P, Rausch C, Rasnick D, & Hehlmann R (1998) Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc Natl Acad Sci USA* 95(23):13692-13697.
15. Gordon DJ, Resio B, & Pellman D (2012) Causes and consequences of aneuploidy in cancer. *Nat Rev Genet* 13(3):189-203.
16. D'Hont A (2005) Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenet Genome Res* 109(1-3):27-33.
17. Piferrer F, et al. (2009) Polyploid fish and shellfish: production, biology and applications to aquaculture for performance improvement and genetic containment. *Aquaculture* 293(3):125-156.
18. Lee S-S, Lee S-A, Yang J, & Kim J (2011) Developing stable progenies of *×Brassicoraphanus*, an intergeneric allopolyploid between *Brassica rapa* and *Raphanus sativus*, through induced mutation using microspore culture. *Theor Appl Genet* 122(5):885-891.
19. Leitch AR & Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. *Science* 320(5875):481-483.
20. Sheltzer JM & Amon A (2011) The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends Genet* 27(11):446-453.
21. Santaguida S & Amon A (2015) Short- and long-term effects of chromosome mis-segregation and aneuploidy. *Nat Rev Mol Cell Biol* 16(8):473-485.
22. Mulla W, Zhu J, & Li R (2014) Yeast: a simple model system to study complex phenomena of aneuploidy. *FEMS Microbiol Rev* 38(2):201-212.
23. van den Broek M, et al. (2015) Chromosomal copy number variation in *Saccharomyces pastorianus* is evidence for extensive genome dynamics in industrial lager brewing strains. *Appl Environ Microbiol* 81(18):6253-6267.
24. Haase SB & Reed SI (2002) Improved flow cytometric analysis of the budding yeast cell cycle. *Cell cycle* 1(2):117-121.
25. Pfosser M, Amon A, Lelley T, & Heberle-Bors E (1995) Evaluation of sensitivity of flow cytometry in detecting aneuploidy in wheat using disomic and ditelosomic wheatrye addition lines. *Cytometry* 21:387-393.
26. Chu G, Vollrath D, & Davis RW (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* 234(4783):1582-1585.
27. Török T, Rockhold D, & King AD (1993) Use of electrophoretic karyotyping and DNA-DNA hybridization in yeast identification. *Int J Food Microbiol* 19(1):63-80.
28. Waghmare SK & Bruschi CV (2005) Differential chromosome control of ploidy in the yeast *Saccharomyces cerevisiae*. *Yeast* 22(8):625-639.
29. Pavelka N, et al. (2010) Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468(7321):321-325.
30. Bozdag GO & Greig D (2014) The genetics of a putative social trait in natural populations of yeast. *Mol Ecol* 23(20):5061-5071.
31. Hindson BJ, et al. (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical chemistry* 83(22):8604-8610.

32. Dunn B & Sherlock G (2008) Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res* 18(10):1610-1623.
33. Gresham D, Dunham MJ, & Botstein D (2008) Comparing whole genomes using DNA microarrays. *Nat Rev Genet* 9(4):291-302.
34. Xie C & Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC bioinformatics* 10(1):80.
35. Zhu YO, Sherlock GJ, & Petrov DA (2016) Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3 (Bethesda)* 6(8):2121-34.
36. van Dijk EL, Jaszczyszyn Y, & Thermes C (2014) Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res* 322(1):12-20.
37. Yoon S, Xuan Z, Makarov V, Ye K, & Sebat J (2009) Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res* 19(9):1586-1592.
38. Zhao M, Wang Q, Wang Q, Jia PL, & Zhao Z (2013) Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* 14(Suppl 11):S1.
39. Chin C-S, *et al.* (2016) Phased diploid genome assembly with single molecule real-time sequencing. *Nat Methods* 13(12):1050-1054.
40. Goodwin S, *et al.* (2015) Oxford Nanopore sequencing, hybrid error correction, and *de novo* assembly of a eukaryotic genome. *Genome Res* 25(11):1750-1756.
41. Istace B, *et al.* (2016) *de novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6(2):1-13.
42. Rhoads A & Au KF (2015) PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13(5):278-289.
43. McIlwain SJ, *et al.* (2016) Genome sequence and analysis of a stress-tolerant, wild-derived strain of *Saccharomyces cerevisiae* used in biofuels research. *G3* 6(6):1757-1766.
44. Branton D, *et al.* (2008) The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26(10):1146-1153.
45. Thompson SL, Bakhoun SF, & Compton DA (2010) Mechanisms of chromosomal instability. *Curr Biol* 20(6):R285-295.
46. Parry JM, Sharp D, & Parry EM (1979) Detection of mitotic and meiotic aneuploidy in the yeast *Saccharomyces cerevisiae*. *Environmental health perspectives* 31:97-111.
47. Compton DA (2011) Mechanisms of aneuploidy. *Current opinion in cell biology* 23(1):109-113.
48. Zhu YO, Siegal ML, Hall DW, & Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci USA* 111(22):E2310-2318.
49. Parry EM & Cox BS (1970) The tolerance of aneuploidy in yeast. *Genetical research* 16(3):333-340.
50. Adams J, Puskas-Rozsa S, Simlar J, & Wilke CM (1992) Adaptation and major chromosomal changes in populations of *Saccharomyces cerevisiae*. *Curr Genet* 22(1):13-19.
51. Chen G, Rubinstein B, & Li R (2012) Whole chromosome aneuploidy: big mutations drive adaptation by phenotypic leap. *BioEssays : news and reviews in molecular, cellular and developmental biology* 34(10):893-900.
52. Parry JM, Sharp D, Tippins RS, & Parry EM (1979) Radiation-induced mitotic and meiotic aneuploidy in the yeast *Saccharomyces cerevisiae*. *Mutat Res* 61(1):37-55.
53. Albertini S & Zimmermann FK (1991) The detection of chemically induced chromosomal malsegregation in *Saccharomyces cerevisiae* D61.M: a literature survey (1984-1990). *Mutat Res* 258(3):237-258.
54. Liu M, Grant SG, Macina OT, Klopman G, & Rosenkranz HS (1997) Structural and mechanistic bases for the induction of mitotic chromosomal loss and duplication ('malsegregation') in the yeast *Saccharomyces cerevisiae*: relevance to human carcinogenesis and developmental toxicology. *Mutat Res* 374(2):209-231.
55. Wood JS (1982) Genetic effects of methyl benzimidazole-2-yl-carbamate on *Saccharomyces cerevisiae*. *Mol Cell Biol* 2(9):1064-1079.
56. Zimmermann FK, Mayer VW, Scheel I, & Resnick MA (1985) Acetone, methyl ethyl ketone, ethyl acetate, acetonitrile and other polar aprotic solvents are strong inducers of aneuploidy in *Saccharomyces cerevisiae*. *Mutat Res* 149(3):339-351.
57. Crebelli R, Conti G, Conti L, & Carere A (1989) A comparative study on ethanol and acetaldehyde as inducers of chromosome malsegregation in *Aspergillus nidulans*. *Mutat Res Fund Mol Mech Mut* 215(2):187-195.
58. Storchova Z (2014) Ploidy changes and genome stability in yeast. *Yeast* 31(11):421-430.
59. Mayer VW & Aguilera A (1990) High levels of chromosome instability in polyploids of *Saccharomyces cerevisiae*. *Mutat Res* 231(2):177-186.
60. Delneri D, *et al.* (2003) Engineering evolution to study speciation in yeasts. *Nature* 422(6927):68-72.
61. Sheltzer JM, *et al.* (2011) Aneuploidy Drives Genomic Instability in Yeast. *Science* 333(6045):1026-1030.
62. Reid RJD, *et al.* (2008) Chromosome-scale genetic mapping using a set of 16 conditionally stable *Saccharomyces cerevisiae* chromosomes. *Genetics* 180(4):1799-1808.
63. Anders KR, *et al.* (2009) A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. *BMC genetics* 10(1):36.
64. Torres EM, *et al.* (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317(5840):916-924.

65. Sheltzer JM, Torres EM, Dunham MJ, & Amon A (2012) Transcriptional consequences of aneuploidy. *Proc Natl Acad Sci USA* 109(31):12644-12649.
66. Sunshine AB, *et al.* (2016) Aneuploidy shortens replicative lifespan in *Saccharomyces cerevisiae*. *Aging cell* 15(2):317-324.
67. Oromendia AB & Amon A (2014) Aneuploidy: implications for protein homeostasis and disease. *Dis Model Mech* 7(1):15-20.
68. Beach RR (2016) Insights into the consequences of chromosome gains and losses in *S. cerevisiae*. (Massachusetts Institute of Technology).
69. Tang Y-C & Amon A (2013) Gene copy-number alterations: a cost-benefit analysis. *Cell* 152(3):394-405.
70. Oromendia AB, Dodgson SE, & Amon A (2012) Aneuploidy causes proteotoxic stress in yeast. *Genes & development* 26(24):2696-2708.
71. Janssen A, van der Burg M, Szuhai K, Kops GJPL, & Medema RH (2011) Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* 333(6051):1895-1898.
72. Crasta K, *et al.* (2012) DNA breaks and chromosome pulverization from errors in mitosis. *Nature* 482(7383):53-58.
73. Hoffelder DR, *et al.* (2004) Resolution of anaphase bridges in cancer cells. *Chromosoma* 112(8):389-397.
74. Terradas M, Martín M, Tusell L, & Genescà A (2009) DNA lesions sequestered in micronuclei induce a local defective-damage response. *DNA repair* 8(10):1225-1234.
75. Blank HM, Sheltzer JM, Meehl CM, & Amon A (2015) Mitotic entry in the presence of DNA damage is a widespread property of aneuploidy in yeast. *Molecular biology of the cell* 26(8):1440-1451.
76. Skoneczna A, Kaniak A, & Skoneczny M (2015) Genetic instability in budding and fission yeast—sources and mechanisms. *FEMS Microbiol Rev* 39(6):917-967.
77. Anderson E & Martin PA (1975) The sporulation and mating of brewing yeasts. *J Inst Brew* 81(3):242-247.
78. Thorburn RR, *et al.* (2013) Aneuploid yeast strains exhibit defects in cell growth and passage through START. *Molecular biology of the cell* 24(9):1274-1289.
79. Niwa O, Tange Y, & Kurabayashi A (2006) Growth arrest and chromosome instability in aneuploid yeast. *Yeast* 23(13):937-950.
80. Hughes TR, *et al.* (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19(4):342-347.
81. Dephoure N, *et al.* (2014) Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *eLife* 3:e03023.
82. Torres EM, *et al.* (2010) Identification of aneuploidy-tolerating mutations. *Cell* 143(1):71-83.
83. Li GW, Burkhardt D, Gross C, & Weissman JS (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157(3):624-635.
84. Tang YC, Williams BR, Siegel JJ, & Amon A (2011) Identification of aneuploidy-selective antiproliferation compounds. *Cell* 144(4):499-512.
85. Geiler-Samerotte KA, *et al.* (2011) Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc Natl Acad Sci USA* 108(2):680-685.
86. Birchler JA, Bhadra U, Bhadra MP, & Auger DL (2001) Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Developmental biology* 234(2):275-288.
87. Berman J (2016) Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. *FEMS Yeast Res* 16(3):fow020.
88. Gasch AP, *et al.* (2016) Further support for aneuploidy tolerance in wild yeast and effects of dosage compensation on gene copy-number evolution. *eLife* 5:e14409.
89. Veitia RA, Bottani S, & Birchler JA (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet* 24(8):390-397.
90. Dodgson SE, *et al.* (2016) Chromosome-specific and global effects of aneuploidy in *Saccharomyces cerevisiae*. *Genetics* 202(4):1395-1409.
91. Hose J, *et al.* (2015) Dosage compensation can buffer copy-number variation in wild yeast. *eLife* 4:e05462.
92. Cromie GA & Dudley AM (2015) Aneuploidy: Tolerating Tolerance. *Curr Biol* 25(17):R771-773.
93. Gallone B, *et al.* (2016) Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166(6):1397-1410. e1316.
94. Rancati G, *et al.* (2008) Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* 135(5):879-893.
95. Rancati G & Pavelka N (2013) Karyotypic changes as drivers and catalyzers of cellular evolvability: a perspective from non-pathogenic yeasts. *Seminars in cell & developmental biology* 24(4):332-338.
96. Voordeckers K & Verstrepen KJ (2015) Experimental evolution of the model eukaryote *Saccharomyces cerevisiae* yields insight into the molecular mechanisms underlying adaptation. *Current opinion in microbiology* 28:1-9.
97. Kohn LM (2005) Mechanisms of fungal speciation. *Annual review of phytopathology* 43:279-308.
98. Dragosits M & Mattanovich D (2013) Adaptive laboratory evolution—principles and applications for biotechnology. *Microb Cell Fact* 12(1):1.
99. Sauer U (2001) Evolutionary engineering of industrially important microbial phenotypes. *Metab Eng*, (Springer), pp 129-169.

100. Çakar ZP, Turanlı-Yıldız B, Alkim C, & Yılmaz Ü (2012) Evolutionary engineering of *Saccharomyces cerevisiae* for improved industrially important properties. *FEMS Yeast Res* 12(2):171-182.
101. Oud B, van Maris AJA, Daran J-MG, & Pronk JT (2012) Genome-wide analytical approaches for reverse metabolic engineering of industrially relevant phenotypes in yeast. *FEMS Yeast Res* 12(2):183-196.
102. Bachmann H, Pronk JT, Kleerebezem M, & Teusink B (2015) Evolutionary engineering to enhance starter culture performance in food fermentations. *Curr Opin Biotechnol* 32:1-7.
103. Yona AH, et al. (2012) Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad Sci USA* 109(51):21010-21015.
104. Gerstein AC, et al. (2015) Too much of a good thing: the unique and repeated paths toward copper adaptation. *Genetics* 199(2):555-571.
105. Oud B, et al. (2013) Genome duplication and mutations in *ACE2* cause multicellular, fast-sedimenting phenotypes in evolved *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 110(45):E4223-E4231.
106. Voordeekers K, et al. (2015) Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genet* 11(11):e1005635.
107. Venkataram S, et al. (2016) Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell* 166(6):1585-1596. e1522.
108. Gresham D, et al. (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* 4(12):e1000303.
109. González-Ramos D, et al. (2016) A new laboratory evolution approach to select for constitutive acetic acid tolerance in *Saccharomyces cerevisiae* and identification of causal mutations. *Biotechnol Biofuels* 9(1):173.
110. Payen C, et al. (2014) The dynamics of diverse segmental amplifications in populations of *Saccharomyces cerevisiae* adapting to strong selection. *G3 (Bethesda)* 4(3):399-409.
111. Dunham MJ, et al. (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99(25):16144-16149.
112. de Kok S, et al. (2012) Laboratory evolution of new lactate transporter genes in a *jen1Δ* mutant of *Saccharomyces cerevisiae* and their identification as *ADY2* alleles by whole-genome resequencing and transcriptome analysis. *FEMS Yeast Res* 12(3):359-374.
113. Chen GB, Bradford WD, Seidel CW, & Li R (2012) *Hsp90* stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature* 482(7384):246-250.
114. Sunshine AB, et al. (2015) The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS biology* 13(5):e1002155.
115. Natesuntorn W, et al. (2015) Genome-wide construction of a series of designed segmental aneuploids in *Saccharomyces cerevisiae*. *Sci Rep* 5.
116. Sugiyama M, et al. (2008) PCR-mediated one-step deletion of targeted chromosomal regions in haploid *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* 80(3):545-553.
117. Borneman AR, et al. (2011) Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet* 7(2):e1001287.
118. Codon AC, Benitez T, & Korhola M (1998) Chromosomal polymorphism and adaptation to specific industrial environments of *Saccharomyces* strains. *Appl Microbiol Biotechnol* 49(2):154-163.
119. Zheng DQ, et al. (2014) Genomic structural variations contribute to trait improvement during whole-genome shuffling of yeast. *Appl Microbiol Biotechnol* 98(7):3059-3070.
120. Naumov GI, Naumova ES, & Louis EJ (1995) Genetic mapping of the  $\alpha$ -galactosidase *MEL* gene family on right and left telomeres of *Saccharomyces cerevisiae*. *Yeast* 11(5):481-483.
121. Carlson M, Celenza JL, & Eng FJ (1985) Evolution of the dispersed *SUC* gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biology* 5(11):2894-2902.
122. Brown CA, Murray AW, & Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Current biology : CB* 20(10):895-903.
123. Ibáñez C, et al. (2014) Comparative genomic analysis of *Saccharomyces cerevisiae* yeasts isolated from fermentations of traditional beverages unveils different adaptive strategies. *Int J Food Microbiol* 171:129-135.
124. Zhang K, et al. (2015) Genomic reconstruction to improve bioethanol and ergosterol production of industrial yeast *Saccharomyces cerevisiae*. *J Ind Microbiol Biotechnol* 42(2):207-218.
125. Querol A & Bond U (2009) The complex and dynamic genomes of industrial yeasts. *FEMS Microbiol Lett* 293(1):1-10.
126. Tadami H, Shikata-Miyoshi M, & Ogata T (2014) Aneuploidy, copy number variation and unique chromosomal structures in bottom-fermenting yeast revealed by array-CGH. *J Inst Brew* 120(1):27-37.
127. Gonzalez SS, Barrio E, Gafner J, & Querol A (2006) Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res* 6(8):1221-1234.
128. Peris D, et al. (2012) The molecular characterization of new types of *Saccharomyces cerevisiae* x *S. kudriavzevii* hybrid yeasts unveils a high genetic diversity. *Yeast* 29(2):81-91.
129. Dunn B, Richter C, Kvitsek DJ, Pugh T, & Sherlock G (2012) Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res* 22(5):908-924.

130. Guijo S, Mauricio J, Salmon J, & Ortega J (1997) Determination of the relative ploidy in different *Saccharomyces cerevisiae* strains used for fermentation and 'flor' film ageing of dry sherry-type wines. *Yeast* 13(2):101-117.
131. Nakao Y, et al. (2009) Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Res* 16(2):115-129.
132. Libkind D, et al. (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA* 108(35):14539-14544.
133. Gayevskiy V & Goddard MR (2016) *Saccharomyces eubayanus* and *Saccharomyces arboricola* reside in North Island native New Zealand forests. *Environ Microbiol*.
134. Peris D, et al. (2014) Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Mol Ecol* 23(8):2031-2045.
135. Bing J, Han P-J, Liu W-Q, Wang Q-M, & Bai F-Y (2014) Evidence for a Far East Asian origin of lager beer yeast. *Current biology : CB* 24(10):R380-R381.
136. Peris D, et al. (2016) Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast *Saccharomyces eubayanus*. *PLoS Genet* 12(7):e1006155.
137. Gibson B & Liti G (2015) *Saccharomyces pastorianus*: genomic insights inspiring innovation for industry. *Yeast* 32(1):17-27.
138. Wendland J (2014) Lager yeast comes of age. *Eukaryot Cell* 13(10):1256-1265.
139. Heblly M, et al. (2015) *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res* 15(3):fov005.
140. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2015) New lager yeast strains generated by interspecific hybridization. *J Ind Microbiol Biotechnol* 42(5):769-778.
141. Okuno M, et al. (2016) Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA research : an international journal for rapid publication of reports on genes and genomes* 23(1):67-80.
142. Walther A, Hesselbart A, & Wendland J (2014) Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. *G3 (Bethesda)* 4(5):783-793.
143. Hewitt SK, Donaldson IJ, Lovell SC, & Delneri D (2014) Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. *PLoS One* 9(3):e92203.
144. Baker EC, et al. (2015) The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol Biol Evol* 32(11):2818-2831.
145. Gordon JL, Byrne KP, & Wolfe KH (2009) Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet* 5(5):e1000485.
146. Dunn B, et al. (2013) Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet* 9(3):e1003366.
147. Strobe PK, et al. (2015) The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res* 25(5):762-774.
148. Krogerus K, et al. (2016) Ploidy influences the functional attributes of *de novo* lager yeast hybrids. *Appl Microbiol Biotechnol*:1-20.
149. Bellon JR, et al. (2011) Newly generated interspecific wine yeast hybrids introduce flavour and aroma diversity to wines. *Appl Microbiol Biotechnol* 91(3):603-612.
150. Zambonelli C, et al. (1997) Technological properties and temperature response of interspecific *Saccharomyces* hybrids. *J Sci Food Agr* 74(1):7-12.
151. Coloretti F, Zambonelli C, & Tini V (2006) Characterization of flocculent *Saccharomyces* interspecific hybrids for the production of sparkling wines. *Food Microbiol* 23(7):672-676.
152. Tronchoni J, Gamero A, Arroyo-Lopez FN, Barrio E, & Querol A (2009) Differences in the glucose and fructose consumption profiles in diverse *Saccharomyces* wine species and their hybrids during grape juice fermentation. *Int J Food Microbiol* 134(3):237-243.
153. Hou LH, Meng M, Guo L, & He JY (2015) A comparison of whole cell directed evolution approaches in breeding of industrial strain of *Saccharomyces cerevisiae*. *Biotechnol Lett* 37(7):1393-1398.
154. Zheng D, et al. (2013) Construction of novel *Saccharomyces cerevisiae* strains for bioethanol active dry yeast (ADY) production. *PLoS One* 8(12):e85022.
155. Mans R, et al. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 15(2):fov004.
156. DiCarlo JE, et al. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res*:gkt135.
157. Sato TK, et al. (2014) Harnessing genetic diversity in *Saccharomyces cerevisiae* for fermentation of xylose in hydrolysates of alkaline hydrogen peroxide-pretreated biomass. *Appl Environ Microbiol* 80(2):540-554.
158. Sirr A, et al. (2015) Allelic variation, aneuploidy, and nongenetic mechanisms suppress a monogenic trait in yeast. *Genetics* 199(1):247-262.
159. Gasch AP, et al. (2001) Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Molecular biology of the cell* 12(10):2987-3003.



160. Hughes TR, *et al.* (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet* 25(3):333-337.
161. Millet C & Makovets S (2016) Aneuploidy as a mechanism of adaptation to telomerase insufficiency. *Curr Genet*:1-8.
162. Ludlow CL, *et al.* (2016) Independent origins of yeast associated with coffee and cacao fermentation. *Curr Biol* 26(7):965-971.
163. Zhang K, *et al.* (2016) Genomic structural variation contributes to phenotypic change of industrial bioethanol yeast *Saccharomyces cerevisiae*. *FEMS Yeast Res* 16(2):fov118.
164. Heid CA, Stevens J, Livak KJ, & Williams PM (1996) Real time quantitative PCR. *Genome Res* 6(10):986-994.
165. Musacchio A & Salmon ED (2007) The spindle-assembly checkpoint in space and time. *Nat Rev Mol Cell Biol* 8(5):379-393.
166. Ganem NJ, Godinho SA, & Pellman D (2009) A mechanism linking extra centrosomes to chromosomal instability. *Nature* 460(7252):278-282.

## Chapter 3: Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D

Alex N. Salazar <sup>#</sup>, Arthur R. Gorter de Vries <sup>#</sup>, Marcel van den Broek, Melanie Wijsman, Pilar de la Torre Cortés, Anja Brickwedde, Nick Brouwers, Jean-Marc G. Daran and Thomas Abeel

<sup>#</sup> These authors contributed equally to this publication and should be considered co-first authors.

The haploid *Saccharomyces cerevisiae* strain CEN.PK113-7D is a popular model system for metabolic engineering and systems biology research. Current genome assemblies are based on short-read sequencing data scaffolded based on homology to strain S288C. However, these assemblies contain large sequence gaps, particularly in subtelomeric regions, and the assumption of perfect homology to S288C for scaffolding introduces bias. In this study, we obtained a near-complete genome assembly of CEN.PK113-7D using only Oxford Nanopore Technology's MinION sequencing platform. 15 of the 16 chromosomes, the mitochondrial genome, and the 2-micron plasmid are assembled in single contigs and all but one chromosome starts or ends in a telomere cap. This improved genome assembly contains 770 Kbp of added sequence containing 248 gene annotations in comparison to the previous assembly of CEN.PK113-7D. Many of these genes encode functions determining fitness in specific growth conditions and are therefore highly relevant for various industrial applications. Furthermore, we discovered a translocation between chromosomes III and VIII which caused misidentification of a *MAL* locus in the previous CEN.PK113-7D assembly. This study demonstrates the power of long-read sequencing by providing a high-quality reference assembly and annotation of CEN.PK113-7D and places a caveat on assumed genome stability of microorganisms.

Essentially as published in FEMS Yeast Research 2017;17(7):fox074

Supplementary materials available online

<https://doi.org/10.1093/femsyr/fox074>

## Introduction

Whole Genome Sequencing (WGS) reveals important genetic information of an organism which can be linked to specific phenotypes and enable genetic engineering approaches (1, 2). Short-read sequencing has become the standard method for WGS in the past years due to its low cost, high sequencing accuracy and high output of sequence reads. In most cases, the obtained read data is used to reassemble the sequenced genome either by *de novo* assembly or by mapping the reads to a previously-assembled closely-related genome. However, the sequence reads obtained are relatively short: between 35 and 1000 bp (3). This poses challenges as genomes have long stretches of repetitive sequences of several thousand nucleotides in length and can only be characterized if a read spans the repetitive region and has a unique fit to the flanking ends (4). As a result, *de novo* genome assembly based on short-read technologies “break” at repetitive regions preventing reconstruction of whole chromosomes. The resulting assembly consists of dozens to hundreds of sequence fragments, commonly referred to as *contigs*. These contigs are then either analysed independently or ordered and joined together adjacently based on their alignment to a closely-related reference genome. However, referenced based joining of contigs into so-called *scaffolds*, is based on the assumption that the genetic structure of the sequenced strain is identical to that of the reference genome—potentially concealing existing genetic variation.

Previous genome assemblies of the *Saccharomyces cerevisiae* strain CEN.PK113-7D have been based on homology with the fully-assembled reference genome of *S. cerevisiae* strain S288C (5, 6). CEN.PK113-7D is a haploid strain used as a model organism in biotechnology-related research and systems biology because of its convenient growth characteristics, its robustness under industrially-relevant conditions, and its excellent genetic accessibility (6-9). CEN.PK113-7D was sequenced using a combination of 454 and Illumina short-read libraries and a draft genome was assembled consisting of over 700 contigs (6). After scaffolding using MAIA (10) and linking based on homology with the genome of S288C, it was possible to reconstruct all 16 chromosomes. However, there were large sequence gaps within chromosomes and the subtelomeric regions were left unassembled, both of which could contain relevant open reading frames (ORFs) (6). Assuming homology to S288C, more than 90 % of missing sequence was located in repetitive regions corresponding mostly to subtelomeric regions and Ty-elements. These regions are genetically unstable as repeated sequences promote recombination events (11); therefore the assumption of homology with S288C could be unjustified. Ty-elements are present across the genome: repetitive sequences with varying length (on average ~6 Kbp) resulting from introgressions of viral DNA (12). Subtelomeric regions are segments towards the end of chromosomes consisting of highly repetitive elements making them notoriously challenging to reconstruct using only short-read sequencing data (13). While Ty-elements are likely to have limited impact on gene expression, subtelomeric regions harbour various so-called subtelomeric genes. Several gene families are present mostly in subtelomeric regions and typically have functions determining the cell’s interaction with its environment; such as nutrient uptake (14-16), sugar utilisation (17), and inhibitor tolerance (18). Many of these subtelomeric gene families therefore contribute to the adaptation of industrial strains to the specific environment they are used in. For example, the *RTM* and *SUC* gene families are relevant for bioethanol production as they increase inhibitor-tolerance in molasses and utilization of extracellular sucrose, respectively (14, 18). Similarly, *MAL* genes enable utilization of maltose and maltotriose and *FLO* genes enable calcium-dependent flocculation, both of which are crucial for the beer brewing industry (19-21). As is the case for Ty-elements, subtelomeric regions are unstable due

to repetitive sequences and homology to various regions of the genome, which is likely to cause diversity across strains (6, 11, 21). Characterizing and accurately localizing subtelomeric gene families is thus crucial for associating strain performance to specific genomic features and for targeted engineering approaches for strain improvement (13).

In contrast to short-read technologies, single-molecule sequencing technologies can output sequence reads of several thousand nucleotides in length. Recent developments of long-read sequencing technologies have decreased the cost and increased the accuracy and output, yielding near-complete assemblies of diverse yeast strains (22, 23). For example, *de novo* assembly of a biofuel production *S. cerevisiae* strain using PacBio reads produced a genome assembly consisting of 25 chromosomal contigs scaffolded into 16 chromosomes. This assembly revealed 92 new genes relative to S288C amongst which 28 previously uncharacterized and unnamed genes. Interestingly, many of these genes had functions linked to stress tolerance and carbon metabolism which are functions critical to the strains industrial application (22). In addition, rapid technological advances in nanopore sequencing have matured as a competitive long-read sequencing technology and the first yeast genomes assembled using nanopore reads are appearing (22-26). For example, Istace *et al.* sequenced 21 wild *S. cerevisiae* isolates and their genome assemblies ranged between 18 and 105 contigs enabling the detection of 29 translocations and 4 inversions relative to the chromosome structure of reference S288C. In addition, large variations were found in several difficult to sequence subtelomeric genes such as *CUP1*, which was correlated to large differences in copper tolerance (25). Nanopore sequencing has thus proven to be a potent technology for characterizing yeast.

In this study, we sequenced CEN.PK113-7D using Oxford Nanopore Technology's (ONT) MinION sequencing platform. This nanopore *de novo* assembly was compared to the previous short-read assembly of CEN.PK113-7D (6) with particular attention for previously, poorly-assembled subtelomeric regions and for structural variation potentially concealed due to the assumption of homology to S288C.

## Materials and methods

### Yeast strains

The *Saccharomyces cerevisiae* strain "CEN.PK113-7D Frankfurt" (*MATa MAL2-8c*) was kindly provided by dr. P. Kötter in 2016 (6, 27). It was plated on solid YPD (containing 10 g/l yeast extract, 20 g/l peptone and 20 g/l glucose) upon arrival and a single colony was grown once until stationary phase in liquid YPD medium and 1 mL aliquots with 30 % glycerol were stored at -80°C since. The previously sequenced CEN.PK113-7D sample was renamed "CEN.PK113-7D Delft" (6). It was obtained from the same source in 2001 and 1 mL aliquots with 30 % glycerol were stored at -80°C with minimal propagation since (no more than three cultures on YPD as described above).

### Yeast cultivation and genomic DNA extraction

Yeast cultures were incubated in 500-ml shake-flasks containing 100 ml liquid YPD medium at 30°C on an orbital shaker set at 200 rpm until the strains reached stationary phase with an OD<sub>660</sub> between 12 and 20. Genomic DNA of CEN.PK113-7D Delft and CEN.PK113-7D Frankfurt for whole genome sequencing was isolated using the Qiagen 100/G kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and quantified using a Qubit® Fluorometer 2.0 (ThermoFisher Scientific, Waltham, MA).

### **Short-read Illumina sequencing**

Genomic DNA of CEN.PK113-7D Frankfurt was sequenced on a HiSeq2500 sequencer (Illumina, San Diego, CA) with 150 bp paired-end reads using PCR-free library preparation by Novogene Bioinformatics Technology Co., Ltd (Yuen Long, Hong Kong). All Illumina sequencing data are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the bioproject accession number PRJNA393501.

### **MinION Sequencing**

MinION genomic libraries were prepared using either nanopore Sequencing Kit SQK-MAP006 (2D-ligation for R7.3 chemistry), SQK-RAD001 (Rapid library prep kit for R9 chemistries) or SQK-MAP007 (2D-ligation for R9 chemistries) (Oxford Nanopore Technologies, Oxford, United Kingdom). Two separate libraries of SQK-MAP006 and one library of SQK-RAD001 were used to sequence CEN.PK113-7D Delft. Only one SQK-MAP007 library was used to sequence CEN.PK113-7D Frankfurt. With the exception of the SQK-RAD001 library, all libraries used 2-3 µg of genomic DNA fragmented in a Covaris g-tube (Covaris) with the “8-10 kbp fragments” settings according to manufacturer’s instructions. The SQK-RAD001 library used 200 ng of unsheared genomic DNA. Libraries for SQK-MAP006 and SQK-MAP007 were constructed following manufacturer’s instructions with the exception of using 0.4x concentration of AMPure XP Beads (Beckman Coulter Inc., Brea, CA) and 80 % EtOH during the “End Repair/dA-tailing module” step. The SQK-RAD001 library was constructed following manufacturer’s instructions. Prior to sequencing, flow cell quality was assessed by running the MinKNOW platform QC (Oxford Nanopore Technology). All flow cells were primed with priming buffer and the libraries were loaded following manufacturer’s instructions. The mixture was then loaded into the flow cells for sequencing. The SQK-MAP006 library of CEN.PK113-7D Delft was sequenced twice on a R7.3 chemistry flow cell (FLO-MIN103) and the SQK-RAD001 library was sequenced on a R9 chemistry flow cell (FLO-MIN105)—all for 48 hours. The SQK-MAP007 library for CEN.PK113-7D Frankfurt was sequenced for 48 hours on a R9 chemistry flow cell (FLO-MIN104). Reads from all sequencing runs were uploaded and base-called using Metrichor desktop agent (<https://metrichor.com/s/>). The error rate of nanopore reads in the CEN.PK113-7D Frankfurt and Delft was determined by aligning them to the final CEN.PK113-7D assembly (see section below) using Graphmap (28) and calculating mismatches based on the CIGAR strings of reads with a mapping quality of at least 1 and no more than 500 nt of soft/hard clipping on each end of the alignment to avoid erroneous read-alignments due to repetitive regions (i.e. paralogous genes, genes with copy number variation). All nanopore sequencing data are available at NCBI under the bioproject accession number PRJNA393501.

### **De novo genome assembly**

FASTA and FASTQ files were extracted from base-called FAST5 files using Poretools (version 0.6.0) (29). Raw nanopore reads were filtered for lambda DNA by aligning to the *Enterobacteria phage lambda* reference genome (RefSeq assembly accession: GCF\_000840245.1) using Graphmap (28) with *--no-end2end* parameter and retaining only unmapped reads using Samtools (30). All reads obtained from the Delft and the Frankfurt CEN.PK113-7D stock cultures were assembled *de novo* using Canu (version 1.3) (31) with *--genomesize* set to 12 Mbp. The assemblies were aligned using the MUMmer tool package: Nucmer with the *--maxmatch* parameter and filtered for the best one-to-one alignment using Delta-filter (32). The genome assemblies were visualized using Mummerplot (32) with the *--fat* parameter. Gene annotations were performed using MAKER2 annotation pipeline (version 2.31.9) using SNAP (version 2013-11-29) and Augustus (version 3.2.3) as *ab initio* gene

predictors (33). 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+) (34). Translated protein sequence of the final gene model were aligned using BLASTP to S288C protein Swiss-Prot database. Custom made Perl scripts were used to map systematic names to the annotated gene names. Telomere cap sequences (TEL07R of size 7,306 bp and TEL07L of size 781 bp) from the manually-curated and complete reference genome for *S. cerevisiae* S288C (version R64, Genbank ID: 285798) obtained from SGD were aligned to the assembly as a proxy to assess completeness of each assembled chromosome. SGIDs for TEL07R and TEL07L are S000028960 and S000028887, respectively. The Tablet genome browser (35) was used to visualize nanopore reads aligned to the nanopore *de novo* assemblies. Short assembly errors in the Frankfurt assembly were corrected with Nanopolish (version 0.5.0) using default parameters (36). Two contigs, corresponding to chromosome XII, were manually scaffolded based on homology to S288C. To obtain the 2-micron native plasmid in CEN.PK113-7D, we aligned S288C's native plasmid to the "unassembled" contigs file provided by Canu (31) and obtained the best aligned contig in terms of size and sequence similarity. Duplicated regions due to assembly difficulties in closing circular genomes were identified with Nucmer and manually corrected. BWA (37) was used to align Illumina reads to the scaffolded Frankfurt assembly using default parameters. Pilon (38) was then used to further correct assembly errors by aligning Illumina reads to the scaffolded Frankfurt assembly using correction of only SNPs and short indels (`--fix bases` parameter) using only reads with a minimum mapping quality of 20 (`--minmq 20` parameter). Polishing with structural variant correction in addition to SNP and short indel correction was benchmarked, but not applied to the final assembly (Additional File 1).

#### **Analysis of added information in the CEN.PK113-7D nanopore assembly**

Gained and lost sequence information in the nanopore assembly of CEN.PK113-7D was determined by comparing it to the previous short-read assembly (6). Contigs of at least 1 Kbp of short-read assembly were aligned to the nanopore CEN.PK113-7D Frankfurt assembly using the MUMmer tool package (32) using `show-coords` to extract alignment coordinates. For multi-mapped contigs, overlapping alignments of the same contig were collapsed and the largest alignment length as determined by Nucmer was used. Unaligned coordinates in the nanopore assembly were extracted and considered as added sequence. Added genes were retrieved by extracting the gene annotations in these unaligned regions from the annotated nanopore genome; mitochondria and 2-micron plasmid genes were excluded. For the lost sequence, unaligned sequences were obtained by aligning the contigs of the nanopore assembly to the short-read contigs of at least 1 kb using the same procedure as described above. Lost genes were retrieved by aligning the unaligned sequences to the short-read CEN.PK113-7D assembly with BLASTN (version 2.2.31+) (34) and retrieving gene annotations. BLASTN was used to align DNA sequences of YHRCTy1-1, YDRCTy2-1, YLWTy3-1, YHLWTy4-1, and YCLWTy5-1 (obtained from the *Saccharomyces Genome Database*; SGIDs: S000007006, S000006862, S000007020, S000006991, and S000006831, respectively) as proxies for the location of two known groups of Ty-elements in *Saccharomyces cerevisiae*, *Metaviridae* and *Pseudoviridae* (12), in the CEN.PK113-7D Frankfurt assembly. Non-redundant locations with at least a 2 Kbp alignment and an E-value of 0.0 as determined by BLASTN were then manually inspected.

#### **Comparison of the CEN.PK113-7D assembly to the S288C genome**

The nanopore assembly of CEN.PK113-7D and the reference genome of S2888C (Accession number GCA\_000146045.2) were annotated using the MAKER2 pipeline described in the "*De novo* genome assembly" section. For each genome a list of gene names per chromosome was constructed and

compared strictly on their names to identify genes names absent in the corresponding chromosome in the other genome. The ORFs of genes identified as absent in either genome were aligned using BLASTN (version 2.2.31+) to the total set of ORFs of the other genome and matches with an alignment length of half the query and with a sequence identity of at least 95 % were listed. If one of the unique genes aligned to an ORF on the same chromosome, it was manually inspected to check if it was truly absent in the other genome. Merged ORFs and misannotations were not considered in further analysis. These alignments were also used to identify copies and homologues of the genes identified as truly absent in the other genome.

Gene ontology analysis was performed using the Gene Ontology term finder of SGD using the list of unique genes as the query set and all annotated genes as the background set of genes for each genome (Additional file 2A and 2C). The ORFs of genes identified as present in S288C but absent in CEN.PK113-7D in previously made lists (6, 39) were obtained from SGD. The ORFs were aligned both ways to ORFs from SGD identified as unique to S288C in this study using BLASTN. Genes with alignments of at least half the query length and with a sequence identity of at least 95 % were interpreted as confirmed by the other data set. In order to analyze the origin of genes identified as unique to S288C, these ORFs were aligned using BLASTN to 481 genome assemblies of various *S. cerevisiae* strains obtained from NCBI (Additional file 3) and alignments of at least 50 % of the query were considered. The top alignments were selected based on the highest sequence ID and only one alignment per strain was counted per gene.

### **Chromosome translocation analysis**

Reads supporting the original and translocated genomic architectures of chromosomes III and VIII were identified via read alignment of raw nanopore reads. First, the translocation breakpoints coordinates were calculated based on whole-genome alignment of CEN.PK113-7D Delft assembly to S288C with MUMmer. A modified version of S288C was created containing the normal architectures of all 16 chromosomes and the mitochondrial genome plus the translocated architecture of chromosomes III-VIII and VIII-III. The first nearest unique flanking genes at each breakpoint were determined using BLASTN (version 2.2.31+) (40, 41) in reference to both S288C and the Delft CEN.PK113-7D nanopore assembly. Raw nanopore reads from CEN.PK113-7D Delft and Frankfurt were aligned to the modified version of S288C and nanopore reads that spanned the translocation breakpoints as well as the unique flanking sequences were extracted. Supporting reads were validated by re-aligning them to the modified version of S288C using BLASTN.

## **Results**

### **Sequencing on a single nanopore flow cell enables near-complete genome assembly**

To obtain a complete chromosome level *de novo* assembly of *Saccharomyces cerevisiae* CENPK113-7D, we performed long read sequencing on the Oxford Nanopore Technology's (ONT) MinION platform. A fresh sample of CEN.PK113-7D was obtained from the original distributor dr. P. Kötter (further referred to as "CEN.PK113-7D Frankfurt"), cultured in a single batch on YPD medium and genomic DNA was extracted. CEN.PK113-7D Frankfurt was sequenced on a single R9 (FLO-MIN104) chemistry flow cell using the 2D ligation kit for the DNA libraries producing more than 49x coverage of the genome with an average read-length distribution of 10.0 Kbp (Supplementary Figure S1) and an estimated error rate of 10 % (Supplementary Figure S2). We used Canu (31) to produce high-quality *de novo* assemblies using only nanopore data. Before correcting for misassemblies, the assembly contained a total of 21 contigs with an N50 of 756 Kbp

(Supplementary Table S1). This represented a 19-fold reduction in the number of contigs and a 15-fold increase of the N50 in comparison to the short-read-only assembly of the first CEN.PK113-7D draft genome version (6) (Table 1).

**Table 1. Comparison of 454/Illumina and nanopore *de novo* assemblies of CEN.PK113-7D.** Summary of *de novo* assembly metrics of CEN.PK113-7D Delft and CEN.PK113-7D Frankfurt. For the short-read assembly, only contigs of at least 1 Kbp are shown (6). The nanopore assembly of CEN.PK113-7D Delft is uncorrected for misassemblies while CEN.PK113-7D Frankfurt was corrected for misassemblies.

Data	CEN.PK113-7D Delft		CEN.PK113-7D Frankfurt
	Short read	Nanopore	Nanopore
<b>Contigs (<math>\geq</math> 1 Kbp)</b>	414	24	20
<b>Largest contig</b>	0.210 Mbp	1.08 Mbp	1.50 Mbp
<b>Smallest contig</b>	0.001 Mbp	0.013 Mbp	0.085 Mbp
<b>N50</b>	0.048 Mbp	0.736 Mbp	0.912 Mbp
<b>Total assembly size</b>	11.4 Mbp	11.9 Mbp	12.1 Mbp

Most chromosomes of the nanopore *de novo* assembly are single contigs and are flanked by telomere caps. Genome completeness was determined by alignment to the manually-curated reference genome of the strain S288C (version R64, Genbank ID: 285798) (Supplementary Table S2). The two largest yeast chromosomes, IV and XII, were each split in two separate contigs, and two additional contigs (31 and 38 Kbp in length) corresponded to unplaced subtelomeric fragments. In particular, the assembly for chromosome XII was interrupted in the *RDN1* locus—a repetitive region consisting of gene encoding ribosomal RNA estimated to be more than 1-Mbp long (42). Since no reads were long enough to span this region, the contigs were joined with a gap.

Manual curation resolved chromosome III, chromosome IV and the mitochondrial genome. Chromosome IV was fragmented into two contigs at locus of 11.5 Kbp containing two Ty-elements in S288C (coordinates 981171-992642). Interestingly, the end of the first contig and the start of the second contig have 8.8 Kbp of overlap (corresponding to the two Ty-elements) and one read spans the repetitive Ty-elements and aligns to unique genes on the left and right flanks (*EXG2* and *DIN7*, respectively). We therefore joined the contigs without missing sequence resulting in a complete assembly of chromosome IV. For chromosome III, the last ~27 Kbp contained multiple telomeric caps next to each other. The last ~10 Kbp had little to no coverage when re-aligning raw nanopore reads to the assembly (Supplementary Figure S3). The coordinates for the first telomeric cap were identified and the remaining sequence downstream was removed resulting in a final contig of size of 347 Kbp. The original contig corresponding to the mitochondrial genome had a size of 104 Kbp and contained a nearly identical ~20 Kbp overlap corresponding to start of the *S. cerevisiae* mitochondrial genome (i.e. origin of replication) (Supplementary Figure S4). This is a common artifact as assembly algorithms generally have difficulties reconstructing and closing circular genomes (22, 42). The coordinates of the overlaps were determined with Nucmer (32) and manually joined resulting to a final size of 86,616 bp.

Overall, the final CEN.PK113-7D Frankfurt assembly contained 15 chromosome contigs, 1 chromosome scaffold, the complete mitochondrial contig, the complete 2-micron plasmid and two unplaced telomeric fragments, adding up to a total of 12.1 Mbp (Table 1 and Supplementary Table S3). Of the 16 chromosomes, 11 were assembled up until both telomeric caps, four were



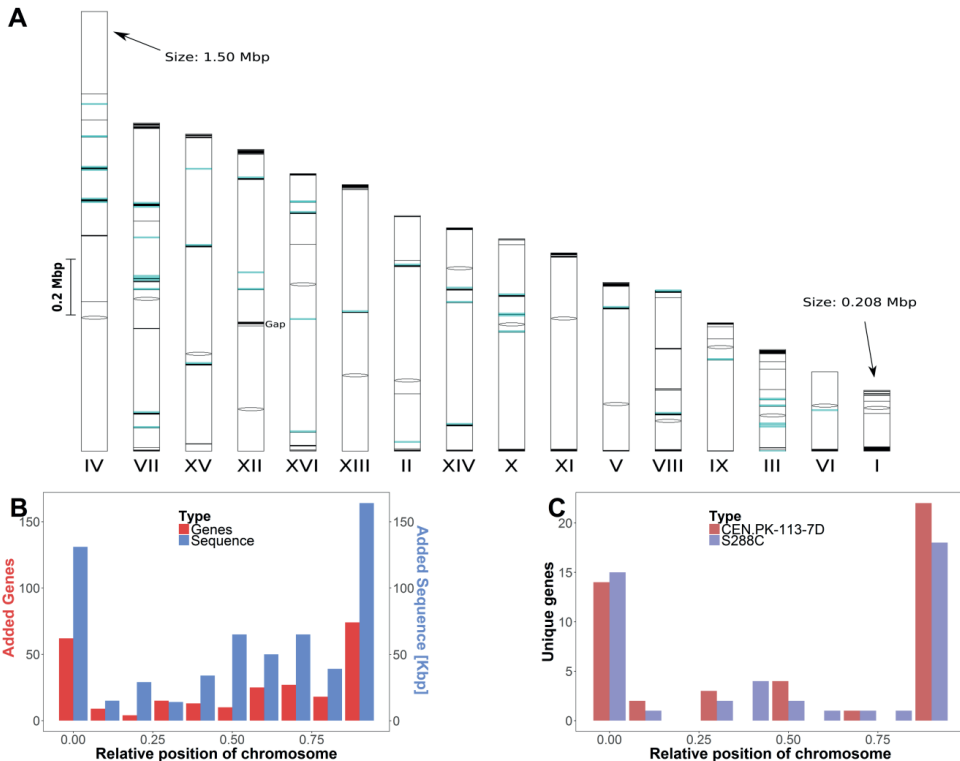
missing one of the telomere caps and only chromosome X was missing both telomere caps. Based on homology with S288C, the missing sequence was estimated not to exceed 12 kbp for each missing (sub)telomeric region. Furthermore, we found a total of 46 retrotransposons Ty-elements: 44 were from the *Pseudoviridae* group (30 *Ty1*, 12 *Ty2*, 1 *Ty4*, and 1 *Ty5*) and 2 from *Metaviridae* group (*Ty3*). The annotated nanopore assembly of CEN.PK113-7D Frankfurt is available at NCBI under the bioproject accession number PRJNA393501.

### **Comparison of the nanopore and short-read assemblies of CEN.PK113-7D**

We compared the nanopore assembly of CEN.PK113-7D to a previously published version to quantify the improvements over the current state-of-the art (6). Alignment of the contigs of the short-read assembly to the nanopore assembly revealed 770 Kbp of previously unassembled sequence, including the previously unassembled mitochondrial genome (Additional file 4A). This gained sequence is relatively spread out over the genome (Figures 1A and 1B) and contained as much as 284 chromosomal gene annotations (Additional file 4B). Interestingly, 69 out of 284 genes had paralogs, corresponding to a fraction almost twice as high as the 13 % found in the whole genome of S288C (43). Gene ontology analysis revealed an enrichment in the biological process of cell aggregation ( $9.30 \times 10^{-4}$ ); in the molecular functions of mannose binding ( $P=3.90 \times 10^{-4}$ ) and glucosidase activity ( $P=7.49 \times 10^{-3}$ ); and in the cellular components of the cell wall ( $P=3.41 \times 10^{-7}$ ) and the cell periphery component ( $P=5.81 \times 10^{-5}$ ). Some newly-assembled genes are involved in central carbon metabolism, such as *PDC5*. In addition, many of the added genes are known to be relevant in industrial applications including hexose transporters such as *HXT* genes and sugar polymer hydrolases such as *IMA* and *MALx2* genes; several genes relevant for cellular metal homeostasis, such as *CUP1-2* (linked to copper ion tolerance) and *FIT1* (linked to iron ion retention); genes relevant for nitrogen metabolism in medium rich or poor in specific amino acids, including amino acid transporters such as *VBA5*, amino acid catabolism genes such as *ASP3-4* and *LEU2* and amino-acid limitation response genes such as many *PAU* genes; several *FLO* genes which are responsible for calcium-dependent flocculation; and various genes linked to different environmental stress responses, such as *HSP* genes increasing heat shock tolerance and *RIM101* increasing tolerance to high pH.

To evaluate whether some previously assembled sequence was missing in the nanopore assembly, we aligned the nanopore contigs to the short-read assembly (6). Less than 6 Kbp of sequence of the short-read assembly was not present in the nanopore assembly, distributed over 13 contigs (Additional file 4C). Only two ORFs were missing: the genes *BIO1* and *BIO6* (Additional file 4D). Alignment of *BIO1* and *BIO6* sequences to the nanopore assembly showed that the right-end of the chromosome I contig contains the first ~500 nt of *BIO1*. While *BIO1* and *BIO6* were present in the nanopore sequences, they are absent in the final assembly likely due to the lack of long-enough reads to resolve the repetitive nature of this subtelomeric region.

Overall an additional 770 Kbp sequence containing 284 genes was gained, while 6 Kbp containing two genes was not captured compared to the previous assembly. In addition, the reduction from over 700 to only 20 contigs clearly shows that the nanopore assembly is much less fragmented than the short-read assembly (Table 1).



**Figure 1: Overview of gained and lost sequence and genes in the CEN.PK113-7D Frankfurt nanopore assembly relative to the short-read CEN.PK113-7D assembly and to the genome of S288C.** The two unplaced subtelomeric contigs and the mitochondrial DNA were not included in this figure. (1A) Chromosomal location of sequence assembled in the nanopore assembly which was not assembled using short-read data. The sixteen chromosome contigs of the nanopore assembly are shown. Chromosome XII has a gap at the *RDN1* locus, a region estimated to contain more than 1 Mbp worth of repetitive sequence (42). Centromeres are indicated by black ovals, gained sequence relative to the short-read assembly is indicated by black marks and 46 identified retrotransposon Ty-elements are indicated by blue marks. The size of all chromosomes and marks is proportional to their corresponding sequence size. In total 611 Kbp of sequence was added within the chromosomal contigs. (1B) Relative chromosome position of sequences and genes assembled on chromosome contigs of the nanopore assembly which were not assembled using short-read data. The positions of added sequence and genes were normalized to the total chromosome size. The number of genes (red) and the amount of sequence (cyan) over all chromosomes are shown per tenth of the relative chromosome size. (1C) Relative chromosome position of gene presence differences between S288C and CEN.PK113-7D. The positions of the 45 genes identified as unique to CEN.PK113-7D and of the 44 genes identified as unique to S288C were normalized to the total chromosome size. The number of genes unique to CEN.PK113-7D (red) and to S288C (purple) are shown per tenth of the relative chromosome position.

### Comparison of the Nanopore assembly of CEN.PK113-7D to S288C

To identify unique and shared genes between CEN.PK113-7D and S288C, we compared annotations made using the same method for both genomes (Additional Files 2A and 2C). We identified a total of 45 genes unique to CEN.PK113-7D and 44 genes unique to S288C (Additional Files 2B and 2D). Genes located in regions that had no assembled counterpart in the other genome were excluded; 20 for S288C and 27 for CEN.PK113-7D. Interestingly, the genes unique to either strain and genes present on different chromosomes were found mostly in the outer 10 % of the chromosomes, indicating that

the subtelomeric regions harbor most of the genetic differences between CEN.PK113-7D and S288C (Figure 1C).

In order to validate the genes identified as unique to S288C, we compared them to genes identified as absent in CEN.PK113-7D in previous studies (Additional file 2D, Table 2). 25 genes of S288C were identified as absent in CEN.PK113-7D by array comparative genomic hybridization (aCGH) analysis (39) and 21 genes were identified as absent in CEN.PK113-7D based on short-read WGS (6). Of these genes, 19 and 10 respectively were identified as genes in S288C by our annotation pipeline and could be compared to the genes we identified as unique to S288C. While 19 of these 29 genes were also absent in the nanopore assembly, the remaining 10 genes were fully assembled and annotated, indicating they were erroneously identified as missing (Table 2).

**Table 2: Presence in the nanopore assembly of genes identified as absent in CEN.PK113-7D in previous research.** For genes identified as absent in CEN.PK113-7D in two previous studies, the absence or presence in the nanopore assembly of CEN.PK113-7D is shown. 25 genes were identified previously by array comparative genome hybridisation (39) and 21 genes were identified by short-read genome assembly (6). Genes which were not annotated by MAKER2 in S288C could not be analysed. Genes with an alignment to genes identified as missing in the nanopore assembly of at least 50 % of the query length and 95 % sequence identity were confirmed as being absent, while those without such an alignment were identified as present. The presence of these genes was verified manually, which revealed the misannotation of YPL277C as YOR389W.

	Not analysed	Absent in assembly	Present in assembly
Daran-Lapujade <i>et al</i>	YAL064C-A, YAL066W, YAR047C, YHL046W-A, YIL058W, YOL013W-A	YAL065C, YAL067C, YBR093C, YCR018C, YCR105W, YCR106W, YDR038C, YDR039C, YHL047C, YHLO48W, YNR070W, YNR071C and YNR074C	YAL069W, YDR036C, YDR037W, YJL165C, YNR004W, and YPL277C (misannotated as YOR389W)
Nijkamp <i>et al</i>	Q0140, YDR543C, YDR544C, YDR545W, YIL046W-A, YLR154C-H, YLR156C-A, YLR157C-C, YLR159C-A, YOR029W and YOR082C	YBR093C, YCR040W, YCR041W, YDR038C, YDR039C and YDR040C	YDR036C, YHL008C, YHR056C and YLR055C

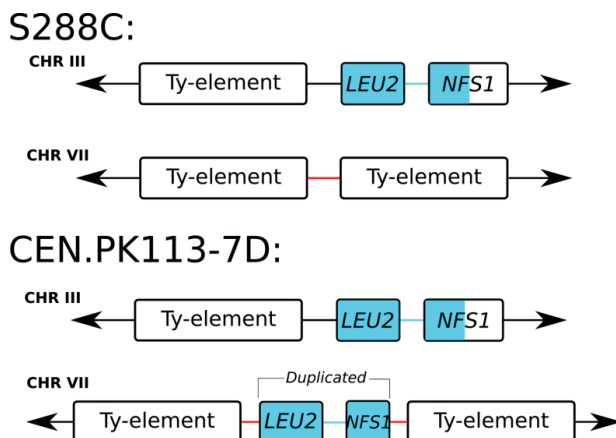
In order to determine if the genes unique to S288C have homologues elsewhere in the genome of CEN.PK113-7D or if they are truly unique, we aligned the ORFs of the 44 genes identified as unique in S288C to the ORFs in the nanopore CEN.PK113-7D assembly. 26 genes were completely absent in the CEN.PK113-7D assembly, while the remaining 18 genes aligned to between 1 and 20 ORFs each in the genome of CEN.PK113-7D with more than 95 % sequence identity, indicating they may have close homologues or additional copies in S288C (Additional file 2D). Gene ontology analysis revealed no enrichment in biological process, molecular functions or cell components of the 26 genes without homologues in CEN.PK113-7D. Five genes without homologues were labelled as putative. However, there were many genes encoding proteins relevant for fitness under specific industrial conditions, such as *PHO5* which is part of the response to phosphate scarcity, *COS3* linked to salt tolerance, *ADH7* linked to acetaldehyde tolerance, *RDS1* linked to resistance to cycloheximide, *PDR18* linked to ethanol tolerance and *HXT17* which is involved in hexitol uptake (Additional file 2D). In addition, we confirmed the complete absence of *ENA2* and *ENA5* in CEN.PK113-7D which are responsible for lithium sensitivity of CEN.PK113-7D (44).

Conversely, to determine if the genes unique to CEN.PK113-7D have homologues elsewhere in the genome of S288C or if they are truly unique, we aligned the ORFs of the 45 genes identified as unique in CEN.PK113-7D to the ORFs of S288C. A set of 16 genes were completely absent in S288C,

while the remaining 29 aligned to between one and 16 ORFs each in the genome of S288C with more than 95 % sequence (Additional File 2D). Gene ontology analysis revealed no enrichment in biological processes, molecular functions or cell components of the 16 genes unique to CEN.PK113-7D without homologues. However, among the genes without homologues a total of 13 were labelled as putative. The presence of an additional copy of *IMA1*, *MAL31* and *MAL32* on chromosome III was in line with the presence of the *MAL2* locus which was absent in S288C. Interestingly the sequence of *MAL13*, which belongs to this locus, was divergent enough from other *MAL*-gene activators to not be identified as homologue. Additionally, when performing the same analysis on the 27 genes on the two unplaced contigs of the CEN.PK113-7D assembly, 7 of them did not align to any gene of S288C with more than 95 % sequence identity, indicating these unplaced telomeric regions are highly unique to CEN.PK113-7D.

Since the genome of CEN.PK113-7D contains 45 ORFs which are absent in S288C, we investigated their origin by aligning them against all available *S. cerevisiae* nucleotide data at NCBI (Additional File 3). For each ORF, we report the strains to which they align with the highest sequence identity and the sequence identity relative to S288C in Additional File 2B. For most genes, several strains aligned equally well with the same sequence identity. For 13 ORFs S288C is among the best matches, indicating these ORFs may come from duplications in the S288C genome. However, S288C is not among the best matches for 32 ORFs. In these, laboratory strain “SK1” is among the best matches 9 times, the west African wine isolate “DBVPG6044” appears 8 times, laboratory strain “W303” appears 7 times, the Belgian beer strain “beer080” appears 3 times and the Brazilian bioethanol strain “bioethanol005” appears 3 times. Interestingly, some grouped unique genes are most related to specific strains. For example, the unique genes identified on the left subtelomeric regions of chromosome XVI (YBL109W, YHR216W and YOR392) and of chromosome VIII (YJL225C and YOL161W) exhibited the highest similarity to DBVPG6044. Similarly, the right end of the subtelomeric region of chromosome III (YPL283W-A and YPR202) and of chromosome XI ((YPL283W-A and YLR466W) were most closely related to W303.

Interestingly, the nanopore assembly revealed a duplication of *LEU2*, a gene involved in synthesis of leucine which can be used as an auxotrophy marker. In the complete reference genome of *S. cerevisiae* S288C, both *LEU2* and *NFS1* are unique, neighboring genes located chromosome III. However, gene annotations of the assemblies and raw nanopore reads support additional copies of *LEU2* and *NFS1* in CEN.PK113-7D located on chromosome VII (Figure 2). The additional copy contained the complete *LEU2* sequence but only ~0.5 kb of the 5' end of *NFS1*. In CEN.PK113-7D and S288C, the *LEU2* and *NFS1* loci in chromosome III were located adjacent to Ty-elements. Two such Ty-elements were also found flanking the additional *LEU2* and *NFS1* loci in chromosome VII (Figure 2). It is likely that the duplication was the result of a translocation based on homology of the Ty-elements which resulted in local copy number increase during its strain development program (27).



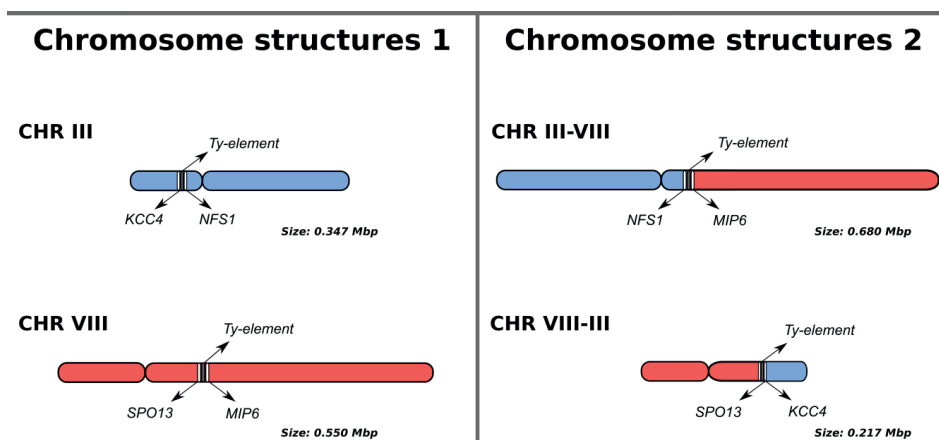
**Figure 2: *LEU2* and *NFS1* duplication in chromosome VII of CEN.PK113-7D.** The nanopore assembly contains a duplication of *LEU2* and part of *NFS1* in CEN.PK113-7D. In S288C, the two genes are located in chromosome III next to a Ty element. In CEN.PK113-7D, the two genes are present in chromosome III and in chromosome VII. The duplication appears to be mediated by Ty-elements. Note that the additional copy in chromosome VII is present in between two Ty-elements and contains only the first ~500 bp of *NFS1*. The duplication is supported by long-read data that span across the *LEU2*, *NFS1*, the two Ty-elements, and the neighboring flanking genes (not shown).

#### Long-read sequencing data reveals chromosome structure heterogeneity in CEN.PK113-7D Delft

CEN.PK113-7D has three confirmed *MAL* loci encoding genes for the uptake and hydrolysis of maltose: *MAL1* on chromosome VII, *MAL2* on chromosome III and *MAL3* on chromosome II (Additional file 2A). A fourth *MAL* locus was identified in previous research on chromosome XI based on contour-clamped homogeneous electric field electrophoresis (CHEF) and southern blotting with a probe for *MAL* loci (6). However, the nanopore assembly revealed no additional *MAL* locus despite the complete assembly of Chromosome XI. The CEN.PK113-7D stock in which the fourth *MAL* locus was obtained from Dr P. Kötter in 2001 and stored at  $-80^{\circ}\text{C}$  since (further referred to as “CEN.PK113-7D Delft”). In order to investigate the presence of the potential *MAL* locus, we sequenced CEN.PK113-7D Delft using nanopore MinION sequencing. Two R7.3 flow cells (FLO-MIN103) produced 55x coverage with an average read-length distribution of 8.5 Kbp and an R9 flow cell (FLO-MIN103) produced 47x coverage with an average read-length distribution of 3.2 Kbp (Supplementary Figure S1). The error rate was estimated to be 13 % (Supplementary Figure S4) after aligning the raw nanopore reads to the CEN.PK113-7D Frankfurt assembly. These reads were assembled into 24 contigs with an N50 of 736 Kbp (Supplementary Table S1).

Alignment of the assembly of CEN.PK113-7D Delft to the Frankfurt assembly showed evidence of a translocation between chromosomes III and VIII (Supplementary Figure S5). The assembly thus suggested the presence of two new chromosomes: chromosomes III-VIII of size 680 Kbp and chromosome VIII-III of size 217 Kbp (Figure 3). The translocation occurred between Ty-element YCLWTy2-1 on chromosome III and long terminal repeats YHRCdelta5-7 on chromosome VIII. These repetitive regions are flanked by unique genes *KCC4* and *NFS1* on chromosome III and *SPO13* and *MIP6* on chromosome VIII (Figure 3). Nanopore reads spanning the whole translocated or non-translocated sequence anchored in the unique genes flanking them were extracted for CEN.PK113-7D Delft and Frankfurt. A total of eight reads from CEN.PK113-7D Delft supported the translocated chromosome III-VIII architecture (largest read was 39 Kbp) and one 19 Kbp read

supported the normal chromosome III architecture. For CEN.PK113-7D Frankfurt, we found only one read of size 23 Kbp that supported the normal chromosome III architecture but we found no reads that supported the translocated architectures. This data suggested that CEN.PK113-7D Delft is in fact a heterogeneous population containing cells with recombined chromosomes III and VIII and cells with original chromosomes III and VIII. As a result, in addition to the *MAL2* locus on chromosome III, CEN.PK113-7D Delft harboured a *MAL2* locus on recombined chromosome III-VIII. As the size of recombined chromosome III-VIII was close to chromosome XI, the *MAL2* locus on chromosome III-VIII led to misidentification of a *MAL4* locus on chromosome XI (6). By repeating the CHEF gel and southern blotting for *MAL* loci on several CEN.PK113-7D stocks, the *MAL2* on the translocated chromosomes III-VIII was shown to be present only in CEN.PK113-7D Delft, demonstrating that there was indeed chromosome structure heterogeneity (Additional File 5).



**Figure 3: Overview of chromosome structure heterogeneity in CEN.PK113-7D Delft for CHRIII and CHRVIII which led to the misidentification of a fourth *MAL* locus in a previous short-read assembly study of the genome of CEN.PK113-7D.** Nanopore reads support the presence of two chromosome architectures: the normal chromosomes III and VIII (left panel) and translocated chromosomes III-VIII and VIII-III (right panel). The translocation occurred in Ty-elements, large repetitive sequences known to mediate chromosomal translocations in *Saccharomyces* species (50). Long-reads are required to diagnose the chromosome architecture via sequencing: the repetitive region between *KCC4* to *NFS1* in chromosome III exceeds 15 Kbp, while the region between *SPO13* and *MIP6* in chromosome VIII is only 1.4 Kbp long. For the translocated architecture, the region from *NFS1* to *MIP6* in chromosome III-VIII exceeds 16 Kbp and the distance from *SPO13* to *KCC4* in chromosome VIII-III is nearly 10 Kbp.

## Discussion

In this study, we obtained a near-complete genome assembly of *S. cerevisiae* strain CEN.PK113-7D using only a single R9 flow cell on ONT's MinION sequencing platform. 15 of the 16 chromosomes as well as the mitochondrial genome and the 2-micron plasmid were assembled in single, mostly telomere-to-telomere, contigs. This genome assembly is remarkably unfragmented, even when compared with other *S. cerevisiae* assemblies made with several nanopore technology flow cells, in which 18 to 105 chromosomal contigs were obtained (22, 25). Despite the long read lengths obtained by Nanopore sequencing, the ribosomal DNA locus in chromosome XII could not be completely resolved. In practice, this would require reads exceeding 1 Mb in length, which current technology cannot yet deliver.

The obtained nanopore assembly is of vastly superior quality to the previous short-read-only assembly of CEN.PK113-7D that was fragmented into over 700 contigs (6). In addition to the lesser fragmentation, the addition of 770 Kbp of previously unassembled sequence led to the identification and accurate placement of 284 additional ORFs spread out over the genome. These newly assembled genes showed overrepresentation for cell wall and cell periphery compartmentalization and relate to functions such as sugar utilization, amino acid uptake, metal ion metabolism, flocculation and tolerance to various stresses. While many of these genes are already present in the short-read assembly of CEN.PK113-7D, copy number was shown to be an important factor determining the adaptation of strains to specific growth conditions (21). The added genes may therefore be very relevant for the specific physiology of CEN.PK113-7D under different industrial conditions (21). The ability of nanopore sequencing to distinguish genes with various similar copies is crucial in *S. cerevisiae* as homologues are frequent particularly in subtelomeric regions, and paralogues are widespread due to a whole genome duplication in its evolutionary history (43). Besides the added sequence, 6 Kbp of sequence of the short-read assembly was not present in the nanopore assembly, mostly consisting of small unplaced contigs. Notably the absence of *BIO1* and *BIO6* in the assembly was unexpected, as it constituted a marked difference between CEN.PK113-7D and many other strains which enables biotin prototrophy (45). Both genes were present in the nanopore reads, but were unassembled likely due to the lack of reads long-enough to resolve this subtelomeric region (a fragment of *BIO1* is located at the right-end of chromosome I). Targeted long-read sequencing in known gaps of a draft assembly followed by manual curation could provide an interesting tool to obtain complete genome assemblies (46). Alternatively, a more complete assembly could be obtained by maximizing read length. The importance of read length is illustrated by the higher fragmentation of the CEN.PK113-7D Delft assembly compared to the Frankfurt one, which was based on reads with lower length distribution despite higher coverage and similar error rate (Table 1, Supplementary Figures S1 and S5). Read-length distribution in nanopore sequencing is highly influenced by the DNA extraction method and library preparation (Supplementary Figure S1). The mitochondrial genome was completely assembled, which is not always possible with nanopore sequencing (22, 23, 25). Even with identical DNA extraction and assembly methods, the mitochondrial genome cannot always be assembled, as illustrated by its absence in the assembly of CEN.PK113-7D Delft. Overall, the gained sequence in the nanopore assembly far outweighs the lost sequence relative to the previous assembly, and the reduction in number of contigs presents an important advantage.

The use of long read sequencing enabled the discovery of a translocation between chromosomes III and VIII, which led to the misidentification of a fourth MAL locus on chromosome XI of CEN.PK113-7D (6). Identification of this translocation required reads to span at least 12 Kbp due to the large repetitive elements surrounding the translocation breakpoints, explaining why it was previously undetected. While the translocation did not disrupt any coding sequence and is unlikely to cause phenotypical changes (47), there may be decreased spore viability upon mating with other CEN.PK strains. Our ability to detect structural heterogeneity within a culture shows that nanopore sequencing could also be valuable in detecting structural variation within a genome between different chromosome copies, which occurs frequently in aneuploid yeast genomes (48). These results highlight the importance of minimal propagation of laboratory microorganisms to warrant genome stability and avoid heterogeneity which could at worst have an impact on phenotype and interpretation of experimental results.

The nanopore assembly of CEN.PK113-7D constitutes a vast improvement of its reference genome which should facilitate its use as a model organism. The elucidation of various homologue and paralogue genes is particularly relevant as CEN.PK113-7D is commonly used as a model for industrial *S. cerevisiae* applications for which gene copy number frequently plays an important role (21, 48). Using the nanopore assembly as a reference for short-read sequencing of strains derived from CEN.PK113-7D will yield more complete and more accurate lists of SNPs and other mutations, facilitating the identification of causal mutations in laboratory evolution or mutagenesis experiments. Therefore, the new assembly should accelerate elucidation of the genetic basis underlying the fitness of *S. cerevisiae* in various environmental conditions, as well as the discovery of new strain improvement strategies for industrial applications (49).

## Acknowledgements

The authors would like to thank dr. P. Kötter for sending CEN.PK113-7D Frankfurt, dr. Kirsten Benjamin for sending CEN.PK113-7D Amyris and dr. Verena Siewers for sending CEN.PK113-7D Chalmers. We are thankful to Prof. Jack T. Pronk (Delft University of Technology) and dr. Niels Kuijpers (HEINEKEN Supply Chain B.V.) for their critical reading of the manuscript.

This work was performed within the BE-Basic R&D Program (<http://www.be-basic.org/>), which was granted an FES subsidy from the Dutch Ministry of Economic Affairs, Agriculture and Innovation (EL&I). Anja Brickwedde was funded by the Seventh Framework Programme of the European Union in the frame of the SP3 people support for training and career development of researchers (Marie Curie), Networks for Initial Training (PITN-GA-2013 ITN-2013-606795) YeastCell.

## References

- Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet* 24(3):133-141.
- Ng PC & Kirkness EF (2010) Whole genome sequencing. *Genetic variation*, (Springer), pp 215-226.
- van Dijk EL, Auger H, Jaszczyszyn Y, & Thermes C (2014) Ten years of next-generation sequencing technology. *Trends Genet* 30(9):418-426.
- Matheson K, Parsons L, & Gammie A (2017) Whole-genome sequence and variant analysis of W303, a widely-used strain of *Saccharomyces cerevisiae*. *G3 (Bethesda)*:g3. 117.040022.
- Cherry JM, et al. (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40(Database issue):D700-705.
- Nijkamp JF, et al. (2012) *De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microb Cell Fact* 11(1):36.
- Canelas AB, et al. (2010) Integrated multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains. *Nat Commun* 1(9):145.
- González-Ramos D, et al. (2016) A new laboratory evolution approach to select for constitutive acetic acid tolerance in *Saccharomyces cerevisiae* and identification of causal mutations. *Biotechnol Biofuels* 9(1):173.
- Papapetridis I, Dijk M, Maris AJA, & Pronk JT (2017) Metabolic engineering strategies for optimizing acetate reduction, ethanol yield and osmotolerance in *Saccharomyces cerevisiae*. *Biotechnol Biofuels* 10(1):107.
- Nijkamp J, et al. (2010) Integrating genome assemblies with MAIA. *Bioinformatics* 26(18):i433-i439.
- Pryde FE, Huckle TC, & Louis EJ (1995) Sequence analysis of the right end of chromosome XV in *Saccharomyces cerevisiae*: an insight into the structural and functional significance of sub-telomeric repeat sequences. *Yeast* 11(4):371-382.
- Kim JM, Vanguri S, Boeke JD, Gabriel A, & Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 8(5):464-478.
- Bergström A, et al. (2014) A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* 31(4):872-888.
- Carlson M, Celenza JL, & Eng FJ (1985) Evolution of the dispersed *SUC* gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biol*. 5(11):2894-2902.
- Naumov GI, Naumova ES, & Louis EJ (1995) Genetic mapping of the  $\alpha$ -galactosidase *MEL* gene family on right and left telomeres of *Saccharomyces cerevisiae*. *Yeast* 11(5):481-483.
- Jordan P, Choe J-Y, Boles E, & Oreb M (2016) Hxt13, Hxt15, Hxt16 and Hxt17 from *Saccharomyces cerevisiae* represent a novel type of polyol transporters. *Sci Rep* 6.



17. Teste M-A, François JM, & Parrou J-L (2010) Characterization of a new multigene family encoding isomaltases in the yeast *Saccharomyces cerevisiae*, the IMA family. *J Biol Chem* 285(35):26815-26824.
18. Denayrolles M, de Villechenon EP, Lonvaud-Funel A, & Aigle M (1997) Incidence of *SUC-RTM* telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Curr Genet* 31(6):457-461.
19. Teunissen AW & Steensma HY (1995) Review: The dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* 11(11):1001-1013.
20. Lodolo EJ, Kock JLF, Axcell BC, & Brooks M (2008) The yeast *Saccharomyces cerevisiae* - the main character in beer brewing. *Fems Yeast Res* 8(7):1018-1036.
21. Brown CA, Murray AW, & Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol* 20(10):895-903.
22. McIlwain SJ, et al. (2016) Genome sequence and analysis of a stress-tolerant, wild-derived strain of *Saccharomyces cerevisiae* used in biofuels research. *G3 (Bethesda)* 6(6):1757-1766.
23. Giordano F, et al. (2017) *De novo* yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci Rep* 7(1):3935.
24. Goodwin S, et al. (2015) Oxford Nanopore sequencing, hybrid error correction, and *de novo* assembly of a eukaryotic genome. *Genome Res* 25(11):1750-1756.
25. Istace B, et al. (2017) *De novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6(2):1-13.
26. Jansen H, et al. (2017) *De novo* whole-genome assembly of a wild type yeast isolate using nanopore sequencing. *F1000Research* 6.
27. Entian K-D & Kötter P (2007) 25 Yeast genetic strain and plasmid collections. *Method Microbiol* 36:629-666.
28. Sović I, et al. (2016) Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat Commun* 7.
29. Loman NJ & Quinlan AR (2014) Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* 30(23):3399-3401.
30. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
31. Koren S, et al. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27(5):722-736.
32. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12.
33. Holt C & Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12(1):491.
34. Camacho C, et al. (2009) BLAST+: architecture and applications. *BMC bioinformatics* 10(1):421.
35. Milne I, et al. (2012) Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform*:bbs012.
36. Loman NJ, Quick J, & Simpson JT (2015) A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat Methods* 12(8):733-U751.
37. Li H & Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589-595.
38. Walker BJ, et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
39. Daran-Lapujade P, et al. (2003) Comparative genotyping of the *Saccharomyces cerevisiae* laboratory strains S288C and CEN. PK113-7D using oligonucleotide microarrays. *FEMS yeast res* 4(3):259-269.
40. Zhang Z, Schwartz S, Wagner L, & Miller W (2000) A greedy algorithm for aligning DNA sequences. *Journal of computational biology : a journal of computational molecular cell biology* 7(1-2):203-214.
41. English AC, et al. (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7(11):e47768.
42. Venema J & Tollervey D (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. *Annual review of genetics* 33(1):261-311.
43. Wolfe KH & Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387(6634):708.
44. Daran-Lapujade P, et al. (2009) An atypical *PMR2* locus is responsible for hypersensitivity to sodium and lithium cations in the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D. *FEMS yeast res* 9(5):789-792.
45. Bracher JM, et al. (2017) Laboratory evolution of a biotin-requiring *Saccharomyces cerevisiae* strain for full biotin prototrophy and identification of causal mutations. *Appl Environ Microbiol*:AEM. 00892-00817.
46. Loose M, Malla S, & Stout M (2016) Real-time selective sequencing using nanopore technology. *Nat Methods* 13(9):751.
47. Naseeb S, et al. (2016) Widespread Impact of Chromosomal Inversions on Gene Expression Uncovers Robustness via Phenotypic Buffering. *Mol Biol Eval* 33(7):1679-1696.
48. Gorter de Vries AR, Pronk JT, & Daran J-MG (2017) Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts. *Appl Environ Microbiol* 83(11):e03206-03216.
49. Oud B, Maris AJA, Daran JM, & Pronk JT (2012) Genome-wide analytical approaches for reverse metabolic engineering of industrially relevant phenotypes in yeast. *FEMS Yeast Res* 12(2):183-196.
50. Fischer G, James SA, Roberts IN, Oliver SG, & Louis EJ (2000) Chromosomal evolution in *Saccharomyces*. *Nature* 405(6785):451-454.

## Chapter 4: Nanopore sequencing and comparative genome analysis confirm lager-brewing yeasts originated from a single hybridization

Alex N. Salazar #, Arthur R. Gorter de Vries #, Marcel van den Broek, Nick Brouwers, Pilar de la Torre Cortés, Niels G. A. Kuijpers, Jean-Marc G. Daran and Thomas Abeel

# These authors contributed equally to this publication and should be considered co-first authors.

**Background:** The lager brewing yeast, *S. pastorianus*, is a hybrid between *S. cerevisiae* and *S. eubayanus* with extensive chromosome aneuploidy. *S. pastorianus* is subdivided into Group 1 and Group 2 strains, where Group 2 strains have higher copy number and a larger degree of heterozygosity for *S. cerevisiae* chromosomes. As a result, Group 2 strains were hypothesized to have emerged from a hybridization event distinct from Group 1 strains. Current genome assemblies of *S. pastorianus* strains are incomplete and highly fragmented, limiting our ability to investigate their evolutionary history.

**Results:** To fill this gap, we generated a chromosome-level genome assembly of the *S. pastorianus* strain CBS 1483 using MinION sequencing and analysed the newly assembled subtelomeric regions and chromosome heterozygosity. To analyse the evolutionary history of *S. pastorianus* strains, we developed Alpaca: a method to compute sequence similarity between genomes without assuming linear evolution. Alpaca revealed high similarities between the *S. cerevisiae* subgenomes of Group 1 and 2 strains, and marked differences from sequenced *S. cerevisiae* strains.

**Conclusions:** Our findings suggest that Group 1 and Group 2 strains originated from a single hybridization involving a heterozygous *S. cerevisiae* strain, followed by different evolutionary trajectories. The clear differences between both groups may originate from a severe population bottleneck caused by the isolation of the first pure cultures. Alpaca provides a computationally inexpensive method to analyse evolutionary relationships while considering non-linear evolution such as horizontal gene transfer and sexual reproduction, providing a complementary viewpoint beyond traditional phylogenetic approaches.

Essentially as published on bioRxiv

Supplementary materials available online at <http://tinyurl.com/y5ypd6yu>

<https://doi.org/10.1101/603480>

## Background

The lager-brewing yeast *Saccharomyces pastorianus* is an interspecies hybrid between *S. cerevisiae* and *S. eubayanus*. Lager brewing emerged in the late middle ages and was carried out during winter months at temperatures between 8 and 15 °C, followed by a prolonged maturation period referred to as lagering (1, 2). While *S. cerevisiae* is a well-studied species frequently used in biotechnological processes (3), *S. eubayanus* was only discovered in 2011 and has thus far only been isolated from the wild (4). Therefore, the ancestral *S. pastorianus* hybrid likely emerged from a spontaneous hybridization between an ale brewing *S. cerevisiae* yeast and a wild *S. eubayanus* contaminant, and took over lager brewing due to increased fitness under these conditions (4-6). Indeed, laboratory-made *S. cerevisiae* x *S. eubayanus* hybrids demonstrated hybrid vigour by combining the fermentative capacity and sugar utilisation of *S. cerevisiae* and the ability to grow at lower temperatures of *S. eubayanus* (7, 8).

The genomes of *S. pastorianus* strains are highly aneuploid, containing 0 to 5 copies of each chromosome (5, 9-13). Between 45 and 79 individual chromosomes were found in individual *S. pastorianus* genomes, compared to a normal complement of 32 chromosomes in euploid *Saccharomyces* hybrids. The degree of aneuploidy of *S. pastorianus* is exceptional in the *Saccharomyces* genera, and likely evolved during its domestication in the brewing environment (9). Nevertheless, two groups can be distinguished based on their genome organisation: Group 1 strains, which have approximately haploid *S. cerevisiae* and diploid *S. eubayanus* chromosome complements; and Group 2 strains, which have approximately diploid to tetraploid *S. cerevisiae* and diploid *S. eubayanus* chromosome complements (5, 10, 11, 14).

Group 1 and Group 2 strains in *S. pastorianus* were initially thought to have originated from two different hybridization events. Some lager-specific genes from Group 2 strains are absent in Group 1 strains, and the subtelomeric regions of Group 1 and Group 2 strains differ substantially (15, 16). Based on these differences, Group 1 and Group 2 strains were hypothesized to have emerged from different independent hybridization events, involving a haploid *S. cerevisiae* for Group 1 strains and a higher ploidy *S. cerevisiae* strain for Group 2 strains (5, 17). Indeed, crosses between *S. cerevisiae* and *S. eubayanus* strains with varying ploidies could be made in the laboratory, all of which performed well in the lager brewing process (18). Comparative genome analysis between Group 1 and Group 2 strains revealed that there were more synonymous nucleotide differences in the *S. cerevisiae* subgenome than in the *S. eubayanus* subgenome (19). As accumulation of synonymous mutations was presumed to equally affect both genomes, the authors hypothesized that Group 1 and 2 strains originated from two hybridizations, with a similar *S. eubayanus* parent and different *S. cerevisiae* parents.

More recent studies now support that Group 1 and Group 2 strains originated from the same hybridization event. Identical recombinations between the *S. cerevisiae* and *S. eubayanus* subgenomes were found at the *ZUO1*, *MAT*, *HSP82* and *XRN1/KEM1* loci in all analysed *S. pastorianus* strains (11, 13, 14), which did not emerge when such hybrids were evolved under laboratory conditions (20). These conserved recombinations indicate that all *S. pastorianus* strains share a common *S. cerevisiae* x *S. eubayanus* hybrid ancestor, and that the differences between Group 1 and Group 2 strains emerged subsequently. Sequence analysis of ten *S. pastorianus* genomes revealed that the *S. cerevisiae* sub-genome in Group 1 strains is relatively homozygous, while Group 2 strains possess heterozygous sub-regions (11). Moreover, heterozygous nucleotide stretches in Group 2

strains were composed of sequences highly similar to Group 1 genomes and of sequences from a different *S. cerevisiae* genome with a 0.5 % lower sequence identity. As a result, the authors formulated two hypotheses to explain the emergence of Group 1 and Group 2 strains from a shared ancestral hybrid: (i) the ancestral hybrid had a heterozygous *S. cerevisiae* sub-genome, and Group 1 strains underwent a massive reduction of the *S. cerevisiae* genome content while Group 2 did not, or (ii) the ancestral hybrid had a homozygous Group 1-like genome and Group 2 strains were formed by a subsequent hybridization event of such a Group 1-like strain with another *S. cerevisiae* strain, resulting in a mixed *S. cerevisiae* genome content in Group 2 strains.

Since the exact *S. cerevisiae* and *S. eubayanus* ancestors of *S. pastorianus* are not available, the evolutionary history of *S. pastorianus* has so far been based on the sequence analysis using available *S. cerevisiae* and *S. eubayanus* reference genomes (5, 11). However, these reference genomes are not necessarily representative of the original parental genomes of *S. pastorianus*. Although *S. pastorianus* genomes are available, they were sequenced with short-read sequencing technology (10-13) preventing assembly of large repetitive stretches of several thousand base pairs, such as TY-elements or paralogous genes often found in *Saccharomyces* genomes (21). The resulting *S. pastorianus* genomes assemblies are thus incomplete and fragmented into several hundred or thousand contigs (10-13).

Single-molecule sequencing technologies can generate reads of several thousand base pairs and span entire repetitive regions, enabling near complete chromosome-level genome assemblies of *Saccharomyces* yeasts (22-27). In addition to the lesser fragmentation, the assembly of regions containing repetitive sequences reveals large numbers of previously unassembled open reading frames, particularly in the sub-telomeric regions of chromosomes (24, 25, 27). Sub-telomeric regions are relatively unstable (28), and therefore contain much of the genetic diversity between different strains (29, 30). In *S. pastorianus*, notable differences were found between the sub-telomeric regions of Group 1 and Group 2 strains (15, 16), which could be used to understand their origin. Moreover, repetitive regions are enriched for genes with functions determining the cell's interaction with its environment, such as nutrient uptake, sugar utilization, inhibitor tolerance and flocculation (31-34). As a result, the completeness of sub-telomeric regions is critical for understanding genetic variation and evolutionary relationships between strains, as well as for understanding their performance in industrial applications (24, 29, 30).

Here, we used nanopore sequencing to obtain a chromosome-level assembly of the Group 2 *S. pastorianus* strain CBS 1483 and analysed the importance of new-found sequences relative to previous genome assemblies, with particular focus on industrially-relevant subtelomeric gene families. As the CBS 1483 genome contains multiple non-identical copies for many chromosomes, we analysed structural and sequence-level heterozygosity using short- and long-read data. Moreover, we developed a method to investigate the evolutionary origin of *S. pastorianus* by evaluating the genome similarity of several Group 1 and Group 2 *S. pastorianus* strains relative to a large dataset of *S. cerevisiae* and *S. eubayanus* genomes, including an isolate of the Heineken A-yeast lineage which was isolated by dr. Elion in 1886 and is still used in beer production today.

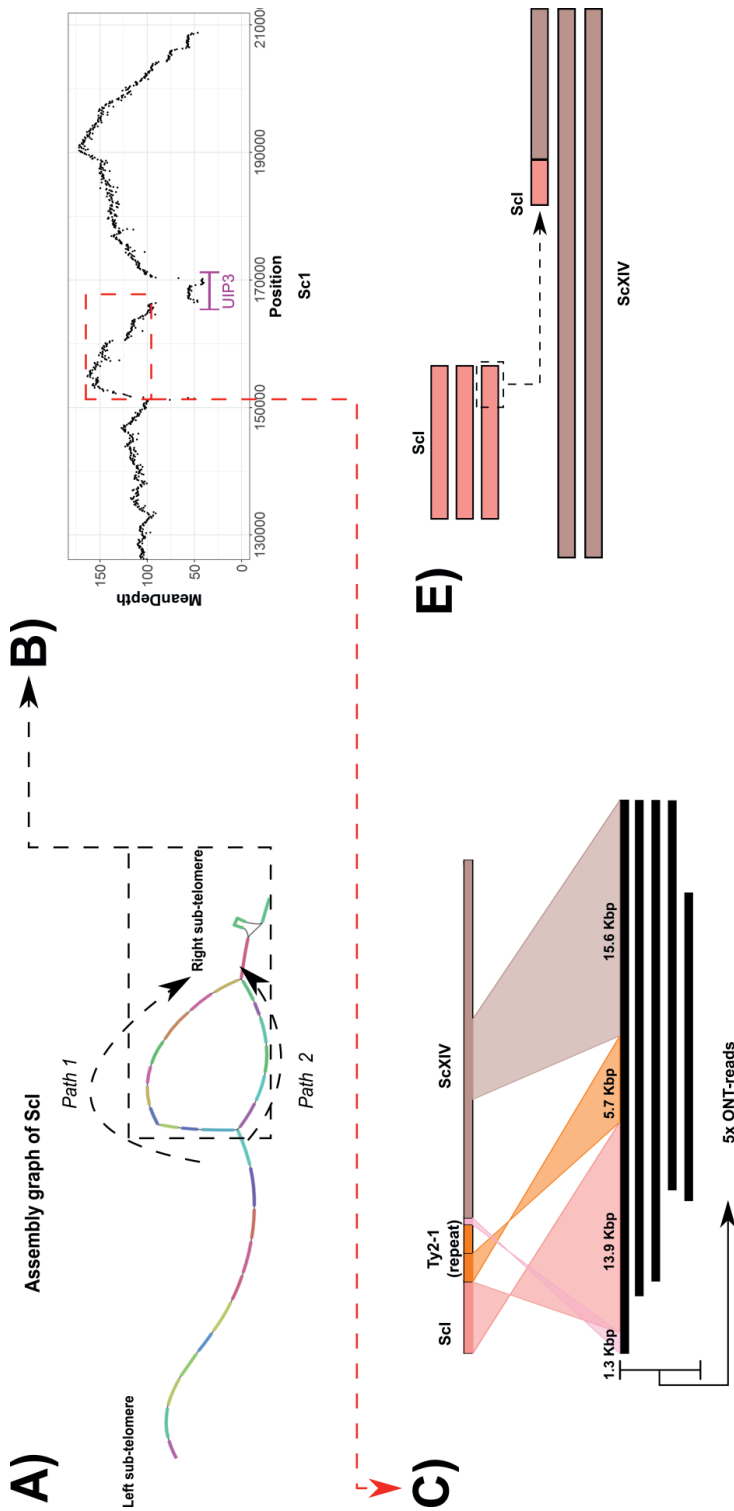
## Results

### Near-complete haploid assembly of CBS 1483

We obtained 3.3 Gbp of whole genome sequencing data of the *Saccharomyces pastorianus* strain CBS 1483 using 4 flow cells on Oxford Nanopore Technology's MinION platform. Based on a genome size of 46 Mbp accounting for all chromosome copy numbers, the combined coverage was 72x with an average read length of 7 Kbp (Figure S1). We assembled the reads using Canu and performed manual curation involving circularization of the mitochondrial DNA, scaffolding of ScXII (chromosome XII of the *S. cerevisiae* sub-genome) and resolution of assembly problems due to inter- and intra-chromosomal structural heterozygosity in ScI and ScXIV (Figure 1). Assembly errors were corrected with Pilon (35) using paired-end Illumina reads with 159x coverage. We obtained a final assembly of 29 chromosome contigs, 2 chromosome scaffolds, and the complete mitochondrial contig leading to a total size of 23.0 Mbp (Figure 2 and Table 1). The assembly was remarkably complete: of the 31 chromosomes (in CBS 1483 ScIII and SeIII recombined into a chimeric SeIII-ScIII chromosome(10)), 29 were in single contigs; 21 of the chromosomes contained both telomere caps; 8 contained one of the caps; and 2 were missing both caps. Some chromosomes contain sequence from both parental sub-genomes due to recombinations; those chromosomes were named SeIII-ScIII, SeVII-ScVII, ScX-SeX, SeX-ScX and SeXIII-ScXIII, in accordance with previous nomenclature (10). Annotation of the assembly resulted in the identification of 10,632 genes (Additional File 1A). We determined chromosome copy number based on coverage analysis of short-read alignments to the genome assembly of CBS 1483 (Figure 2 and S2).

**Table 1. Length and gaps of each assembled chromosome of the *S. cerevisiae* and *S. eubayanus* subgenome in the *de novo* assembly of Group 2 *S. pastorianus* strain CBS 1483.** The mitochondrial DNA assembly is also shown.

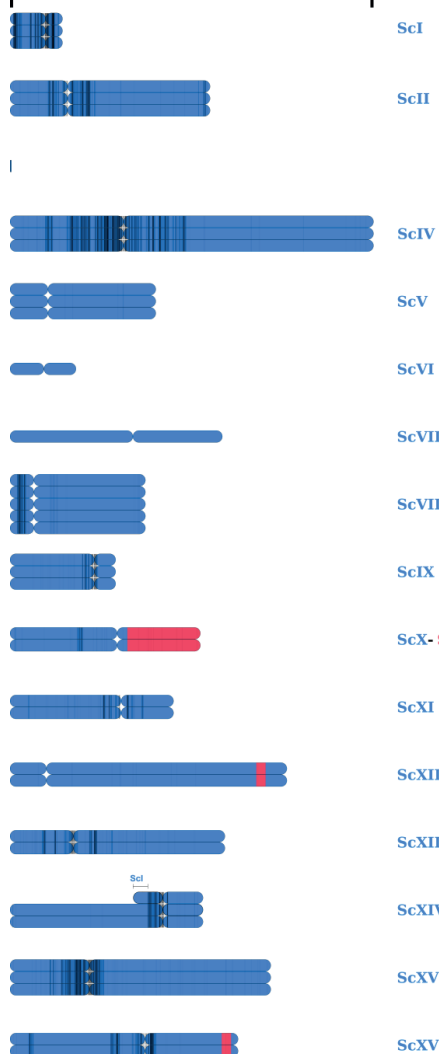
<i>S. cerevisiae</i> sub-genome			<i>S. eubayanus</i> sub-genome		
Contig/Scaffold	Size	Gaps	Contig/Scaffold	Size	Gaps
ScI	208794	0	SeI	183365	0
ScII	812290	0	SeII	1284912	0
ScIII	0	0	SeIII	311639	0
ScIV	1480484	0	SeIV	995872	0
ScV	590259	0	SeV	580717	0
ScVI	263951	0	SeVI	268897	0
ScVII	862436	0	SeVII	1048199	0
ScVIII	547874	0	SeVIII	813607	0
ScIX	426203	0	SeIX	413986	0
ScX	772632	0	SeX	698708	0
ScXI	662864	0	SeXI	658371	0
ScXII	1128411	2	SeXII	1043408	0
ScXIII	872991	0	SeXIII	966749	0
ScXIV	783474	0	SeXIV	765784	1
ScXV	1060500	0	SeXV	754183	0
Sc XVI	926828	0	SeXVI	788293	0
Unplaced	36198	0	Mitochondria	68765	0



**Figure 1. Structural heterozygosity within multiple copies of the *S. cerevisiae* chromosome I of CBS 1483.** (A) Layout of *S. cerevisiae* chromosome I in the assembly graph. Paths 1 and 2 represent alternative contigs in the right-end of the chromosome—the gene *UIP3* is deleted in path 2. (B) Sequencing coverage of ONT read-alignments of CBS 1483 in the right-end of chromosome I after joining path 1 and discarding path 2. The location of the *UIP3* gene is indicated. (C) Alignment overview of five raw ONT reads supporting the introgression of a ~14 Kbp in chromosome I (salmon colour) to a region at the right-end of chromosome XIV (brown colour) in the *S. cerevisiae* sub-genome. The additional alignments (pink and orange) are alignments to computationally confirmed Ty-2 repetitive elements. (D) Schematic representation of the two chromosome architectures of *S. cerevisiae* chromosome XIV (brown colour) due to translocation of an additional copy of the right arm of chromosome I (salmon colour) to the left arm of chromosome XIV.

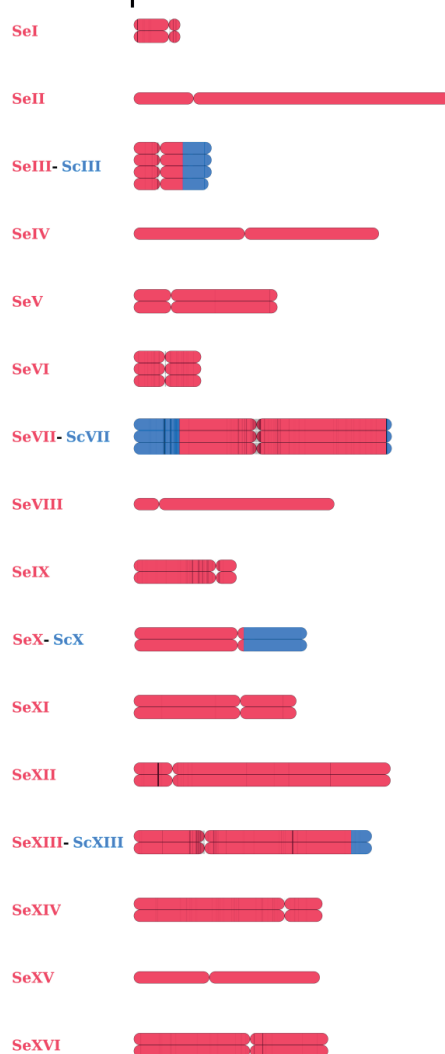
## *S.cerevisiae* sub-genome

max size = 1.48 Mbp



## *S.eubayanus* sub-genome

max size = 1.28 Mbp



**Figure 2. Overview of Nanopore-only *de novo* genome assembly of the *S. pastorianus* strain, CBS 1483.** For each chromosome, all copies are represented. Genomic material originating from *S. cerevisiae* (blue) and from *S. eubayanus* (red) are shown, and the position of the centromere is indicated. Heterozygous SNP calls are represented as vertical, black lines and are drawn with transparency to depict the density of SNP calls in a given region. Underlying chromosome copy number data and the list of heterozygous SNPs is available in Figure S2 and Additional File 1F.

### Comparison between ONT and Illumina assemblies

In order to compare our novel nanopore assembly of CBS 1483 to the previous assembly generated using short-read data, we aligned contigs of CBS 1483 from van den Broek *et al.* (10) to our current ONT-assembly, revealing a total 1.06 Mbp of added sequence. The added sequence overlapped with 323 ORFs (Additional File 1B). Conversely, aligning the nanopore assembly to the van den Broek *et al.* 2017 assembly revealed that only 14.9 Kbp of sequence was lost, affecting 15 ORFs (Additional File 1C). Gene ontology analysis of the added genes showed enrichment of several biological processes, functions, and components such as flocculation (P-value =  $7.44 \times 10^{-3}$ ) as well as transporter activity for several sugars including mannose, fructose and glucose (P-value  $\leq 1.5 \times 10^{-5}$ ) (Additional File 1D). Among the added genes were various members of subtelomeric gene families such as the *FLO*, *SUC*, *MAL*, *HXT* and *IMA* genes (Additional file 1E). Due to their role in the brewing-relevant traits such as carbohydrate utilization and flocculation, the complete assembly of subtelomeric gene families is crucial to capture different gene versions and copy number effects.

The assembly of CBS 1483 contained 9 *MAL* transporters, which encode for the ability to import maltose and maltotriose (36-38), constituting 85 % of fermentable sugar in brewer's wort (39). The *S. cerevisiae* subgenome harboured *ScMAL31* on ScII, *ScMAL11* on ScVII and on SeVII-ScVII, and *ScMAL41* on ScXI (Additional File 1B and 1E). However, the *ScMAL11* gene, also referred to as *AGT1*, was truncated, and there was no *ScMAL21* gene due to the complete absence of ScIII, as reported previously (10, 12). In the *S. eubayanus* subgenome, *MAL31*-type transporter genes were found in SeII, SeV, and SeXIII-ScXIII, corresponding to the location of the *S. eubayanus* transporter genes *SeMALT1*, *SeMALT2* and *SeMALT3*, respectively (25). In addition, a *MAL11*-like transporter was found on SeXV. Consistently with previous reports, no *MTY1*-like maltotriose transporter was found in CBS 1483 (10). Due to the absence of *MTY1* and the truncation of *ScMAL11*, maltotriose utilisation is likely to rely on the *SeMAL11* transporter in CBS 1483. Indeed, a *MAL11*-like transporter was recently shown to confer maltotriose utilisation in an *S. eubayanus* isolate from North Carolina (40).

The assembly also contained 14 *FLO* genes encoding flocculins which cause cell mass sedimentation upon completion of sugar consumption (34, 41, 42). The heavy flocculation of *S. pastorianus* cells simplifies biomass separation at the end of the brewing process, and resulted in their designation as bottom-fermenting yeast (43). Flocculation is mediated by flocculins: lectin-like cell wall proteins which effect cell-to-cell adhesion. In CBS 1483, we identified 12 flocculin genes, in addition to two *FLO8* transcriptional activators of flocculins (Additional File 1E). Flocculation intensity has been correlated to the length of flocculin genes (44-46). Specifically, increased length and number of tandem repeats within the *FLO* genes caused increased flocculation (46, 47). We therefore analysed tandem repeats in *S. cerevisiae*, *S. eubayanus* and *S. pastorianus* genomes and found that most *FLO* genes contain a distinct repeat pattern: two distinct, adjacent sequences each with variable copy number (Table 2). The repeats in *FLO1*, *FLO5*, and *FLO9* of the *S. cerevisiae* strain S288C have the same repeats of 135 bp and 15 bp; while repeats are of 189 bp and 15 bp for *FLO10* and of 132 bp and 45 bp for *FLO11*. The same repeat structures can be found in the *S. eubayanus* strain CBS 12357 as *FLO1*, *FLO5*, and *FLO9* contain repeats of 156 and 30 bp; although were unable to find clear repeat patterns for *FLO10* and *FLO11* in this genome. In *S. pastorianus* CBS 1483, the repeat lengths of *FLO* genes corresponded to the subgenome they were localized in (Table 2). Compared to the non-flocculent S288C and CBS 12357 strains, *FLO* genes were systematically shorter in CBS 1483, contrasting with available theory (41-49). The intense flocculation phenotype of *S. pastorianus* was previously attributed to a gene referred to as *LgFLO1* (48, 50, 51). However, alignment of previously



published partial and complete *LgFLO1* sequences did not confirm the presence of a similar ORF in CBS 1483. Moreover, the annotated *FLO* genes had higher identity with *S. eubayanus* and *S. cerevisiae* *FLO* genes, than with *LgFLO1*. Therefore, flocculation is likely to rely on one or several of the identified *FLO* genes from *S. cerevisiae* or *S. eubayanus* (Table 2).

**Table 2. Tandem repeat analysis in *FLO* genes. We found seven repeat sequences when analysing flocculation genes *FLO1*, *FLO5*, *FLO9*, *FLO10*, and *FLO11* in *S. cerevisiae* (S288C) and *S. eubayanus* (CBS 12357) genomes.** These sequences are referred to as sequence A (135 nt), B (15 nt), C (189 nt), D (45 nt), E (132 nt), F (156 nt), and G (30 nt). We used these sequences to analyse the copy numbers of each repeat within all *FLO* genes in our ONT-only assembly of CBS 1483 using the ONT-only S288C assembly as a control. Their respective copy numbers are shown below. Repeat sequences are indicated in Additional File 1H.

Species	(Sub)genome	Gene	Gene size (nt)	A	B	C	D	E	F	G	
<i>S. cerevisiae</i>	S288C	<i>FLO1</i>	4614	18.0	9.4	-	-	-	-	-	
		<i>FLO5</i>	3228	8.0	9.6	-	-	-	-	-	
		<i>FLO9</i>	3969	13.0	8.3	-	-	-	-	-	
		<i>FLO10</i>	3510	-	3.8	4.4	-	-	-	-	
		<i>FLO11</i>	4104	-	-	-	38.7	6.6	-	-	
	S288C (ONT)	<i>FLO1</i>	4615	18.0	9.4	-	-	-	-	-	
		<i>FLO5</i>	3228	8.0	9.6	-	-	-	-	-	
		<i>FLO9</i>	3978	13.0	8.3	-	-	-	-	-	
		<i>FLO10</i>	3508	-	3.8	4.4	-	-	-	-	
		<i>FLO11</i>	4104	-	-	-	38.7	6.6	-	-	
	CBS 1483	<i>FLO1</i> (ScVI)	1038	-	-	-	-	-	-	-	
		<i>FLO5</i> (ScI)	2603	1.0	11.1	-	-	-	-	-	
		<i>FLO9</i> (ScI)	2967	5.0	15.4	-	-	-	-	-	
		<i>FLO11</i> (ScIX)	2787	-	-	-	14.1	5.6	-	-	
	<i>S. eubayanus</i>	CBS 12357	<i>FLO1</i>	5517	-	-	-	-	-	24.7	2.8
			<i>FLO5</i>	1325	-	-	-	-	-	-	-
<i>FLO9</i> (SeI)			4752	-	-	-	-	-	8.3	45.9	
<i>FLO9</i> (SeVI)			3480	-	-	-	-	-	-	-	
<i>FLO9</i> (SeX)			4041	-	-	-	-	-	7.4	20.1	
<i>FLO9</i> (SeXII)			3321	-	-	-	-	-	-	10.2	
<i>FLO10</i> (SeXI)			4128	-	-	-	-	-	-	-	
<i>FLO11</i> (SeIX)			4149	-	-	-	-	-	-	-	
CBS 1483		<i>FLO5</i> (SeI)	1945	-	-	-	-	-	4.9	2.8	
		<i>FLO5</i> (SeI)	391	-	-	-	-	-	-	-	
		<i>FLO5</i> (SeVI)	3765	-	-	-	-	-	-	-	
		<i>FLO5</i> (SeXI)	2582	-	-	-	-	-	4.9	2.8	
		<i>FLO9</i> (SeI)	2100	-	-	-	-	-	3.0	3.8	
		<i>FLO9</i> (SeXII)	2892	-	-	-	-	-	-	6.3	
		<i>FLO10</i> (SeVI)	3378	-	-	-	-	-	-	-	
		<i>FLO11</i> (SeIX)	3909	-	-	-	-	-	-	-	

### Sequence heterogeneity in CBS 1483

As other Group 2 *S. pastorianus* strains, CBS 1483 displays heterozygosity between different copies of its *S. cerevisiae* subgenome (11). We therefore systematically identified heterozygous nucleotides in

its genome and investigated the ORFs with allelic variation. Using 156x coverage of paired-end Illumina library of CBS 1483, we found a total of 6,367 heterozygous SNPs across the genome (Additional File 1F). Although the heterozygous SNPs are present across the whole genome, they affect primarily the *S. cerevisiae* sub-genome, with the majority clustered around centromeres (Figure 2). Of these positions, 58 % were located within ORFs, resulting in 896 ORFs with allelic variation consisting of 1 to 30 heterozygous nucleotides. A total of 685 ORFs showed heterozygosity which would result in amino acid sequence changes, including 16 premature stop codons, 4 lost stop codons and 1566 amino acid substitutions (Additional File 1F). Gene ontology analysis of the ORFs affected by heterozygous calls revealed no significant enrichment in processes, functions of compartments. However, it should be noted that several industrially-relevant genes encoded more than one protein version, such as: the *BDH1* and *BDH2* genes, encoding butane-diol dehydrogenases involved in reduction of the off flavour compound diacetyl (52), the *FLO5* and *FLO9* genes encoding flocculins (49), and the *OAF1* gene encoding a regulator of ethyl-ester production pathway (53).

### Structural heterogeneity in CBS 1483 chromosomes

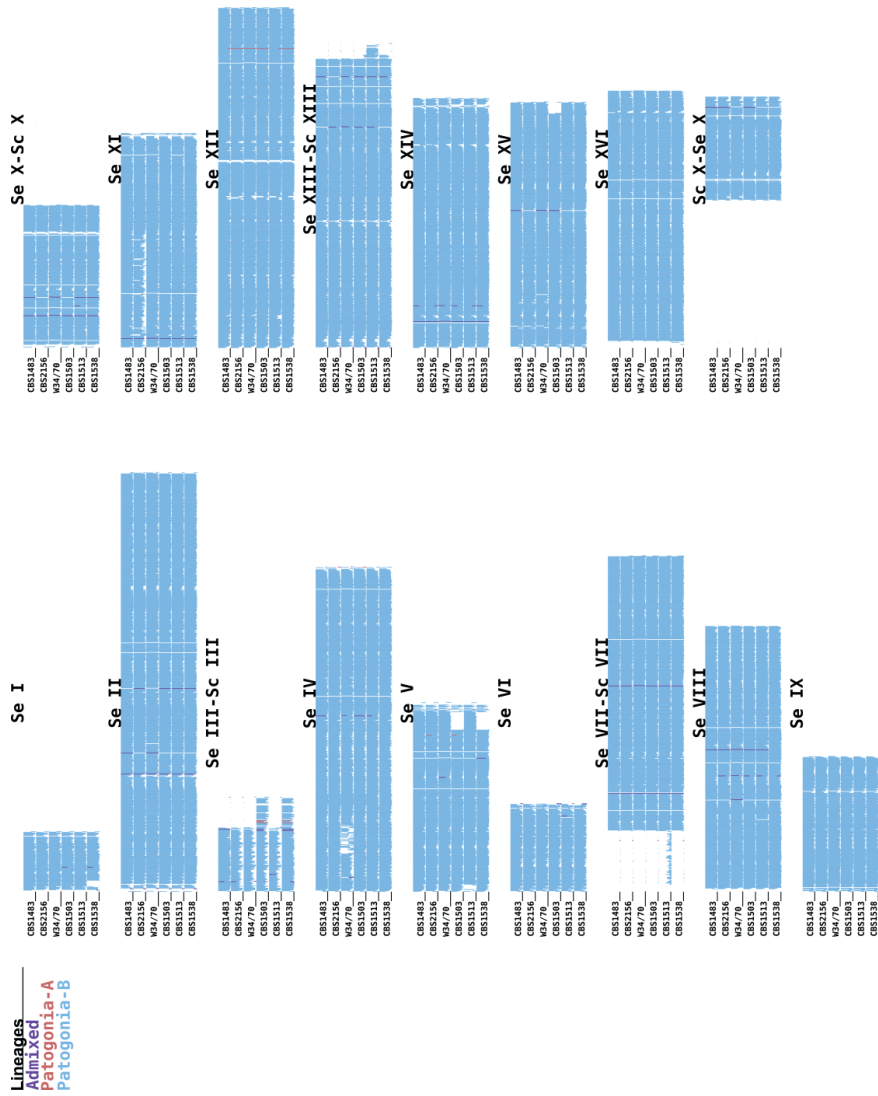
We investigated whether information about structural heterogeneity between chromosome copies could be recovered despite the fact that current assembly algorithms reduce genome assemblies to consensus sequences. Information about structural and sequence variation between different chromosome haplotypes is not captured by consensus assemblies. However, raw read data contains information for each chromosome copy. To identify structural heterogeneity, we identified ORFs whose predicted copy number deviated from that of the surrounding region in the chromosome based on read coverage analysis (Figure S3). We found 213 ORFs with deviating copy number (Additional File 1G). While no enrichment was found by gene ontology analysis, many of these ORFs are located in subtelomeric regions (29). Nevertheless, a few regions contained adjacent ORFs with deviating copy number, indicating larger structural variation between chromosome copies. For example, 21 consecutive ORFs in the right-end of the ScXV appear to have been deleted in 2 of the 3 chromosome copies (Figure S3). *UIP3*, one of the genes with deviating copy number, was located on the right arm of chromosome ScI. This region was previously identified as having an additional copy in CBS 1483, although it could not be localized based on short read data (10). The assembly graph showed two possible structures for ScI, which were collapsed into a single contig in the final assembly (Figure 1A). Sequence alignment, gene annotations and sequencing coverage indicated two versions of the ScI contigs: one with and one without the gene *UIP3* (Figure 1B). Sequence alignments of raw-ONT reads revealed five reads (from 20.6 to 36.7 Kbp) linking the right arm of ScI to the left arm of ScXIV at position ~561 Kbp (Figure 1C). This location corresponded to a Ty-2 repetitive element; known to mediate recombination within *Saccharomyces* genomes (21). In addition to the increased coverage of the right arm of ScI, the left arm of ScXIV showed decreased sequencing coverage up until the ~561 Kbp position. Together, these results suggest that the left arm of one copy of ScXIV was replaced with an additional copy of the right arm of ScI (Figure 1D). As no reads covered both the recombination locus and the *UIP3* locus, it remained unclear if *UIP3* is present in the ScI copy translocated to chromosome ScXIV. The resolution of two alternative chromosome architectures of ScI and ScXIV illustrates the ability of long-read alignment to resolve structural heterozygosity.

### **Differences between Group 1 and 2 genomes do not result from separate ancestry**

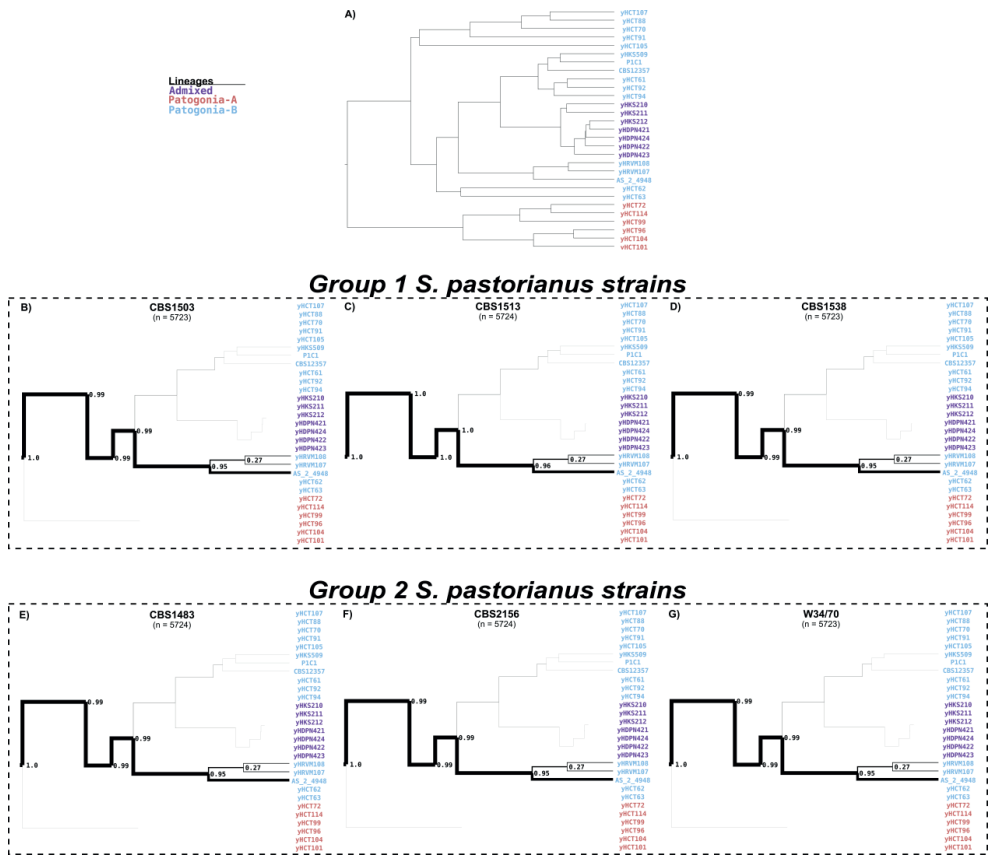
*S. pastorianus* strains can be subdivided into two separate groups—termed Group 1 and Group 2—based on both phenotypic (54) and genomic features (5, 11). However, the ancestral origin of each group remains unclear. The two groups may have emerged by independent hybridization events (19). Alternatively, Group 1 and Group 2 strains may originate from the same hybridization event, but Group 2 strains later hybridized with a different *S. cerevisiae* strain (11). In both cases, analysis of the provenance of genomic material from Group 1 and Group 2 genomes could confirm the existence of separate hybridization events if different ancestries are identified. Pan-genomic analysis of *S. cerevisiae* strains indicated that their evolution was largely non-linear, involving frequent horizontal gene transfer and sexual backcrossing events (55). Especially if the evolutionary ancestry of *S. pastorianus* involves admixture of different *S. cerevisiae* genomes (11), approaches considering only linear evolution such as phylogenetic trees are insufficient (56). Complex, non-linear evolutionary relationships could be addressed with network approaches (57). However, such algorithms are not yet fully mature and would involve extreme computational challenges (58, 59).

Therefore, we developed Alpaca: a simple and computationally inexpensive method to investigate complex non-linear ancestry via comparison of sequencing datasets. Alpaca is based on short-read alignment of a collection of strains to a partitioned reference genome, in which the similarity of each partition to the collection of strains is independently computed using k-mer sets. Reducing the alignments in each partition to k-mer sets prior to similarity analysis is computationally inexpensive. Phylogenetic relationships are also not recalculated, but simply inferred from previously available information on the population structure of the collection of strains. The partitioning of the reference genome enables the identification of strains with high similarity to different regions of the genome, enabling the identification of ancestry resulting from non-linear evolution. Moreover, since similarity analysis is based on read data, heterozygosity is taken into account.

We used Alpaca (60) to identify the most similar lineages for all non-overlapping 2 Kbp sub-regions in the genome of the Group 2 *S. pastorianus* strain CBS 1483 using a reference dataset of 157 *S. cerevisiae* strains (61) and 29 *S. eubayanus* strains (62). We inferred population structures for both reference datasets by using previously defined lineages of each strain along with hierarchical clustering based on genome similarity using MASH (63). For the *S. eubayanus* subgenome, almost all sub-regions of CBS 1483 were most similar to strains from the Patagonia B - Holartic lineage (62) (Figure 3). In fact, 68 % of all sub-regions were most similar to the Tibetan isolate CDFM21L.1 (64) and 27 % to two highly-related North-American isolates (Figure 4), indicating a monophyletic ancestry of the *S. eubayanus* genome. Analysis of *S. pastorianus* strains CBS 2156 and WS 34/70 (Group2), and of CBS 1503, CBS 1513 and CBS 1538 (Group 1), indicated identical ancestry of their *S. eubayanus* subgenomes (Figure 4). Overall, we did not discern differences in the *S. eubayanus* subgenomes of *S. pastorianus* strains, which seem to descend from a strain of the Patagonia B – Holartic lineage and which is most closely related to the Tibetan isolate CDFM21L.1.



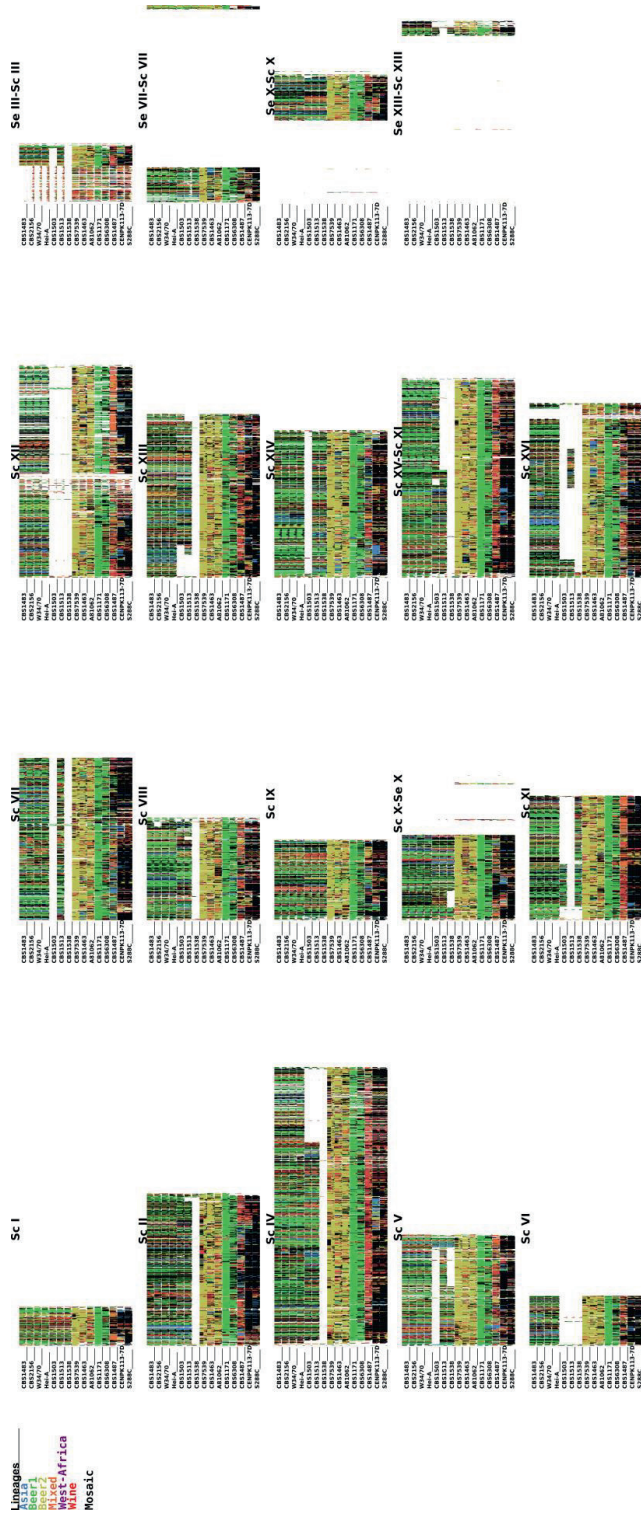
**Figure 3. Similarity profiles of the *S. eubayanus* (sub-)genomes of Group 1 and 2. *S. pastorianus* strains, as determined using Alpaca. Each *S. eubayanus* chromosome of the CBS 1483 assembly was partitioned in non-overlapping sub-regions of 2 Kbp. The colors represent the most similar lineages based on k-mer similarity of 29 *S. eubayanus* strains from Peris *et al* (62): admixed (purple), Patagonia-A (red), Patagonia-B (blue). Similarity patterns are shown for the Group 2 strains CBS 1483, CBS 2156 and WS34/70 and the Group 1 strains CBS 1503, CBS 1513 and CBS 1538.**



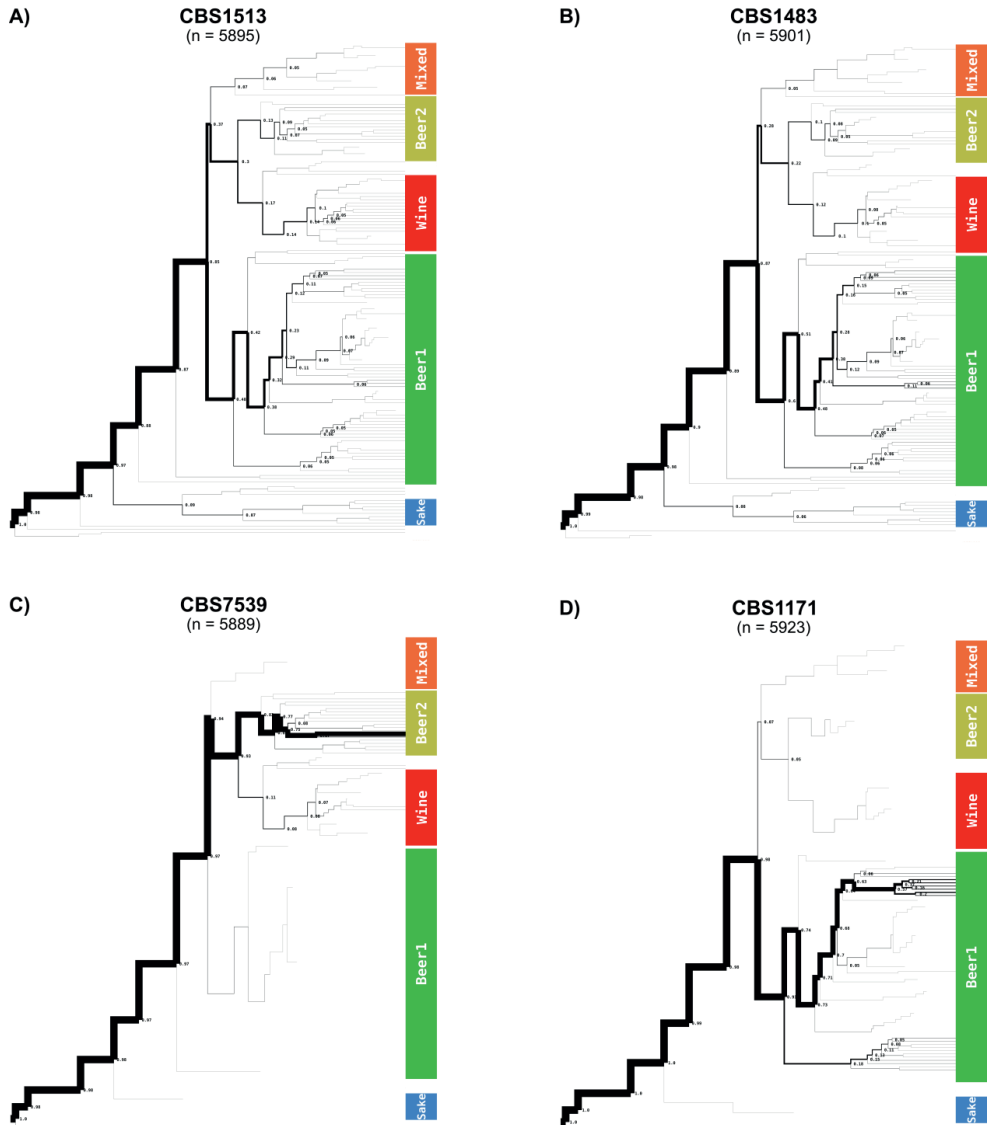
**Figure 4. Tree-tracing of the genome-scale similarity across the *S. eubayanus* (sub-)genomes of Group 1 and 2 *S. pastorianus* strains, as determined using Alpaca.** The frequency at which a genome from the reference data set of 29 *S. eubayanus* genomes from Peris *et al* (62) was identified as most similar for a sub-region of the CBS 1483 genome is depicted. The reference dataset is represented as a population tree, upon which only lineages with similarity are indicated with a thickness proportional to the frequency at which they were found as most similar ('N' being the total sum of the number of times all samples appeared as top-scoring). The complete reference population tree (A), the genomes of Group 1 strains CBS 1503, CBS 1513 and CBS 1538 (B-D) and for the genomes of Group 2 strains CBS 1483, CBS 2156 and WS34/70 (E-G) are shown.

In contrast, for the *S. cerevisiae* sub-genome of CBS 1483, the most similar *S. cerevisiae* strains varied across the sub-regions of every chromosome (Figure 5). No strain of the reference dataset was most similar for more than 5 % of sub-regions, suggesting a high degree of admixture (Figure 6). However, 60 % of sub-regions were most similar to the Beer 1 lineage, 12 % were most similar to the Wine lineage and 10 % to the Beer 2 lineage. In order to determine Alpaca's ability to differentiate genomes with different admixed ancestries, we analysed the genomes of 8 *S. cerevisiae* strains: six ale-brewing strains and the laboratory strains CEN.PK113-7D and S288C. The strains CBS 7539, CBS 1463 and A81062 were identified as similar to the Beer 2 lineage, CBS 1171 and CBS 6308 as similar to the Beer 1 lineage, CBS 1487 as similar to the Wine lineage, and CEN.PK113-7D and S288C as similar to the mosaic laboratory strains (Figure 5). In addition, the distribution of similarity over the *S. cerevisiae* population tree differed per strain (Figure 6 and S4). While no single strain was most similar for more than 8 % of the sub-regions for CBS 1487 and CBS 6308, for CBS 7539 67 % of sub-regions were most similar to the strain beer002. As both beer002 and CBS 7539 are annotated as Bulgarian beer yeast (55, 61), this similarity likely reflects common origin. The different similarity profiles of all *S. cerevisiae* strains indicate that Alpaca can differentiate different ancestry by placement of genetic material within the *S. cerevisiae* population tree, whether a genome has a linear monophyletic origin or a non-linear polyphyletic origin.

To identify possible differences in genome compositions within the *S. cerevisiae* subgenomes of *S. pastorianus*, we analysed other Group 1 and 2 strains using Alpaca, including an isolate of the Heineken A-yeast lineage (Hei-A), which was isolated in 1886 and represents one of the earliest pure yeast cultures. Whole genome sequencing, alignment to the CBS 1483 assembly and sequencing coverage analysis revealed that the ploidy of the Hei-A isolate corresponds to that of a Group 2 strain (Figure S5). Analysis of Hei-A and the other *S. pastorianus* Group 2 strains CBS 2156 and WS 34/70 using Alpaca yielded almost identical patterns of similarity at the chromosome-level as CBS 1483 (Figure 5). Moreover, similarity was distributed across the *S. cerevisiae* population tree almost identically as in CBS 1483 (Figure 6 and S4). The Group 1 *S. pastorianus* strains CBS 1503, CBS 1513 and CBS 1538 displayed different patterns of similarity at the chromosome-level relative to Group 2 strains. While various chromosome regions harboured almost identical similarity patterns, some regions differed significantly, such as: ScI, the middle of ScIV, the left arm of ScV, ScVIII, the right arm of ScIX, ScX-SeX, ScXI and ScXIII (Figure 5). However, at the genome level, similarity was distributed across the *S. cerevisiae* population tree almost identically as in Group 2 strains, except for a slightly higher contribution of the Beer 2 and Wine lineages, at the expense of a lower contribution of the Beer 1 lineage (Figure 6 and S4). The almost identical distribution of all Group 1 and Group 2 strains over the *S. cerevisiae* population tree indicate that they have the same *S. cerevisiae* ancestry. The spread of similarity across the *S. cerevisiae* population tree advocates for an admixed, possibly heterozygous ancestry of the *S. cerevisiae* subgenome of *S. pastorianus*. Furthermore, the different patterns of similarity at the chromosome level between both groups are compatible with an initially heterozygous *S. cerevisiae* subgenome which was subjected to independent loss of heterozygosity events in each group, resulting in differential retention of each haplotype. The lower relative contribution of Beer 1 strains in Group 1 strains may be explained by the complete absence of *S. cerevisiae* chromosomes with high similarity to Beer1 strains, such as ScV, ScXI and ScXv-ScXI.



**Figure 5. Similarity profiles of the *S. cerevisiae* (sub-)genomes of various *Saccharomyces* strains, as determined using Alpacca.** Each *S. cerevisiae* chromosome of the CBS 1483 assembly was partitioned in non-overlapping sub-regions of 2 Kbp. The colors represent the most similar lineages based on k-mer similarity of 157 *S. cerevisiae* strains from Gallone *et al* (61): Asia (blue), Beer1 (green), Beer2, (gold), Mixed (orange), West-Africa (purple), Wine (red). Mosaic strains are shown in black and ambiguous or low-similarity sub-regions in white. Similarity patterns are shown for the Group 2 *S. pastorianus* strains CBS 1483, CBS 2156, WS34/70 and Hei-A, for the Group 1 *S. pastorianus* strains CBS 1503, CBS 1543 and CBS 1538, for *S. cerevisiae* ale-brewing strains CBS 7539, CBS 1463, A81062, CBS 1171, CBS 6308 and CBS 1483, and for *S. cerevisiae* laboratory strains CEN.PK113-7D and S288C.



**Figure 6.** Tree-tracing of the genome-scale similarity across the *S. cerevisiae* (sub-)genomes of various *Saccharomyces* strains, as determined using Alpaca. The frequency at which a genome from the reference data set of 157 *S. cerevisiae* strains from Gallone *et al* (61) was identified as most similar for a sub-region of the CBS 1483 genome is depicted. The reference dataset is represented as a population tree, upon which only lineages with similarity are indicated with a thickness proportional to the frequency at which they were found as most similar ('N' being the total sum of the number of times all samples appeared as top-scoring). The genomes of *S. pastorianus* Group 1 strain CBS 1513 (A), of *S. pastorianus* Group 2 strain CBS 1483 (B), of *S. cerevisiae* strain CBS 7539 and of *S. cerevisiae* strain CBS 1171 are shown. The tree-tracing figures of *S. pastorianus* Group 1 strains CBS 1503 and CBS 1538, of *S. pastorianus* Group 2 strains CBS 2156, WS34/70 and Hei-A, and of *S. cerevisiae* strains CBS 1463, A81062, CBS 6308, CBS 1487, CEN.PK113-7D and S288C are shown in Figure S4.



## Discussion

In this study, we used Oxford Nanopore Technology's (ONT) MinION sequencing platform to study the genome of CBS 1483, an alloaneuploid Group 2 *S. pastorianus* strain. The presence of extensively aneuploid *S. cerevisiae* and *S. eubayanus* subgenomes substantially complicates analysis of *S. pastorianus* genomes (10). We therefore explored the ability of nanopore sequencing to generate a reference genome in the presence of multiple non-identical chromosome copies, and investigated the extent to which structural and sequence heterogeneity can be reconstructed. Despite its aneuploidy, we obtained a chromosome-level genome haploid assembly of CBS 1483 in which 29 of the 31 chromosomes were assembled in a single contig. Comparably to assemblies of euploid *Saccharomyces* genomes (22-27), nanopore sequencing resulted in far lesser fragmentation and in the addition of considerable sequences compared to a short-read based assembly of CBS 1483, notably in the subtelomeric regions (10). The added sequences enabled more complete identification of industrially-relevant subtelomeric genes such as the *MAL* genes, responsible for maltose and maltotriose utilisation (36-38), and the *FLO* genes, responsible for flocculation (34, 41, 42). Due to the instability of subtelomeric regions (28-30), the lack of reference-based biases introduced by scaffolding allows more certainty about chromosome structure (24). Since subtelomeric genes encode various industrially-relevant traits (31-34), their mapping enables further progress in strain improvement of lager brewing yeasts. Combined with recently developed Cas9 gene editing tools for *S. pastorianus* (65), accurate localisation and sequence information about subtelomeric genes is critical to investigate their contribution to brewing phenotypes by enabling functional characterization (66).

Despite the presence of non-identical chromosome copies in CBS 1483, the genome assembly only contained one contig per chromosome. While the assembly did not capture information about heterogeneity, mapping of short-read data enabled identification of sequence heterozygosity across the whole genome. In previous work, two alternative chromosome structures could be resolved within a population of euploid *S. cerevisiae* strain CEN.PK113-7D by alignment of nanopore reads (24). Therefore, we evaluated the ability to identify structural heterogeneity by aligning nanopore read data to the assembly. Indeed, nanopore-read alignments enabled the identification of two versions of chromosome *ScI*: with and without an internal deletion of the gene *UIP3*. Furthermore, the length of nanopore reads enabled them to span a TY-element, revealing that one of the copies of the right arm of *ScI* was translocated to the left arm of *ScXIV*. While the two alternative structures of *ScI* constitute a first step towards the generation of chromosome copy haplotypes, nanopore reads only enabled the hypothesis-based resolution of suspected heterogeneity. Assembly algorithms which do not generate a single consensus sequence per chromosome are emerging (67, 68). However, haplotyping is particularly difficult in aneuploid and polyploid genomes due to copy number differences between chromosomes (67). A further reduction of the relatively high error rate of nanopore reads, or the use of more accurate long-read sequencing technologies, could simplify the generation of haplotype-level genome assemblies in the future by reducing noise (69).

We used the chromosome-level assembly of CBS 1483 to study the ancestry of *S. pastorianus* genomes. Due to the importance of non-linear evolution in the domestication process of *Saccharomyces* strains (55), and to the admixed hybrid nature of *S. pastorianus* (11, 62), we used the newly-developed method Alpaca to analyse the ancestry of CBS 1483 instead of classical phylogenetic approaches using reference datasets of *S. cerevisiae* and *S. eubayanus* strains (61, 62). All *S. pastorianus* genomes displayed identical distribution of similarity across the reference

*S. eubayanus* population tree, both at the chromosome and whole-genome level. All *S. pastorianus* genomes also showed identical distribution of similarity across the reference *S. cerevisiae* population tree at the whole genome level; however, Group 1 and Group 2 strains displayed different similarity patterns at the chromosome level. The absence of differences in the *S. cerevisiae* genome at the whole genome level and recurrence of identical chromosomal break points between Group 1 and 2 strains discredit previous hypotheses of different independent hybridization events in the evolution of Group 1 and 2 strains (11, 19). Instead, these results are compatible with the emergence of Group 1 and 2 strains from a single shared hybridization event between a homozygous *S. eubayanus* genome closely related to the Tibetan isolate CDFM21L.1 and an admixed heterozygous *S. cerevisiae* genome with a complex polyphyletic ancestry. Loss of heterozygosity is frequently observed in *Saccharomyces* genomes (55, 70), and therefore likely to have affected both the genomes of Group 1 and 2 strains (11, 71, 72). The different chromosome-level similarity patterns in both groups likely emerged through different loss of heterozygosity events in Group 1 and 2 strains (71, 72). In addition, the lower *S. cerevisiae* chromosome content of Group 1 is consistent with observed loss of genetic material from the least adapted parent during laboratory evolution of *Saccharomyces* hybrids (73-76). In this context, the lower *S. cerevisiae* genome content of Group 1 strains may have resulted from a rare and serendipitous event. For example, chromosome loss has been observed due to unequal chromosome distribution a sporulation event of an allopolyploid *Saccharomyces* strain (77). Such mutant may have been successful if loss of *S. cerevisiae* chromosomes provided a selective advantage in the low-temperature lager brewing environment (73, 74). The loss of the *S. cerevisiae* subgenome may have affected only Group 1 strains due to different brewing conditions during their domestication. However, the high conservation of similarity within Group 1 and Group 2 strains indicate that the strains within each Group are closely related, indicating a strong population bottleneck in their evolutionary history.

Such a bottleneck could have been caused by the isolation and propagation of a limited number *S. pastorianus* strains, which may have eventually resulted in the extinction of other lineages. The first *S. pastorianus* strains isolated in 1883 by Hansen at the Carlsberg brewery were all Group 1 strains (13, 78). Due to the industry practice of adopting brewing methods and brewing strains from successful breweries, Hansen's Group 1 isolates likely spread to other breweries as these adopted pure culture brewing (1). Many strains which were identified as Group 2 by whole genome sequencing were isolated in the Netherlands (5, 11): Elion isolated the Heineken A-yeast in 1886 (79), CBS 1484 was isolated in 1925 from the Oranjeboom brewery (5), CBS 1483 was isolated in 1927 in a Heineken brewery (10), and CBS 1260, CBS 2156 and CBS 5832 were isolated from unknown breweries in the Netherlands in 1937, 1955 and 1968, respectively (5, 80). Analogously to the spread of Group 1 strains from Hansen's isolate, Group 2 strains may have spread from Elion's isolate. Both Heineken and Carlsberg distributed their pure culture yeast biomass to breweries over Europe and might therefore have functioned as an evolutionary bottleneck by supplanting other lineages with their isolates (81, 82). Overall, our results support that the differences between Group 1 and 2 strains emerged by differential evolution after an initial shared hybridization event, and not by a different *S. eubayanus* and/or *S. cerevisiae* ancestry.

Beyond its application in this study, we introduced Alpaca as a method to evaluate non-linear evolutionary ancestry. The use of short-read alignments enables Alpaca to account for sequence heterozygosity when assessing similarity between two genomes and is computationally inexpensive as they are reduced to k-mer sets. Moreover, Alpaca leverages previously determined phylogenetic

relationships within the reference dataset of strains to infer the evolutionary relationship of the reference genome to the dataset of strains. Due to the presence of non-linear evolutionary processes in a wide range of organisms (83, 84), the applicability of Alpaca extends far beyond the *Saccharomyces* genera. For example, genetic introgressions from *Homo neanderthalensis* constitute about 1 % of the human genome (85). Horizontal gene transfer is even relevant across different domains of life: more than 20 % of ORFs of the extremely thermophilic bacteria *Thermotoga maritima* were more closely related to genomes of Archaea than to genomes of other Bacteria (86). Critically, horizontal gene transfer, backcrossing and hybridization have not only played a prominent role in the domestication of *Saccharomyces* yeasts (55), but also in other domesticated species such as cows, pigs, wheat and citrus fruits (87-90). Overall, Alpaca can significantly simplify the analysis of new genomes in a broad range of contexts when reference phylogenies are already available.

## Conclusions

With 29 of the 31 chromosomes assembled in single contigs and 323 previously unassembled genes, the genome assembly of CBS 1483 presents the first chromosome-level assembly of a *S. pastorianus* strain specifically, and of an alloaneuploid genome in general. While the assembly only consisted of consensus sequences of all copies of each chromosome, sequence and structural heterozygosity could be recovered by alignment of short and long-reads to the assembly, respectively. We developed Alpaca to investigate the ancestry of Group 1 and Group 2 *S. pastorianus* strains by computing similarity between short-read data from *S. pastorianus* strains relative to large datasets of *S. cerevisiae* and *S. eubayanus* strains. In contrast with the hypothesis of separate hybridization events, Group 1 and 2 strains shared similarity with the same reference *S. cerevisiae* and *S. eubayanus* strains, indicating shared ancestry. Instead, differences between Group 1 and Group 2 strains could be attributed to different patterns of loss of heterozygosity subsequent to a shared hybridization event between a homozygous *S. eubayanus* genome closely related to the Tibetan isolate CDFM21L.1 and an admixed heterozygous *S. cerevisiae* genome with a complex polyphyletic ancestry. We identified the Heineken A-Yeast isolate as a Group 2 strain. We hypothesize that the large differences between Group 1 and Group 2 strains and the high similarity within Group 1 and 2 strains result from a strong population bottleneck which occurred during the isolation of the first Group 1 and Group 2 strains, from which all current *S. pastorianus* strains descend. Beyond its application in this study, the ability of Alpaca to reveal non-linear ancestry without requiring heavy computations presents a promising alternative to phylogenetic network analysis to investigate horizontal gene transfer, backcrossing and hybridization.

## Methods

### Yeast strains, cultivation techniques and genomic DNA extraction

*Saccharomyces* strains used in this study are indicated in Table 3. *S. pastorianus* strain CBS 1483, *S. cerevisiae* strain S288C and *S. eubayanus* strain CBS 12357 were obtained from the Westerdijk Fungal Biodiversity Institute (<http://www.westerdijkinstituut.nl/>). *S. eubayanus* strain CDFM21L.1 was provided by Prof. Feng-Yan Bai. An isolate from the *S. pastorianus* Heineken A-Yeast lineage (Hei-A) was obtained from HEINEKEN Supply Chain B.V., Zoeterwoude, the Netherlands. All strains were stored at -80°C in 30 % glycerol (vol/vol). Yeast cultures were inoculated from frozen stocks into 500-mL shake flasks containing 100 mL liquid YPD medium (containing 10 g L<sup>-1</sup> yeast extract, 20 g L<sup>-1</sup> peptone and 20 g L<sup>-1</sup> glucose) and incubated at 12 °C on an orbital shaker set at 200 rpm until the strains reached stationary phase with an OD<sub>660</sub> between 12 and 20. Genomic DNA was isolated using

the Qiagen 100/G kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and quantified using a Qubit® Fluorometer 2.0 (ThermoFisher Scientific, Waltham, MA).

### Short-read Illumina sequencing

Genomic DNA of CBS 1483 and CDFM21L.1 was sequenced on a HiSeq2500 sequencer (Illumina, San Diego, CA) with 125 bp paired-end reads with an insert size of 550 bp using PCR-free library preparation by Keygene (Wageningen, The Netherlands). Genomic DNA of the Heineken A-Yeast isolate Hei-A was sequenced in house on a MiSeq sequencer (Illumina) with 300 bp paired-end reads using PCR-free library preparation. All Illumina sequencing data are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the bioproject accession number PRJNA522669.

### MinION sequencing and basecalling

A total of four MinION genomic libraries of CBS 1483 were created using different chemistries and flow cells: one library using 2D-ligation (Sequencing Kit SQK-MAP006) with a R7.3 chemistry flow cell (FLO-MIN103); two libraries using 2D-ligation (Sequencing Kit SQK-NSK007) with two R9 chemistry flow cells (FLO-MIN105); and one library using 1D-ligation (Sequencing Kit SQK-LASK108) with a R9 chemistry flow cell (FLO-MIN106). All libraries were constructed using the same settings as previously described (24) and reads were uploaded and basecalled using the Metrichor desktop agent (<https://metrichor.com/s/>). All nanopore sequencing data are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the bioproject accession number PRJNA522669.

**Table 3: *Saccharomyces* strains used in this study.** For strains of the reference dataset, please refer to their original publication (61, 62).

Name	Species	Description	Reference
CBS 1483	<i>S. pastorianus</i>	Group 2	(10)
CBS 2156	<i>S. pastorianus</i>	Group 2	(11)
WS 34/70	<i>S. pastorianus</i>	Group 2	(12)
Heineken A-Yeast	<i>S. pastorianus</i>	Group 2	This study
CBS 1503	<i>S. pastorianus</i>	Group 1	(11)
CBS 1513	<i>S. pastorianus</i>	Group 1	(13)
CBS 1538	<i>S. pastorianus</i>	Group 1	(11)
S288C	<i>S. cerevisiae</i>	Laboratory strain	(102)
CEN.PK113-7D	<i>S. cerevisiae</i>	Laboratory strain	(24)
CBS 7539	<i>S. cerevisiae</i>	Ale brewing strain	(55)
CBS 1463	<i>S. cerevisiae</i>	Ale brewing strain	(55)
A81062	<i>S. cerevisiae</i>	Ale brewing strain	(18)
CBS 1171	<i>S. cerevisiae</i>	Ale brewing strain	(55)
CBS 6308	<i>S. cerevisiae</i>	Ale brewing strain	(55)
CBS 1487	<i>S. cerevisiae</i>	Ale brewing strain	(55)
CBS 12357	<i>S. eubayanus</i>	Patagonian Isolate	(4)
CDFM21L.1	<i>S. eubayanus</i>	Tibetan isolate	(64)

### De novo genome assembly

The genome of CBS 1483 was assembled *de novo* using only the ONT sequencing data generated in this study. The assembly was generated using Canu (91), polished using Pilon (35) and annotated using MAKER2 (92), as previously described (24) with some modifications: Pilon (version 1.22) was only used to polish sequencing errors in the ONT-only *de novo* assembly, and Minimap2 (93) (version 2.7) was used as the long-read aligner to identify potential misassemblies and heterozygous structural variants, which were visualized using Ribbon (94). The resulting assembly was manually

curated: (i) a contig of 24 Kbp comprised entirely of “TATATA” sequence was discarded; (ii) three contigs of 592, 465, and 95 Kbp (corresponding to the rDNA locus of the *S. cerevisiae* sub-genome) and complete sequence up and downstream of this locus were joined with a gap; (iii) four contigs corresponding to *S. cerevisiae* chromosome I (referred to as *ScI*) were joined without a gap into a complete 208 Kbp chromosome assembly (Figure 2A); (iv) two contigs corresponding to *ScXIV* were joined with a gap (Figure 2D); and (v) 23 Kbp of overlapping sequence from the mitochondrial contig corresponding to the origin of replication was identified with Nucmer (95) and manually removed when circularizing the contig, leading to the complete a final size of 69 Kbp. The assembled genomes are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the bioproject accession number PRJNA522669. Gene annotations are available in Additional File 1A.

### **Comparison between ONT-only and Illumina-only genome assembly**

Gained and lost sequence information in the nanopore assembly of CBS 1483 was determined by comparing it to the previous short-read assembly (10), as previously described (24) with the addition of using minimum added sequence length of 25 nt.

### **FLO gene analysis**

We used Tandem Repeat Finder (version 4.09) (96) with recommended parameters to identify tandem repeat sequences in *FLO1* (SGDID:S000000084), *FLO5* (SGDID:S000001254), *FLO8* (SGDID:S000000911), *FLO9* (SGDID:S000000059), *FLO10* (SGDID:S000001810), and *FLO11* (SGDID:S000001458) of *S. cerevisiae* strain S288C (97) as well as in *FLO1*, *FLO5*, *FLO8*, *FLO9*, *FLO10* and *FLO11* of *S. eubayanus* strain CBS 12357 (25). The resulting tandem repeat sequences were then used as proxies to characterize *FLO* genes in our assembly of CBS 1483, in a previously generated assembly of *S. cerevisiae* strain CEN.PK113-7D (24) and the *Lg-FLO1* genes previously described in *S. cerevisiae* strain CMBSVM11 (GenBank HM358276) and *S. pastorianus* strain KBY001 (GenBank D89860.1) (50, 51). BLASTN (version 2.2.31+) (98) was then used to align the tandem sequences to each *FLO* gene. The alignments were further processed via an in-house script in the Scala programming language to identify repeat clusters by requiring a minimum alignment coverage of 0.5 and a maximum gap between two repeats of 3x times the repeat sequence length. The total number of copies was estimated by dividing the total size of the cluster by the repeat sequence length.

### **Intra-chromosomal heterozygosity**

Sequence variation was identified by aligning the short-read Illumina reads generated in this study to the ONT-only assembly with BWA (99) and calling variants with Pilon (35) using the `--fix "bases"`, `"local"` and `--diploid` parameters. To restrict false positive calls, SNPs were disregarded within 10 Kbp of the ends of the chromosomes, if minor alleles had a frequency below 15 % allele frequency, and if the coverage was below 3 reads.

Copy number variation for all chromosomes were estimated by aligning all short-reads to the ONT-only assembly. Reads were trimmed of adapter sequences and low-quality bases with Trimmomatic (100) (version 0.36) and aligned with BWA (99) (version 0.7.12). The median coverage was computed using a non-overlapping window of 100 nt, copy number was determined by comparing the coverage to that of the chromosome with the smallest median coverage. Additionally, copy number variation at the gene-level was also investigated based on whether the coverage of an individual gene significantly deviated from the coverage of the surrounding region. First, we defined contiguous chromosomal sub-regions with fixed copy number (Table S1). The mean and standard

deviation of coverages of these sub-regions were then computed using ONT-only alignments. Mean coverages of every gene was then computed and an uncorrected Z-test (101) was performed by comparing a gene's mean coverage and the corresponding mean and standard deviation of the pre-defined sub-region that the gene overlapped with.

### **Similarity analysis and lineage tracing of *S. pastorianus* sub-genomes using Alpaca**

We developed Alpaca (60) to investigate non-linear ancestry of a reference genome based on large sequencing datasets. Briefly, Alpaca partitions a reference genome into multiple sub-regions, each reduced to a k-mer set representation. Sequence similarities of the sub-regions are then independently computed against the corresponding sub-regions in a collection of target genomes. Non-linear ancestry can therefore be inferred by tracing the population origin of the most similar genome(s) in each sub-region. Detailed explanation Alpaca can be found in our method description (60).

Alpaca (version 1.0) was applied to the ONT CBS 1483 genome assembly to investigate the similarity of sub-regions from both sub-genomes to previously defined population lineages. For partitioning the CBS 1483 genome into sub-regions, we used a k-mer size of 21 and a sub-region size of 2 Kbp and used the short-read Illumina data of CBS 1483 produced in this study to assure accurate k-mer set construction. For investigating mosaic structures in the *S. cerevisiae* subgenome, we used 157 brewing-related *S. cerevisiae* genomes (project accession number PRJNA323691) which were subdivided in six major lineages: Asia, Beer1, Beer2, Mixed, West-Africa, Wine and Mosaic (61). For the *S. eubayanus* subgenome, we used 29 available genomes (project accession number PRJNA290017) which were subdivided in three major lineages: Admixed, Patagonia-A, and Patagonia-B (62). Raw-reads of all samples were trimmed Trimmomatic and filtered reads were aligned to CBS 1483 genome using BWA (99). Alpaca was also applied to several *Saccharomyces* genomes to investigate evolutionary similarities and differences between Group 1 and Group 2 *S. pastorianus* genomes. We used Group 1 strains CBS 1503, CBS 1513, and CBS 1538, and Group 2 strains CBS 2156 and WS34/70 (project accession number PRJDB4073) (11). As a control, eight *S. cerevisiae* genomes were analysed: ale strains CBS 7539, CBS 1463, CBS 1171, CBS 6308, and CBS 1487 (project accession number PRJEB13017) (55) and A81062 (project accession number PRJNA408119) (18), and laboratory strains CEN.PK113-7D (project accession number PRJNA393501) (24) and S288C (project accession number PRJEB14774) (23). Similarly, raw-reads for all strains were trimmed with Trimmomatic and aligned to the ONT CBS 1483 genome assembly using BWA. Partitioning of the additional *S. pastorianus* and *S. cerevisiae* genomes with Alpaca was performed by deriving k-mer sets from read-alignments only, assuring direct one-to-one comparison of all sub-regions across all genomes. K-mer size of 21 and sub-region size of 2 Kbp were used. The *S. cerevisiae* and *S. eubayanus* sequencing data were used to identify potential mosaic structures in these genomes. Lastly, *S. cerevisiae* and *S. eubayanus* strains were subdivided into subpopulations according to previously defined lineages (61, 62). MASH (version 2.1) (63) was then used to hierarchically cluster each genome based on their MASH distance using k-mer size of 21, sketch size of 1,000,000, and minimum k-mer frequency of 2. The resulting trees were used as population reference trees for Alpaca (60).

### **Acknowledgments**

We thank Prof. Feng Yan Bai for kindly providing us *S. eubayanus* strain CDFM21L.1.

## References

1. Meussdoerffer FG (2009) A comprehensive history of beer brewing. *Handbook of brewing: Processes, technology, markets*, ed Eßlinger HM (Wiley-VCH, Weinheim), pp 1-42.
2. Kodama Y, Kielland-Brandt MC, & Hansen J (2006) Lager brewing yeast. *Comparative genomics*, (Springer, Berlin), pp 145-164.
3. Dequin S (2001) The potential of genetic engineering for improving brewing, wine-making and baking yeasts. *Appl Microbiol Biotechnol* 56(5-6):577-588.
4. Libkind D, et al. (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA* 108(35):14539-14544.
5. Dunn B & Sherlock G (2008) Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res*.
6. de Barros Lopes M, Bellon JR, Shirley NJ, & Ganter PF (2002) Evidence for multiple interspecific hybridization in *Saccharomyces sensu stricto* species. *FEMS Yeast Res* 1(4):323-331.
7. Hebly M, et al. (2015) *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res* 15(3).
8. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2015) New lager yeast strains generated by interspecific hybridization. *J Ind Microbiol Biotechnol* 42(5):769-778.
9. Gorter de Vries AR, Pronk JT, & Daran J-MG (2017) Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts. *Appl Environ Microbiol*:AEM. 03206-03216.
10. Van den Broek M, et al. (2015) Chromosomal copy number variation in *Saccharomyces pastorianus* evidence for extensive genome dynamics in industrial lager brewing strains. *Appl Environ Microbiol*:AEM. 01263-01215.
11. Okuno M, et al. (2016) Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res* 23(1):67-80.
12. Nakao Y, et al. (2009) Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Res* 16(2):115-129.
13. Walther A, Hesselbart A, & Wendland J (2014) Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. *G3 (Bethesda)*:g3. 113.010090.
14. Hewitt SK, Donaldson IJ, Lovell SC, & Delneri D (2014) Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. *PLoS One* 9(3):e92203.
15. Liti G, Peruffo A, James SA, Roberts IN, & Louis EJ (2005) Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. *Yeast* 22(3):177-192.
16. Monerawela C, James TC, Wolfe KH, & Bond U (2015) Loss of lager specific genes and subtelomeric regions define two different *Saccharomyces cerevisiae* lineages for *Saccharomyces pastorianus* Group I and II strains. *FEMS Yeast Res* 15(2):fou008.
17. Rainieri S, et al. (2006) Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome. *Appl Environ Microbiol* 72(6):3968-3974.
18. Krogerus K, et al. (2016) Ploidy influences the functional attributes of *de novo* lager yeast hybrids. *Appl Microbiol Biotechnol* 100(16):7203-7222.
19. Baker E, et al. (2015) The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol Biol Evol* 32(11):2818-2831.
20. Gorter de Vries AR, et al. (2019) Laboratory evolution of a *Saccharomyces cerevisiae* × *S. eubayanus* hybrid under simulated lager-brewing conditions, *Frontiers in Genetics* 10:242.
21. Kim JM, Vanguri S, Boeke JD, Gabriel A, & Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 8(5):464-478.
22. Giordano F, et al. (2017) *De novo* yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci Rep* 7(1):3935.
23. Istace B, et al. (2017) *de novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6(2):1-13.
24. Salazar AN, et al. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN. PK113-7D. *FEMS Yeast Res* 17(7).
25. Brickwedde A, et al. (2018) Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357T. *Front Microbiol* 9:1786.
26. Yue J-X, et al. (2017) Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat Genet* 49(6):913.
27. McIlwain SJ, et al. (2016) Genome sequence and analysis of a stress-tolerant, wild-derived strain of *Saccharomyces cerevisiae* used in biofuels research. *G3 (Bethesda)*:g3. 116.029389.
28. Pryde FE, Huckle TC, & Louis EJ (1995) Sequence analysis of the right end of chromosome XV in *Saccharomyces cerevisiae*: an insight into the structural and functional significance of sub-telomeric repeat sequences. *Yeast* 11(4):371-382.

29. Bergström A, *et al.* (2014) A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* 31(4):872-888.
30. Brown CA, Murray AW, & Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol* 20(10):895-903.
31. Jordan P, Choe J-Y, Boles E, & Oreb M (2016) Hxt13, Hxt15, Hxt16 and Hxt17 from *Saccharomyces cerevisiae* represent a novel type of polyol transporters. *Sci Rep* 6:23502.
32. Teste M-A, François JM, & Parrou J-L (2010) Characterization of a new multigene family encoding isomaltases in the yeast *Saccharomyces cerevisiae*: the IMA family. *J Biol Chem*:jbc. M110. 145946.
33. Denayrolles M, de Villechenon EP, Lonvaud-Funel A, & Aigle M (1997) Incidence of *SUC-RTM* telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Curr Genet* 31(6):457-461.
34. Teunissen A & Steensma H (1995) The dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* 11(11):1001-1013.
35. Walker BJ, *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
36. Alves SL, *et al.* (2008) Molecular analysis of maltotriose active transport and fermentation by *Saccharomyces cerevisiae* reveals a determinant role for the *AGT1* permease. *Appl Environ Microbiol* 74(5):1494-1501.
37. Chang Y, Dubin R, Perkins E, Michels C, & Needleman R (1989) Identification and characterization of the maltose permease in genetically defined *Saccharomyces* strain. *J Bacteriol* 171(11):6148-6154.
38. Naumov GI, Naumova ES, & Michels C (1994) Genetic variation of the repeated *MAL* loci in natural populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genetics* 136(3):803-812.
39. Zastrow C, Hollatz C, De Araujo P, & Stambuk B (2001) Maltotriose fermentation by *Saccharomyces cerevisiae*. *J Ind Microbiol Biotechnol* 27(1):34-38.
40. Baker EP & Hittinger CT (2018) Evolution of a novel chimeric maltotriose transporter in *Saccharomyces eubayanus* from parent proteins unable to perform this function, *PLoS genetics* 15(4):e1007786.
41. Van Mulders SE, *et al.* (2009) Phenotypic diversity of Flo protein family-mediated adhesion in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 9(2):178-190.
42. Miki B, Poon NH, James AP, & Seligy VL (1982) Possible mechanism for flocculation interactions governed by gene *FLO1* in *Saccharomyces cerevisiae*. *J Bacteriol* 150(2):878-889.
43. Dengis PB, Nelissen L, & Rouxhet PG (1995) Mechanisms of yeast flocculation: comparison of top-and bottom-fermenting strains. *Appl Environ Microbiol* 61(2):718-728.
44. Fidalgo M, Barrales RR, & Jimenez J (2008) Coding repeat instability in the *FLO11* gene of *Saccharomyces* yeasts. *Yeast* 25(12):879-889.
45. Zara G, Zara S, Pinna C, Marceddu S, & Budroni M (2009) *FLO11* gene length and transcriptional level affect biofilm-forming ability of wild flor strains of *Saccharomyces cerevisiae*. *Microbiology* 155(12):3838-3846.
46. Verstrepen KJ, Jansen A, Lewitter F, & Fink GR (2005) Intragenic tandem repeats generate functional variability. *Nat Genet* 37(9):986.
47. Liu N, Wang D, Wang ZY, He XP, & Zhang B (2007) Genetic basis of flocculation phenotype conversion in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 7(8):1362-1370.
48. Ogata T, Izumikawa M, Kohno K, & Shibata K (2008) Chromosomal location of Lg-*FLO1* in bottom-fermenting yeast and the *FLO5* locus of industrial yeast. *J Appl Microbiol* 105(4):1186-1198.
49. Soares EV (2011) Flocculation in *Saccharomyces cerevisiae*: a review. *J Appl Microbiol* 110(1):1-18.
50. Van Mulders SE, *et al.* (2010) Flocculation gene variability in industrial brewer's yeast strains. *Appl Microbiol Biotechnol* 88(6):1321-1331.
51. Kobayashi O, Hayashi N, Kuroki R, & Sone H (1998) Region of Flo1 proteins responsible for sugar recognition. *J Bacteriol* 180(24):6503-6510.
52. Li P, *et al.* (2017) Reducing diacetyl production of wine by overexpressing *BDH1* and *BDH2* in *Saccharomyces uvarum*. *J Ind Microbiol Biotechnol* 44(11):1541-1550.
53. Saerens S, Thevelein J, & Delvaux F (2008) Ethyl ester production during brewery fermentation, a review. *Cerevisia* 33(2):82-90.
54. Gibson BR, Storgårds E, Krogerus K, & Vidgren V (2013) Comparative physiology and fermentation performance of Saaz and Froberg lager yeast strains and the parental species *Saccharomyces eubayanus*. *Yeast* 30(7):255-266.
55. Peter J, *et al.* (2018) Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556(7701):339.
56. Gogarten JP & Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3(9):679.
57. Solís-Lemus C, Bastide P, & Ané C (2017) PhyloNetworks: a package for phylogenetic networks. *Mol Biol Evol* 34(12):3292-3298.
58. Hejase HA & Liu KJ (2016) A scalability study of phylogenetic network inference methods using empirical datasets and simulations involving a single reticulation. *BMC bioinformatics* 17(1):422.
59. Consortium CP-G (2016) Computational pan-genomics: status, promises and challenges. *Brief Bioinform* 19(1):118-135.
60. Salazar A & Abeel T (2019) Alpaca: a kmer-based approach for investigating mosaic structures in microbial genomes. *bioRxiv*:551234.



61. Gallone B, *et al.* (2016) Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166(6):1397-1410. e1316.
62. Peris D, *et al.* (2016) Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast *Saccharomyces eubayanus*. *PLoS Genet* 12(7):e1006155.
63. Ondov BD, *et al.* (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 17(1):132.
64. Bing J, Han P-J, Liu W-Q, Wang Q-M, & Bai F-Y (2014) Evidence for a Far East Asian origin of lager beer yeast. *Curr Biol* 24(10):R380-R381.
65. Gorter de Vries AR, de Groot PA, Broek M, & Daran J-MG (2017) CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb Cell Fact* 16(1):222.
66. Winzeler EA, *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285(5429):901-906.
67. He D, Saha S, Finkers R, & Parida L (2018) Efficient algorithms for polyploid haplotype phasing. *BMC genomics* 19(2):110.
68. Chin C-S, *et al.* (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13(12):1050.
69. Wenger AM, *et al.* (2019) Highly-accurate long-read sequencing improves variant detection and assembly of a human genome. *bioRxiv*:519025.
70. Magwene PM, *et al.* (2011) Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 108(5):1987-1992.
71. Chambers SR, Hunter N, Louis EJ, & Borts RH (1996) The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol Cell Biol* 16(11):6110-6120.
72. González SS, Barrio E, & Querol A (2008) Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. *Appl Environ Microbiol* 74(8):2314-2320.
73. Smukowski Heil CS, *et al.* (2017) Loss of heterozygosity drives adaptation in hybrid yeast. *Molecular biology and evolution* 34(7):1596-1612.
74. Heil CS, Large CR, Patterson K, & Dunham MJ (2018) Temperature preference biases parental genome retention during hybrid evolution. *bioRxiv*:429803.
75. Pérez Través L, Lopes CA, Barrio E, & Querol A (2014) Study of the stabilization process in *Saccharomyces* intra- and interspecific hybrids in fermentation conditions. *Int Microbiol*. 17(4):213-224.
76. Antunovics Z, Nguyen H-V, Gaillardin C, & Sipiczki M (2005) Gradual genome stabilisation by progressive reduction of the *Saccharomyces uvarum* genome in an interspecific hybrid with *Saccharomyces cerevisiae*. *FEMS Yeast Res* 5(12):1141-1150.
77. Lopandic K, *et al.* (2016) Genotypic and phenotypic evolution of yeast interspecies hybrids during high-sugar fermentation. *Appl Microbiol Biotechnol* 100(14):6331-6343.
78. Hansen EC (1883) Recherches sur la physiologie et la morphologie des ferments alcooliques. V. Methodes pour obtenir des cultures pures de Saccharomyces et de microorganismes analogues. *Compt Rend Trav Lab Carlsberg* 2:92-105.
79. Gélinas P (2010) Mapping early patents on baker's yeast manufacture. *Compr Rev Food Sci Food Saf* 9(5):483-497.
80. Scheda R & Yarrow D (1968) Variation in the fermentative pattern of some *Saccharomyces* species. *Arch Mikrobiol* 61(3):310-316.
81. Hornsey IS (2003) *A history of beer and brewing* (Royal Society of Chemistry, Cambridge, UK).
82. Mendlik F (1937) Some aspects of the scientific development of brewing in Holland. *J Inst Brew* 43(4):294-300.
83. Keeling PJ & Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9(8):605.
84. Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 3(9):711.
85. Racimo F, Sankararaman S, Nielsen R, & Huerta-Sánchez E (2015) Evidence for archaic adaptive introgression in humans. *Nat Rev Genet* 16(6):359.
86. Nelson KE, *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399(6734):323.
87. Larson G, *et al.* (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307(5715):1618-1621.
88. McTavish EJ, Decker JE, Schnabel RD, Taylor JF, & Hillis DM (2013) New World cattle show ancestry from multiple independent domestication events. *Proc Natl Acad Sci USA* 110(15):E1398-E1406.
89. Brenchley R, *et al.* (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491(7426):705.
90. Wu GA, *et al.* (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotechnol* 32(7):656.
91. Koren S, *et al.* (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*:gr. 215087.215116.
92. Holt C & Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics* 12(1):491.
93. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18):3094-3100.

94. Nattestad M, Chin C-S, & Schatz MC (2016) Ribbon: visualizing complex genome alignments and structural variation. *bioRxiv*:082123.
95. Kurtz S, *et al.* (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12.
96. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* 27(2):573-580.
97. Cherry JM, *et al.* (2011) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic acids research* 40(D1):D700-D705.
98. Camacho C, *et al.* (2009) BLAST+: architecture and applications. *BMC bioinformatics* 10(1):421.
99. Li H & Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26(5):589-595.
100. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
101. Fisher RA (1937) *The design of experiments* (Oliver And Boyd; Edinburgh; London).
102. Goffeau A, *et al.* (1996) Life with 6000 genes. *Science* 274(5287):546-567.



## Chapter 5: CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*

Arthur R. Gorter de Vries #, Philip A. de Groot #, Marcel van den Broek, Jean-Marc G. Daran

# These authors contributed equally to this publication and should be considered co-first authors.

**Background:** The ease of use of CRISPR-Cas9 reprogramming, its high efficacy, and its multiplexing capabilities have brought this technology at the forefront of genome editing techniques. *Saccharomyces pastorianus* is an aneuploid interspecific hybrid of *Saccharomyces cerevisiae* and *Saccharomyces eubayanus* that has been domesticated for centuries and is used for the industrial fermentation of lager beer. For yet uncharacterised reasons, this hybrid yeast is far more resilient to genetic alteration than its ancestor *S. cerevisiae*.

**Results:** This study reports a new CRISPR-Cas9 method for accurate gene deletion in *S. pastorianus*. This method combined the *Streptococcus pyogenes cas9* gene expressed from either a chromosomal locus or from a mobile genetic element in combination with a plasmid-borne gRNA expression cassette. While the well-established gRNA expression system using the RNA polymerase III dependent *SNR52* promoter failed, expression of a gRNA flanked with Hammerhead and Hepatitis Delta Virus ribozymes using the RNA polymerase II dependent *TDH3* promoter successfully led to accurate deletion of all four alleles of the *SeILV6* gene in strain CBS 1483. Furthermore the expression of two ribozyme-flanked gRNAs separated by a 10-bp linker in a polycistronic array successfully led to the simultaneous deletion of *SeATF1* and *SeATF2*, genes located on two separate chromosomes. The expression of this array resulted in the precise deletion of all five and four alleles mediated by homologous recombination in the strains CBS 1483 and Weihenstephan 34/70 respectively, demonstrating the multiplexing abilities of this gRNA expression design.

**Conclusion:** These results firmly established that CRISPR-Cas9 is significantly facilitating and accelerating genome editing in *S. pastorianus*.

Essentially as published in *Microbial Cell Factories* 2017;16:222

Supplementary materials available online

<https://doi.org/10.1186/s12934-017-0835-1>

## Background

Lager beer is the most produced fermented beverage: in 2015 the worldwide production reached a global volume of  $170 \times 10^{19}$  L. The fermentation workhorse of lager brewing is *Saccharomyces pastorianus*, a natural interspecific hybrid of *Saccharomyces cerevisiae* and *Saccharomyces eubayanus* (1, 2) whose domestication is thought to have occurred in central Europe (Bohemia, nowadays Czech republic) in the late Middle Ages. Its ability to ferment at low temperature, to flocculate and to produce a vast range of flavour compounds make *S. pastorianus* well suited for the brewing process. In addition to their hybrid nature, *S. pastorianus* strains share a high degree of aneuploidy. While the first strain of *S. pastorianus* Weihenstephan 34/70 was sequenced in 2009 (2), the exact chromosome complement of lager yeast was revealed later with the introduction of next generation sequencing (3-6). Within *S. pastorianus* genomes, chromosomes may be completely absent or present in up to five copies and chromosome copy numbers vary widely across different strains (4). This intricate genome organisation significantly complicates functional gene analysis. Indeed, a simple gene deletion based on double crossover mediated by homologous recombination requires successive removal of all copies of the gene in both sub-genomes by several rounds of transformation. In association with a low propensity to perform homologous recombination, the difficulty to delete high copy number genes may explain the quasi-absence of examples of functional characterisation of *S. pastorianus* genes in the scientific literature based on impact of gene deletion (7-9). Instead a *S. pastorianus* gene or allele is usually cloned in *S. cerevisiae* and characterised based on the impact of the overexpression. However, such approaches do not take into account the role of the orthologous gene harboured by the other sub-genome, the possible occurrence of paralogs, and the gene expression regulation of the gene in its allo-aneuploid genetic background. Therefore, tools are needed to achieve efficient genome editing in allo-aneuploid *S. pastorianus* not only to enable targeted genetic modification, but also to enable functional gene analysis.

The exposed DNA strand ends resulting from a DNA double strand break (DSB) are extremely recombinogenic (10, 11). Even in *Saccharomyces cerevisiae* that exhibits a natural inclination to perform homologous recombination, introduction of a programmed DSB by combining the insertion of an I-SceI restriction site in a chromosomal locus and expression of the endonuclease encoding gene *SCEI* showed substantial stimulation of homologous recombination at the cut site enabling the correct assembly of multiple DNA fragments (12). Although efficient, the use of SceI induced DSB is limited since it requires the insertion of the recognition site prior its utilisation. In the past five years, the advent of the CRISPR (clustered regularly interspaced short palindromic repeat)–Cas9 (CRISPR-associated protein 9) system derived from *Streptococcus pyogenes* has considerably transformed genome engineering approaches (13, 14). The system comprises two elements: a short chimeric RNA that derives from the fusion of the tracr and crRNA called guide RNA (gRNA), and the endonuclease Cas9 (13, 14). By forming a complex with Cas9, the gRNA provides sequence specificity to the system. The hetero-duplex formed by the gRNA and the genomic target places the endonuclease which generates a blunt ended DSB. The system has been successfully implemented in *S. cerevisiae* (15-19), which broadened genome editing possibilities by allowing multiplexing (15, 16, 18) and high precision *in vivo* site-directed mutagenesis (15). The expression of the gRNA has been a point of attention since the gRNA secondary structures are crucial for the formation of the complex with Cas9. Therefore the 5' capping and 3' polyadenylation present in RNA Polymerase II transcripts have to be avoided. By analogy with the expression of gRNA in human cell lines (14), placing the gRNA

behind the control of a RNA Polymerase III dependent promoter (e.g. *SNR52p*) resulted in expression of an active gRNA lacking these modifications (16). In addition, due to the lack of polyadenylation-mediated export to the cytosol, RNA polymerase III transcribed gRNAs reside in the nucleus longer where they can form a complex with Cas9.

However gRNA expression from a RNA polymerase III was shown to result in low and unstable transcript levels (20). To overcome this issue while avoiding inactivation of the gRNA by 5' capping and 3' polyadenylation, the gRNA can be flanked by two ribozymes molecules and expressed by RNA polymerase II. Upon transcription the ribozymes self-cleave, resulting in removal of 5' and 3- ends and release of a mature gRNA (19, 21). Such CRISPR-Cas9 systems have been confirmed to mediate efficient genome editing in multiple cell types already, such as human cell lines (13, 14, 22), mice (23), zebrafish (24), *Caenorhabditis elegans* (25, 26), *Drosophila* (27), yeasts (15, 16, 28, 29), and plants (30-32).

The goal of the present study was to explore the use of CRISPR-Cas9 in *S. pastorianus*, a yeast with low genetic accessibility that is characterised by an unique allo-aneuploid genome. To this end, we present the construction of molecular tools to achieve efficient single and double simultaneous gene deletions. The successful application of this methodology offers an opportunity to get a deeper understanding of hybrid yeast biology.

## Methods

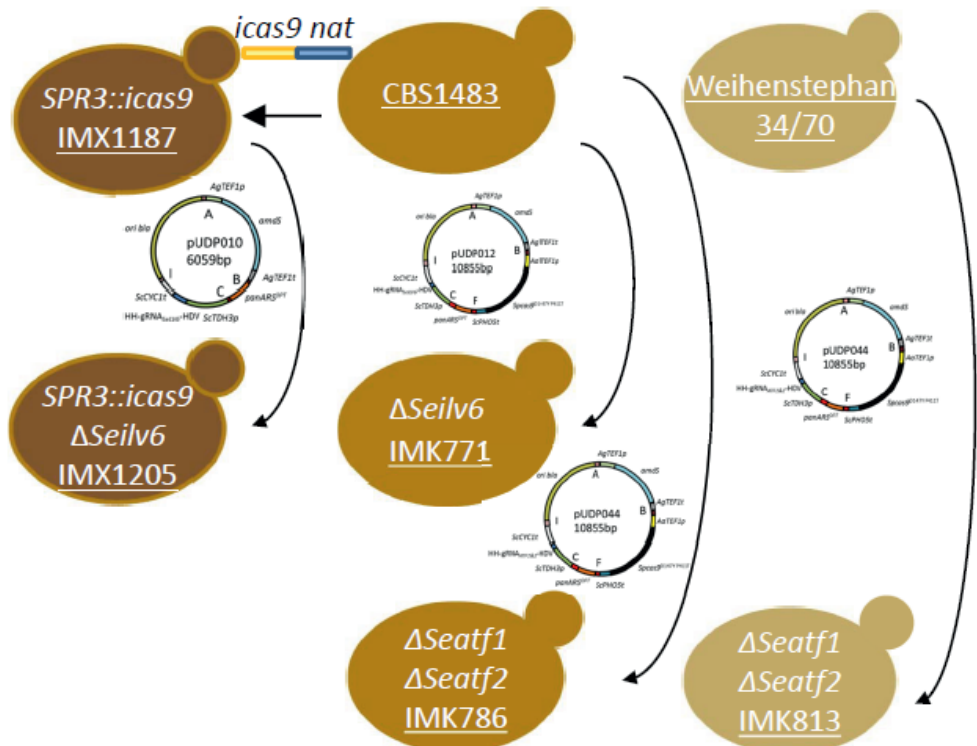
### Strains and growth conditions

The *S. pastorianus* and *cerevisiae* strains used in this study are listed in Table 1 and a construction flow-chart is provided in Figure 1.

**Table 1: Strains used throughout this study.**

Name	Species	Genotype	Source
CBS 1483	<i>S. pastorianus</i>	wildtype	(4)
IMX1187	<i>S. pastorianus</i>	<i>SPR3::AaTEF1p-Spcas9<sup>D147Y P411T</sup>-ScPHO5t</i>	This study
IMX1205	<i>S. pastorianus</i>	<i>SPR3::AaTEF1p-Spcas9<sup>D147Y P411T</sup>-ScPHO5t ΔSeilv6</i>	This study
IMK771	<i>S. pastorianus</i>	<i>ΔSeilv6</i>	This study
IMK786	<i>S. pastorianus</i>	<i>ΔSeatf1 ΔSeatf2</i>	This study
Weihenstephan 34/70	<i>S. pastorianus</i>	wildtype	(2, 66)
IMK813	<i>S. pastorianus</i>	<i>ΔSeatf1 ΔSeatf2</i>	This study
CEN.PK113-7D	<i>S. cerevisiae</i>	<i>MATa MAL2-8c</i>	(67)
IMX585	<i>S. cerevisiae</i>	<i>MATa can1Δ::AaTEF1p-Spcas9<sup>D147Y P411T</sup>-ScPHO5t natNT2</i>	(15)

Under nonselective conditions, *Saccharomyces pastorianus* and *cerevisiae* strains were grown in complex medium (YPD) containing 10 g L<sup>-1</sup> yeast extract, 20 g L<sup>-1</sup> peptone, and 20 g L<sup>-1</sup> glucose. For nourseothricin selection, YPD medium was supplemented with 100 μg.L<sup>-1</sup> of the antibiotic. Synthetic media (SM) containing 20 g L<sup>-1</sup> glucose, 3 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 5 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mL L<sup>-1</sup> of a trace element solution and of a vitamin solution was prepared as previously described (33). For selection of yeast strains harboring an acetamidase marker (34), (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was replaced by 0.6 g L<sup>-1</sup> acetamide as nitrogen source and 6.6 g L<sup>-1</sup> K<sub>2</sub>SO<sub>4</sub> to compensate for sulfate (SM-Ac). Loss of the acetamide marker was selected for on SM containing 2.3 g L<sup>-1</sup> fluoroacetamide (SM-Fac) (34). The pH in all media was adjusted to 6.0 with KOH. Solid media were prepared by adding 2 % agar to the various media. The strains of *S. pastorianus* and *S. cerevisiae* were incubated at 20 °C and 30 °C respectively.



**Figure 1: Strains construction flow-chart.** Schematic representation of the different strain lineages constructed in this study. The strain name is underlined and each arrow indicates a transformation step.

Shake flask cultures of *S. pastorianus* were grown at 20 °C in 500 mL flasks containing 100 mL complete medium (YPD) with 20 g·L<sup>-1</sup> glucose in an Innova 43/43R shaker (Eppendorf, Hauppauge, NY) set at 200 rpm. Frozen stocks were prepared by addition of glycerol (30 % v/v) to exponentially growing shake-flask cultures of *S. cerevisiae*, *S. pastorianus* and overnight cultures of *Escherichia coli* and stored aseptically in 1 mL aliquots at -80 °C.

For growth studies in shake flasks, *S. pastorianus* strains were grown in shake flasks with complete medium YPD. Growth rates were based on optical density at 660 nm (OD<sub>660</sub>) measurements using a Libra S11 spectrophotometer (Biochrom, Cambridge, United Kingdom). Specific growth rates were calculated from exponential fits of the OD<sub>660</sub> against time.

### Plasmid construction

All plasmids and primers used during this study are shown in Tables 2 and 3, respectively. The DNA parts harboured by the plasmids pUD527, pUD528, pUD530, pUD531, pUD532 pUD536 and pUD573 were *de novo* synthesised at GeneArt (Thermo Fisher Scientific, Waltham, MA). Unless specified, plasmids were propagated and stored in *E. coli* strain XL1-blue. Yeast transformation was done by electroporation using 50 µL of competent cells and up to 5 µL DNA as previously described (35) and transformed cells were incubated in 0.5 mL YPD during 1h, after which they were re-suspended in

100  $\mu$ L of sterile demi-water and plated on selective medium. High fidelity PCR amplification was performed using Phusion polymerase (Thermo Fisher Scientific) according to supplier's instructions.

pUD423 was assembled from plasmids pCT, pUD528 and pUC19. The *Streptococcus pyogenes cas9* open reading frame (*cas9*<sup>D147Y P411T</sup> (36)) was amplified from the plasmid pCT (Addgene plasmid #60621) (<https://www.addgene.org/>) using the primers 9390 and 9391. The *AaTEF1* promoter flanked upstream by short homology flank (SHF) B was amplified from the plasmid pUD528 using the primers 3841 and 9394. The *ScPHO5* terminator fragment was amplified from pUD528 using the primers 9392 and 9393, resulting in the addition of SHR F downstream of the terminator. The three fragments together with the pUC19 backbone (37) amplified with the primers 7389 and 9395 were assembled *in vitro* using Ligase Chain Reaction (LCR) with primers 9396-9399 as bridging oligonucleotides as described previously (38) and the resulting plasmid pUD423 was verified using digestion with NdeI.

The cassette for integration of *cas9* into the *SPR3* locus was assembled on pUD526. Flanks for homologous recombination of about 500 bp were amplified from genomic DNA of CBS 1483 using primers sets 10432/10433 and 10434/10435 adding NotI restriction sites upstream of the left homology arm and downstream of the right homology arm and 40 bp homology flanks on both sides of the homology arms for "Gibson" assembly (39). The *cas9* expression cassette was amplified from plasmid pUD423 using primers 10426 and 10427, the nourseothricin marker was amplified from pMEL15 (15) using primers 3597 and 10436 adding a 40 bp homology flank upstream of the *nat* gene, and the plasmid backbone was amplified from plasmid pUC19 using primers 7389 and 9395. Next, 0.2 pmol of each fragment were assembled into pUD526 using NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs, Ipswich, MA), verified by digestion with BamHI and NotI. The integration cassette was obtained by digestion of the plasmid using NotI followed by gel purification.

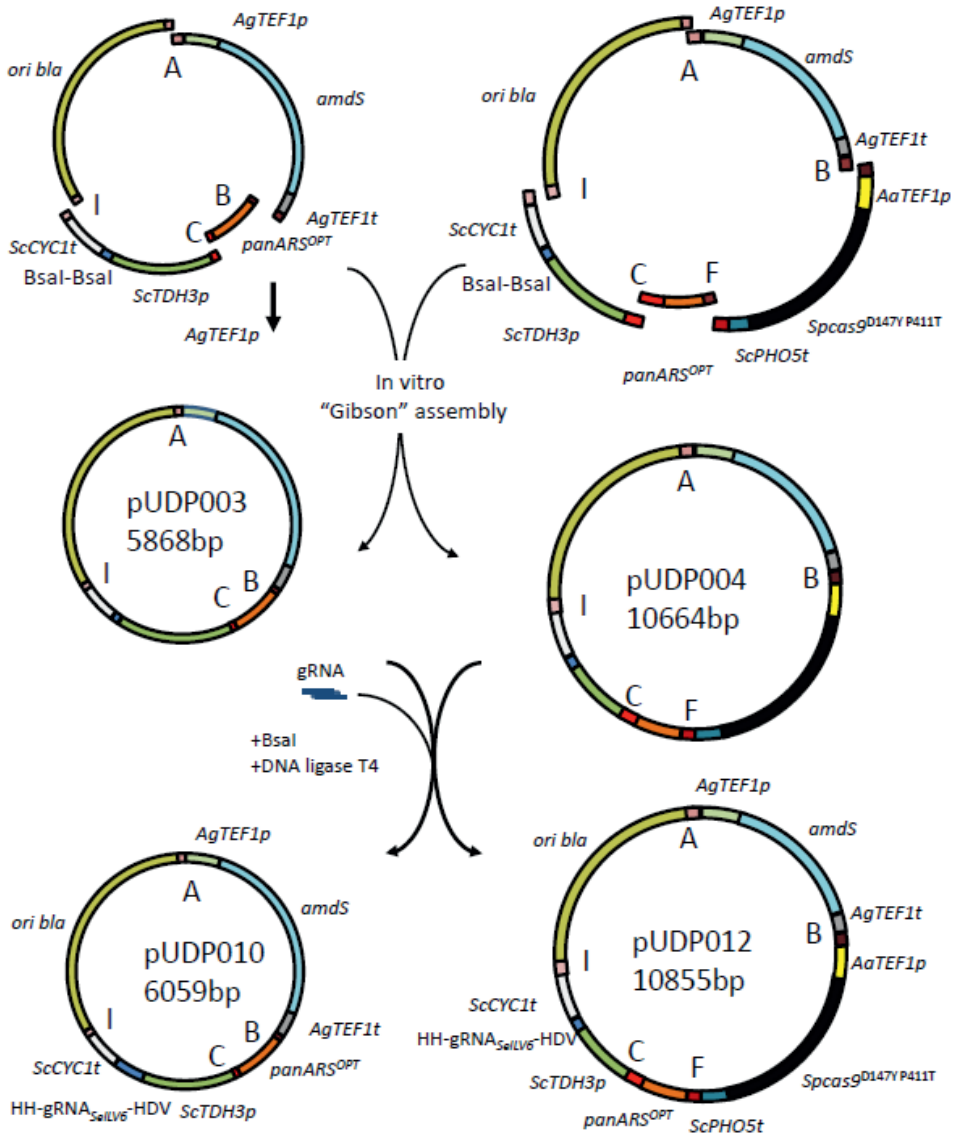
pUDP003 was assembled from plasmids pUD527, pUD530, pUD531 and pUD532 (Figure 2). The *amdS* selection cassette (34) was amplified from pUD527 using primers 3847 and 3276 containing SHF A and B flanks. The synthetic pangenomic yeast replication origin panARSopt (40) was amplified from pUD530 using primers 3841 and 3856 containing SHF B and C flanks. The gRNA introduction site was amplified from pUD531 using primer 3283 and 4068 containing SHF C and I flanks. The *E. coli* replication origin from pBR322 and the *bla* gene conferring resistance to  $\beta$ -lactam antibiotics were amplified from pUDP532 using primers 3274 and 3275 containing SHF I and A flanks. The amplified fragments were digested with DpnI, gel purified and quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). 0.2 pmol of each fragment were assembled into pUDP003 using NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs). The resulting plasmid pUDP003 was verified by restriction analysis using SspI.



**Table 2: Plasmids used throughout the study.** HRL and HRM indicate the left and right homology arms for integration on the *SPR3* locus, SHR stands for synthetic homologous recombination sequence and enzyme digestion sites are indicated in superscript.

Name	Relevant genotype	Source	Addgene ID# <sup>a</sup>
pCT	<i>ori amp'</i> ARS4 CEN6 LEU2 <i>AaTEF1p-Spcas9</i> <sup>D147Y P411T</sup> - <i>ScPHO5</i>	(36)	
pMEL15	<i>ori amp'</i> 2 $\mu$ m <i>natNT2 SNR52p-gRNA<sub>CANI1</sub>-SUP4t</i>	(15)	
pROS12	<i>ori amp'</i> 2 $\mu$ m <i>hphNT1 gRNA<sub>CANI2</sub>-gRNA<sub>ADE2</sub></i>	(15)	
pUC19	<i>ori amp'</i> LacZ multiple cloning site	(68)	
pUD423	<i>ori amp'</i> <i>AaTEF1p-Spcas9</i> <sup>D147Y P411T</sup> - <i>ScPHO5t</i>	GeneArt™	
pUD526	<i>ori amp'</i> <sup>NotI</sup> HRL <i>AaTEF1p-Spcas9</i> <sup>D147Y P411T</sup> - <i>ScPHO5t natNT2 HRM</i> <sup>NotI</sup>	GeneArt™	
pUD527	<i>ori kan'</i> SHRA <i>AaTEF1p-amdS-AgTEF1t</i> SHRB	GeneArt™	
pUD528	<i>ori kan'</i> SHRB <i>AaTEF1p-Spcas9</i> <sup>D147Y P411T</sup> - <i>PHO5t</i>	GeneArt™	
pUD530	<i>ori kan'</i> SHRB panARSopt SHRC	GeneArt™	
pUD531	<i>ori kan'</i> SHRC <i>TDH3p</i> <sup>Bsal</sup> <i>CYC1t</i> SHRI	GeneArt™	
pUD532	<i>ori kan'</i> SHRI <i>bla ori</i> SHRA	GeneArt™	
pUD536	<i>ori amp'</i> <sup>Bsal</sup> <i>gRNA<sub>SeiLV6</sub></i> <sup>Bsal</sup>	GeneArt™	
pUD573	<i>ori amp'</i> <sup>Bsal</sup> <i>gRNA<sub>SeiATF1</sub></i> <sup>Bsal</sup> <i>gRNA<sub>SeiLV6</sub></i> <sup>Bsal</sup>	GeneArt™	
pUDP003	<i>ori amp'</i> panARSopt <i>AgTEF1p-amdS-AgTEF1t</i> <i>TDH3p</i> <sup>Bsal</sup> <i>CYC1t</i>	This study	101164
pUDP004	<i>ori amp'</i> panARSopt <i>AgTEF1p-amdS-AgTEF1t</i> <i>TDH3p</i> <sup>Bsal</sup> <i>CYC1t AaTEF1p-Spcas9</i> <sup>D147Y P411T</sup> - <i>ScPHO5t</i>	This study	101165
pUDP010	<i>ori amp'</i> panARSopt <i>AgTEF1p-amdS-AgTEF1t</i> <i>TDH3p-HH-gRNA<sub>SeiLV6</sub>-HDV-CYC1t</i>	This study	101166
pUDP012	<i>ori amp'</i> panARSopt <i>AgTEF1p-amdS-AgTEF1t</i> <i>TDH3p-HH-gRNA<sub>SeiLV6</sub>-HDV-CYC1t AaTEF1p-Spcas9</i> <sup>D147Y P411T</sup> - <i>ScPHO5t</i>	This study	101167
pUDP044	<i>ori amp'</i> panARSopt <i>AgTEF1p-amdS-AgTEF1t</i> <i>TDH3p-HH-gRNA<sub>SeiATF1</sub>-HDV-CYC1t AaTEF1p-Spcas9</i> <sup>D147Y P411T</sup> - <i>ScPHO5t</i>	This study	101168
pUDR107	<i>ori amp'</i> 2 $\mu$ m <i>hphNT1 gRNA<sub>URA3</sub></i>	This study	

<sup>a</sup> <https://www.addgene.org/>



**Figure 2: Construction of the gRNA expression plasmids pUDP003 and pUDP004.** *In vitro* “Gibson” assembly (39) of functional parts containing an *amdS* selection marker cassette, a synthetic pangenomic yeast replication origin panARSOpt, an *E. coli* replication origin from pBR322 and the *bla* gene conferring resistance to  $\beta$ -lactam antibiotics and a gRNA expression cassette using 60 bp synthetic homologous recombination sequences into pUDP003 and with the addition of a fragment carrying a *Spcas9* expression part into pUDP004. The ribozymes flanked gRNA is next directionally inserted into pUDP003 or pUDP004 using Bsal digestion and ligation yielding the gRNA expressing plasmids pUDP010 and pUDP012 respectively.

**Table 3: Primers used in this study.** SHR sequences are shown in **bold**, gRNA sequences are shown in *italic* and digestion enzyme recognition sites are underlined.

Name	Sequence	Purpose
3274	TATTCACGTAGACGGATAGTATAGC	Amplification SHR I
3275	GTGCCATATGATGATCGGGGAATG	Amplification SHR A
3276	GTTGAACAATCTTAGGCTGTCGAATC	Amplification SHR B
3283	ACGTCACGATCGTATATGC	Amplification SHR C
3597	ATTAAAGGTTCTCGAGAGC	Amplification <i>natN72</i>
3750	GAGGCGTTAGTTGGCTAATGAG	Diagnostic primer
3841	CACCTTCGAGAGGACGATG	Amplification SHR B
3847	ACTATATGAAGGCATGGCTATGG	Amplification SHR A
3856	CTAGCGTCTCTCGCATAGTTC	Amplification SHR C
4068	GCCTACGGTCCCGAAGTATGC	Amplification SHR I
6005	GATCATTTTATCTTCTCACCTCGGGAAG	pROS12 backbone
7389	GGTTCTTAGACGTCAGGTGGC	pUC19 backbone
8076	GTTTAGCTCTATGGTCAAAATTTCCAGAAAAAGGGATCATAGAAAAAGAAATATGCTAAITGAAAAATAGATATGTACCATAAAGTAAAGTGCATGC	Repair DNA SeURA3
8077	TGTCCTGGTTCCGTTATACCGCATGCACCTTACTTATGGTACATATCTATTTTCAATTAGACATAATCTTTCTATGGATCCCTTTTTCTGGAGAAATTTT	Repair DNA SeURA3
8314	GCACATAGAGCTAAAC	Repair DNA SeURA3
8553	TGCCGATGTTTCGGCGTTCCGAACTTCTCCGAGTGAAGATAAATGATCTTGACTGATTTTTCATGGAGTTTTAGAGCTAGAAAATAGCAAGTTAAAAAT	pROS12 + SeURA3 gRNA
8554	AAG	
9310	TGCCAGTATCTTAACCAACTGCACAGAAAAACCTGCAGGAACGAAGATAAAATCAAAAACCTGATTATAAGTAAATGCATGTACTAAACTCAC	Repair DNA SeURA3
9311	AAATTAGAGCTTCAATTTAA	Repair DNA SeURA3
9312	TTAAATTGAAGCTAAATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGATTTTGATTTATCTCGTTTCTGCAGGTTTTGTTCTGTGCAGTTGG	Repair DNA SeURA3
9313	GTTAAGAATACTGGGCA	Repair DNA SeURA3
9314	TCGCCTGCAATCGTCATCG	Diagnostic primer ILV6
9315	CCTTAGAAACATCCGAGCTCCTCGGCCTTATACATC	Repair fragment construction
9316	GATGATAGAGCCAGGAGGAGCTCGGATGTTTCTAAGG	Repair fragment construction
9317	AGCTGGTCGCCAAGGACTAC	Diagnostic primer ILV6
9318	CTACTGGCCAAATTGATGAC	Diagnostic primer SeURA3
9319	GCCCTACACGTTCCGCTATGC	Diagnostic primer SeURA3
9320	GTTGACACAGTCCGTGAAAC	Diagnostic primer SeURA3
9321	AGGCAATTAAGATGACGATGACAAAC	Diagnostic primer SeURA3
9322	ATGGATTATAAGATGACGATGACAAAC	Amplification cas9
9323	CCGCTCAGACCTTTCCTTC	Amplification cas9
9324	TTTTGTATAACTAAATAATATTGGAAACTAAATAGC	Amplification ScPHO5t
9325	<b>TGCCAACTTCCCTTGTATGAAGCGATCTGACCAATCTCTTTGGCAGTATGTTTTCAATTTTGGCATGCCAG</b>	Amplification ScPHO5t + addition SHR F
9326	TGTTGATTGTTTTTAAAGAACTACTCAGAATG	Amplification AaTEF1p
9327	AGGCAGAAACCGTAAAAAG	pUC19 backbone
9328	ATTTCTTCTGATGATGTTCTTAAAAACATAATCAACAATGATTATAAAGATGACGATGACAAAACCTCCAAAAA	LCR bridging oligo
9329	TGACCTCCAAAAAGAAAGAAAGGCTGAGCGGTTTTGTATAACTAAATAATTTGGAAACTAAATAC	LCR bridging oligo

**Table 3 (continued): Primers used in this study.** SHR sequences are shown in **bold**, gRNA sequences are shown in *italic* and digestion enzyme recognition sites are underlined.

Name	Sequence	Purpose
9398	TTCATACAGGGAAAGTTCGGCAGGTTTCTTAGACGTCAGGTGGC	LCR bridging oligo
9399	CTTTTTACGGTTTCTGGCCTCACTTTCGAGAGGACGATG	LCR bridging oligo
9663	<b>CATACGTTGAAACTACGGCAAAAGGATTGGTCAGATCGCTTCATACAGGGAAAGTTCGGCA</b> TCAACATCTTTGGATAAATCAGAAATGAG	Amplification panARSOpt + addition of SHR F
10426	AATCTATAATCAAGTCACTAGTCAACAAGAGCC	Amplification AaTEF1p
10427	TTTTCATTTTTTGGATGCCAGTCTTTTG	Amplification ScPHO5t
10432	AAAAACGCCAAGCGGCCCTTTTACGGTTCCTGGCTGGGGCCGGCTCCAGGTTTGGCACTGTC	40 bp to pUC19 + NotI restriction + Left Homology arm fw
10433	ACTTTGAGGGCTTTGTTGACTATGACTGATTATAGATTTACGAAGGCACCTTTGCATGGG	Left Homology arm rv + 40 bp to AaTEF1p
10434	GACAAACCTGTTGTAATCGAGCTCTCGAAGACCCTTAATCGCGACATCAAATACCTTTGTCC	40 bp to <i>natN72</i> + right homology arm fw
10435	CACATTTCCCGAAAAGTGCCACTGALGCTCTAAGAAACC <b>GGGGCCGG</b> CACGCGGAGGAAGAAAG	Right homology arm rv + NotI restriction + 40 bp to pUC19
10436	ATAAGGAAACTCAAGAACTGGCATCGCAAAAATGAAAATAGGTCTAGAGATCTGTTTAGC	40 bp to ScPHO5t + <i>natN72</i> fw
10686	GAGTAAAGAAGCTCATCTTTATATAGATACGTTATGTAGATGTATAGAGGCCCAAGGGAGCTCGGATGTTCTAAGGCTCTGTATGTACAAACTAC	Repair DNA ILV6
10687	GTATGGACTTATACATTGCT	Repair DNA ILV6
10687	AATGATGAGCTCTTTACTC	Repair DNA ILV6
10992	GTTCAAGATGAATGCTTTGTCAAAGATGATACAGAATGGGCATTCCTCCGGCGTATGGGATCTTCATGGCATCAAGCTTTTTTCAATGGGTGTTTTCTTCGACTA	Repair DNA SeATF1
10993	AATGTTCAITTCCTTCACATTTAGTTCGAAGAAACCCCAATGAAAAAGCTTGATGCCATCCCATACGCCGGAATGCCCAATTCCTGTATCATCTTTG	Repair DNA SeATF1
10993	ACAAGACATTCATCTTGAAC	Repair DNA SeATF2
10994	TTTCTGTTTTTGCCTAGGCAGAAACATGTATTCGAATTTCCGCTGTTTATGGGAACTGAATAACGTTGGTGTATGAACATGGACATGACGCGTATGTTCCAG	Repair DNA SeATF2
10995	GGCACTCTACGGAATCGGGC	Repair DNA SeATF2
10995	GCCCCATTCCTGATAGTGCCTGAACTACGCTCATGTCCATGTTTCATACCACCAACGTTATTCAGTTCCCCATAAACAGCGAAATTCGAATACATGTTTT	Repair DNA SeATF2
10995	GCCTACCAAAACAGCAAA	Repair DNA SeATF2
10996	ATGAGAAAAATCAGGCCCCC	Diagnostic primer
10997	CTAAGGGCTTAAAAGGAGAGC	Diagnostic primer
10998	GAAGGATACGAACACCATATACG	Diagnostic primer
10999	TAAAGCGACGAAATTCGCC	Diagnostic primer
11000	CAGAAAGAAGCCAATTAGCAG	Diagnostic primer
11001	TCAGGGATTTAAAAGCAGAGC	Diagnostic primer
11002	GGATAGTTTAGAGGAATACGAACCG	Diagnostic primer
11003	TATACGAGACCCCGCAGC	Diagnostic primer

pUDP004 was assembled from plasmids pUD423, pUD527, pUD530, pUD531 and pUD532 (Figure 2). The *amdS* selection cassette (34) was amplified from pUD527 using primers 3847 and 3276 introducing SHF A and B flanks. The *cas9* expression cassette was amplified from pUD423 using primers 3841 and 9393 containing SHF B and F flanks. The synthetic pangenomic yeast replication origin panARSopt (40) was amplified from pUD530 using primers 9663 and 3856 containing the SHF C flank and introducing the SHF F flank, thereby replacing the SHF B flank. The gRNA introduction site was amplified from pUD531 using primer 3283 and 4068 containing SHF C and I flanks. The *E. coli* replication origin from pBR322 and the *bla* gene conferring resistance to  $\beta$ -lactam antibiotics were amplified from pUDP532 using primers 3274 and 3275 containing SHF I and A flanks. The amplified fragments were digested with DpnI, gel purified and quantified using a NanoDrop 2000 spectrophotometer (ThermoFischer Scientific). 0.2 pmol of each fragment were assembled into pUDP004 using NEBuilder<sup>®</sup> HiFi DNA Assembly Master Mix (New England BioLabs). The assembled plasmid pUDP004 was verified by restriction analysis using PdmI.

The gRNA sequences for pUDP type plasmids were designed such that they could be synthesized and inserted into pUDP003 or pUDP004 by digestion with BsaI and ligation. From 5' to 3', the sequences were composed of a BsaI recognition site yielding correct sticky ends "GGTCTCGCAA", followed by the hammerhead ribozyme with the first six nucleotides being the reverse complement (°) of the first six nucleotides of the gRNA spacer "°N<sub>6</sub>°N<sub>5</sub>°N<sub>4</sub>°N<sub>3</sub>°N<sub>2</sub>°N<sub>1</sub>CUGAUGAGUCCGUGAGGACGAAACGAGUAAGC UCGUC", followed by the 20 nucleotide gRNA spacer designed as previously (15), followed by the structural gRNA "GUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG CACCGAGUCGGUGCUUUU", followed by the Hepatitis Delta Virus ribozyme "GGCCGGCAUGGCUCCAGCCUCCUGCGCCGGCUGGGCAACAUGCUUCGGCAUGGCGAAUGGGAC", followed again by a BsaI recognition site yielding correct sticky ends "ACAGCGAGACC". For multiplexing, linker "ACAGCGCAA" was added between the HDV ribozyme of the first gRNA and the HH ribozyme of the second gRNA. Plasmids pUD536, containing the gRNA sequence targeting *SeLLV6*, and pUD573, containing a polycistronic array with gRNAs targeting *SeATF1* and *SeATF2*, were *de novo* synthesised at GeneArt (Thermo Fisher Scientific). The plasmid pUDP010, expressing gRNA<sub>*SeLLV6*</sub>, was constructed in a one-pot reaction by digesting pUDP003 and pUD536 using BsaI and ligating with T4 ligase. Similarly pUDP012, expressing gRNA<sub>*SeLLV6*</sub> and *Spcas9*<sup>D147Y P411T</sup>, was assembled from pUDP004 and pUD536. and pUDP044, expressing gRNA<sub>*SeATF1*</sub>::gRNA<sub>*SeATF2*</sub> and *Spcas9*<sup>D147Y P411T</sup> was assembled from pUDP004 and pUD573. Correct assembly of pUDP010 was verified by restriction analysis with SspI and correct assembly of pUDP012 and pUDP044 was verified by restriction analysis using PdmI. Plasmid pUDR107, expressing gRNA<sub>*URA3*</sub>, was constructed using NEBuilder<sup>®</sup> HiFi DNA Assembly Master Mix by assembling the 2  $\mu$ m fragment amplified from pROS12 with primer 8314 and the plasmid backbone amplified from pROS12 with primer 6005 as previously described in (15). Plasmids pUDP003, pUDP004, pUDP010, pUDP012 and pUDP044 were deposited at addgene (<http://www.addgene.org/>) (Table 1).

### Strain construction

The strain IMX1187 was constructed by transforming CBS 1483 with 1  $\mu$ g of the NotI-digested and gel purified integration cassette from pUD526 by electroporation and plated on YPD with nourseothricin (Figure 1). After 5 days, 14 colonies had grown and integration of *cas9* was confirmed using primers 3750 and 9394. One of the colonies was stocked and sequenced.

IMX1205 (Figure 1) was constructed by transforming IMX1187 by electroporation with 500 ng of pUDP010 and 1  $\mu$ g of a 120 bp repair fragment obtained by mixing an equimolar amount of primers 10686 and 10687. Transformants were selected on SM-Ac plates. Transformants were confirmed using primers 9310 and 9313. Prior stocking the isolate was successively streaked out on SM-Ac, YPD

and SM-FAC plates. Genotype was systematically verified after each plating round with primers 9310 and 9313. In the end, one of the colonies was stocked.

IMK771 (Figure 1) was constructed by transforming CBS 1483 by electroporation with 200 ng of pUDP012 and 1 µg of 120 bp repair product obtained by mixing an equimolar amount of primers 10686 and 10687. Transformants were selected on SM-Ac plates. Deletion of *SeLV6* was confirmed using primers 9310 and 9313. Prior stocking the isolate was successively streaked out on SM-Ac, YPD and SM-FAC plates. Genotype was systematically verified after each plating round with primers 9310 and 9313. In the end, one of the colonies was stocked and sequenced.

IMK786 (Figure 1) was constructed by transforming CBS 1483 by electroporation with 200 ng of pUDP044 and 1 µg of 120 bp repair product obtained by mixing an equimolar quantity of primers 10992 and 10993 for *SeATF1* and 1 µg of 120 bp repair product obtained by mixing an equimolar quantity of primers 10994 and 10995 for *SeATF2*. Transformants were selected on SM-Ac plates, deletion of *SeATF1* and *SeATF2* was confirmed using primer 11000/11001 and primers 11002/11003, respectively. Prior stocking the isolate was successively streaked out on SM-Ac, YPD and SM-FAC plates. Genotype was systematically verified after each plating round with primers pairs 11000/11001 and 11002/ 11003 to confirm *SeATF1* and *SeATF2* deletions. In the end, one of the colonies was stocked.

IMK813 (Figure 1) was constructed by transforming Weihenstephan 34/70 by electroporation with 200 ng of pUDP044 and 1 µg of 120 bp repair product obtained by mixing an equimolar quantity of primers 10992 and 10993 for *SeATF1* and 1 µg of 120 bp repair product obtained by mixing an equimolar quantity of primers 10994 and 10995 for *SeATF2*. Transformants were selected on SM-Ac plates, deletion of *SeATF1* and *SeATF2* was confirmed using primer 11000/ 11001 and primers 11002/11003, respectively.

### Next generation sequencing

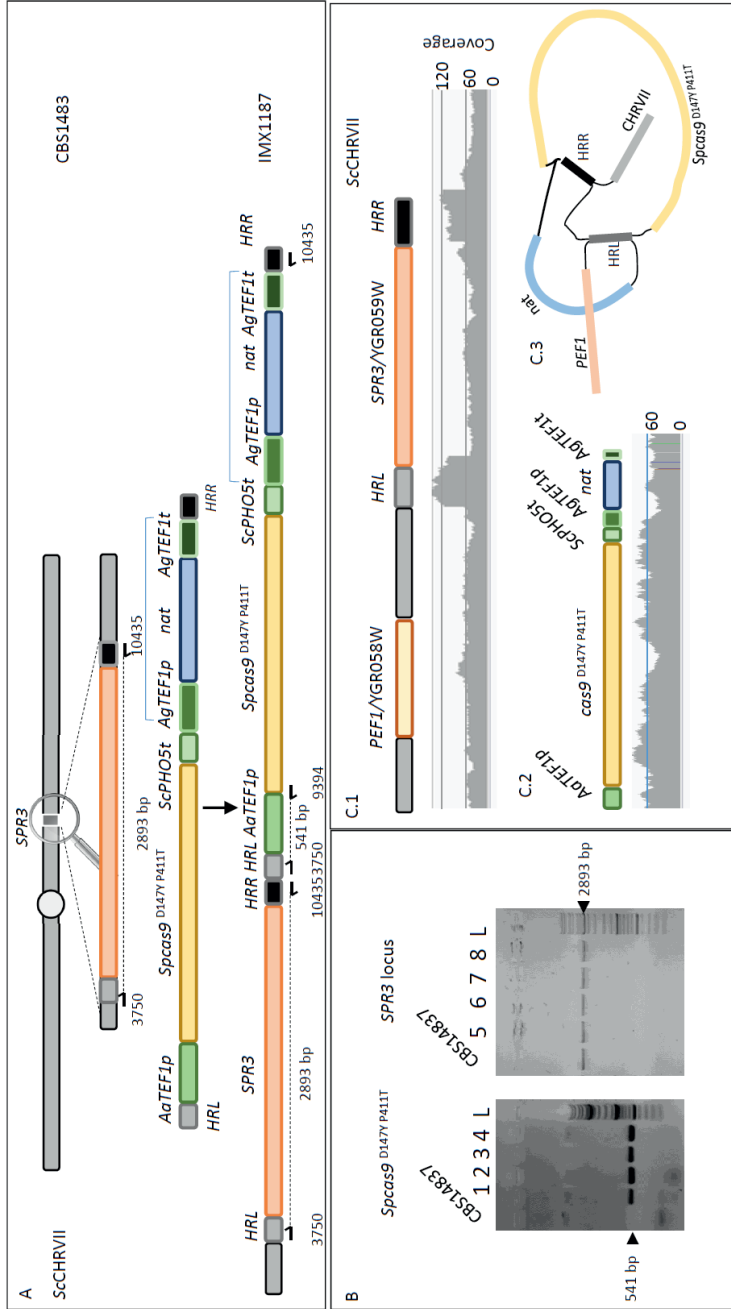
IMX1187 and IMK771 were incubated in 500-ml shake-flasks containing 100 ml liquid YPD medium at 20°C on an orbital shaker set at 200 rpm until the strains reached stationary phase with an OD<sub>660</sub> between 12 and 20. Genomic DNA for whole genome sequencing was isolated using the Qiagen 100/G kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and quantified using a Qubit® Fluorometer 2.0 (ThermoFisher Scientific). 51.57 µg of genomic DNA from IMX1187 and 14.20 µg from IMK771 was sequenced by Novogene Bioinformatics Technology Co., Ltd (Yuen Long, Hong Kong) on a HiSeq 2500 (Illumina, San Diego, CA) with 150 bp paired-end reads using True-seq PCR-free library preparation (Illumina). CRISPR-Cas9 assisted deletions were verified by mapping the sequencing reads onto the *S.pastorianus* CBS 1483 genome (4) using the Burrows–Wheeler Alignment tool (BWA) and further processed using SAMtools (41, 42). The deletions were confirmed by visualising the generated .bam files in the Integrative Genomics Viewer (IGV) software (43). The sequencing data are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the Bioproject PRJNA397648.

## Results

### Construction of a *S. pastorianus* strain expressing *cas9*

To limit construct instability and facilitate successive genome editing events, a copy of the *Streptococcus pyogenes cas9* variant, *cas9*<sup>D147Y P411T</sup> (36) was integrated in the genome of *S. pastorianus* CBS 1483. The *S. cerevisiae* *SPR3/YGR059W* locus is involved in sporulation: a function

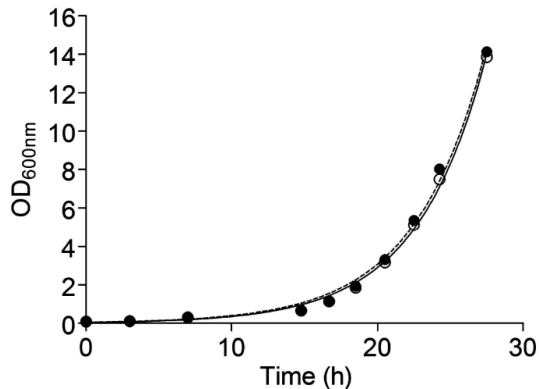
impaired in *S. pastorianus*; therefore it was chosen as integration site as the impact on growth of deletion of *SPR3* should be negligible. Additionally, *SPR3* is located in the middle of the right arm of the *S. cerevisiae* CHR VII which counts only one copy in CBS 1483, which should enable stable integration of a single *cas9* copy (4). To prevent off-target integration driven by homology of the promoter and terminator, *cas9* was placed under the control of the *TEF1* promoter from *Arxula adeninivorans*, which had been shown to be functional in *Saccharomyces* yeast (44). The nourseothricin acetyl transferase expression cassette natNT2 expressed from the *TEF1* promoter from the yeast *Ashbya gossypii* was used as a marker to select for integration (45) (Figure 3). To guide the chromosomal integration of the endonuclease construct, the *cas9* containing fragment was flanked with an homology region of 480-bp targeting the *SPR3* promoter region (HRL, Figure 3A) and a 506-bp targeting the *SPR3* terminator region (HRM) to complete the double cross over integration (Figure 3A). These elements were assembled into a transformation cassette on pUD526 and the purified integration fragment was used to transform *S. pastorianus* CBS 1483 yielding 14 transformants. In comparison, the same transformation in the laboratory *S. cerevisiae* CEN.PK113-7D yielded 476 transformants. Both transformations were performed simultaneously and under identical experimental conditions, therefore the difference in obtained transformants reflected the strong resilience of industrial *S. pastorianus* strains to transformation. The presence of the integrated construct was confirmed in all four tested colonies by PCR using specific primers (3750 and 9394) which amplify between the left homology arm for *SPR3* and the end of the *AaTEF1* promoter. Unexpectedly, a PCR targeting the *SPR3* open reading frame using primers 3750 and 10435 yielded a fragment size corresponding to the wild type. Concomitantly, PCRs targeting *cas9* confirmed the integration in CHR VII in all 4 tested transformants, suggesting that either *SPR3* might have been duplicated prior to replacement of one of the copies by *cas9*, or the cassette was not integrated as intended. To resolve the recombined *SPR3* locus map, one of the transformants, was renamed IMX1187 and re-sequenced using Illumina technology. Mapping of the IMX1187 Illumina pair reads (2X 150bp) on the CBS 1483 reference genome sequence confirmed the presence of the *S. cerevisiae* *SPR3* wild type locus, but it also revealed that the region used for the integration HRL and HRM, exhibited a sequence depth coverage two-fold higher than the *SPR3* open reading frame and the surrounding chromosomal region (Figure 3C). In the meantime, mapping of the IMX1187 reads on the sequence of the deletion cassette including the *cas9* and *nat* genes confirmed the single integration of the transformed fragment. Additionally, absence of reads mapping the  $\beta$ -lactamase gene *bla* present on pUD526 excluded the possibility that the plasmid got mistakenly integrated in the genome. To demonstrate anchoring of the cassette into CHR VII, the reads that mapped to the *SPR3* region and to the integration fragment containing *cas9* and *nat* (including corresponding paired reads) were extracted and assembled using SPAdes (46). The assembly confirmed that the cassette was anchored in CHR VII and the obtained graph suggested that the *cas9/nat* cassette integrated by single crossover resulting in a duplication of the integration site HMR or HML and integration of the *cas9* cassette (Figure 3). However, the integration cassette was fully integrated and should result in expression of Cas9.



**Figure 3: Integration of *Spcas9*<sup>D147Y P411T</sup> at the *SPR3* locus in *S. pastorianus* CBS 1483. A- Schematic representation of the integration of *Spcas9*<sup>D147Y, P411T</sup> (36) and the *nat* selection marker. The integration is directed by homology regions of 480-bp (HRL) and 506-bp (HRM) to complete the double cross over integration. B- Verification of the construction of strain IMX1187. Presence of the integration fragment carrying *Spcas9* and *nat* genes and of the *SPR3* open reading frame and was checked in four transformants with primers 3750 and 9394 (lane 1 to 4) and with primers 3750 and 10435 (lane 5 to 8) respectively. The strain host strain CBS 1483 was used as reference. The transformant in lanes 1 and 5 was renamed IMX1187. The lane labelled with L designated the position of the DNA ladder (Gene ruler DNA ladder Mix (ThermoFischer Scientific #SM0332)). C- 1-Mapping of the 150bp Illumina sequencing reads of IMX1187 onto the reference genome of CBS 1483 (4) at the *SPR3* locus reveals about 120-fold coverage of the homology regions HRL and HRM while the average coverage is about 60-fold. 2- Mapping of the IMX1187 150bp Illumina sequencing reads onto the integrated fragment containing the *Spcas9* and *nat* genes reveals about 60-fold coverage of the cassette. 3- Assembly graph of IMX1187 mapping on *SPR3*, *Spcas9* and their paired reads using SPAdes (46).**



In literature, there are conflicting reports about the physiological consequences of Cas9 expression in *Saccharomyces cerevisiae*, depending on the mode and tuning of expression of the endonuclease gene (15, 19, 47). Therefore, the growth rates of the *S. pastorianus* CBS 1483 and IMX1187 (*AaTEF1-cas9*) were measured in YPD at 20°C. The average maximum specific growth rate derived for biological triplicates for both strains did not deviate more than 2 %. The strains CBS 1483 and IMX1187 exhibited growth rate of  $0.263 \pm 0.002 \text{ h}^{-1}$  and  $0.258 \pm 0.001 \text{ h}^{-1}$  respectively (Figure 4). This result confirmed that single integration of *cas9* in CBS 1483 (IMX1187) did not significantly affect the maximum specific growth rate.

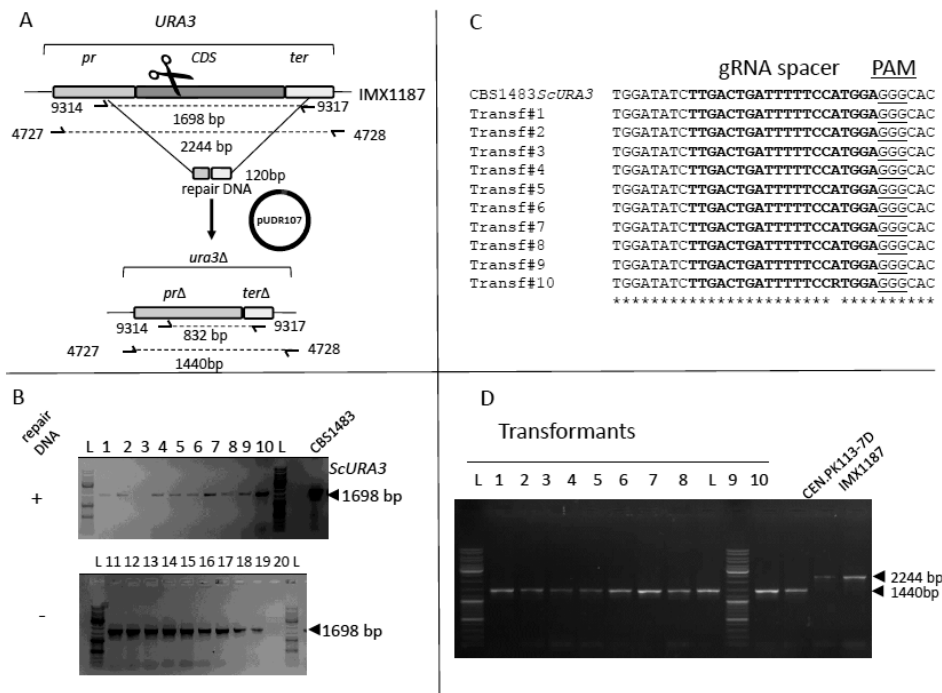


**Figure 4:** Growth curve of the *S. pastorianus* strains CBS 1483 (●) and IMX1187 (*SpCas9*) (○). The strains CBS 1483 and IMX1187 were strains were grown in complex medium (YPD) at 20°C. Growth was monitored based on optical density at 660 nm (OD660) measurements. The data plotted are average and standard deviation of three biological replicates.

#### gRNA delivery systems for efficient editing in *S. pastorianus*

After establishing the chromosomal integration of *cas9* in the genome of CBS 1483, the next step consisted in demonstrating the activity of the RNA-programmed endonuclease. To do so, two gRNA delivery systems were tested, one based on the existing RNA polymerase III dependent system developed for *S. cerevisiae* (15) and one expression system based of ribozyme flanked gRNA expressed from a RNA polymerase II promoter. Firstly, the deletion of *URA3* using the traditional RNA polymerase III system was tested in *S. pastorianus* strain IMX1187 (*AaTEF1p-cas9*). The selected 20-bp spacer to target *URA3* matched the *ScURA3* allele sequence perfectly (TTGACTGATTTTCCATGGA), but carried one mismatch on the 12<sup>th</sup> position from its 3' end (TTGACTGACTTTTCCATGGA) compared to the *S. eubayanus* allele (*SeURA3*). Both alleles shared the same gRNA spacer adjacent motif (PAM) sequence (GGG) and CBS 1483 harbored three *S. cerevisiae* and two *S. eubayanus* alleles. The gRNA<sub>URA3</sub> was expressed by the RNA polymerase III dependent promoter *SNR52p* (16, 48) from the pROS12 plasmid, which carries a hygromycin resistance marker *hph* (15). The resulting plasmid pUDR107 (gRNA<sub>URA3</sub>) was transformed in IMX1187 alone or together with two 120 bp double stranded repair DNA fragments for *ScURA3* and *SeURA3*. In absence of repair DNA, the transformation of the *URA3* gRNA should in theory be lethal and yield few to no transformants, due to the inefficiency of Non-Homologous End Joining (NHEJ) (Figure 5A). However, the transformation of IMX1187 with pUDR107 alone returned several hundred of colonies, a number comparable to when the repair DNA was also provided. A set of ten clones from each transformation were picked and their genotype was diagnosed by specific PCR (9314 and 9317 for *ScURA3* and 9318 and 9321 for

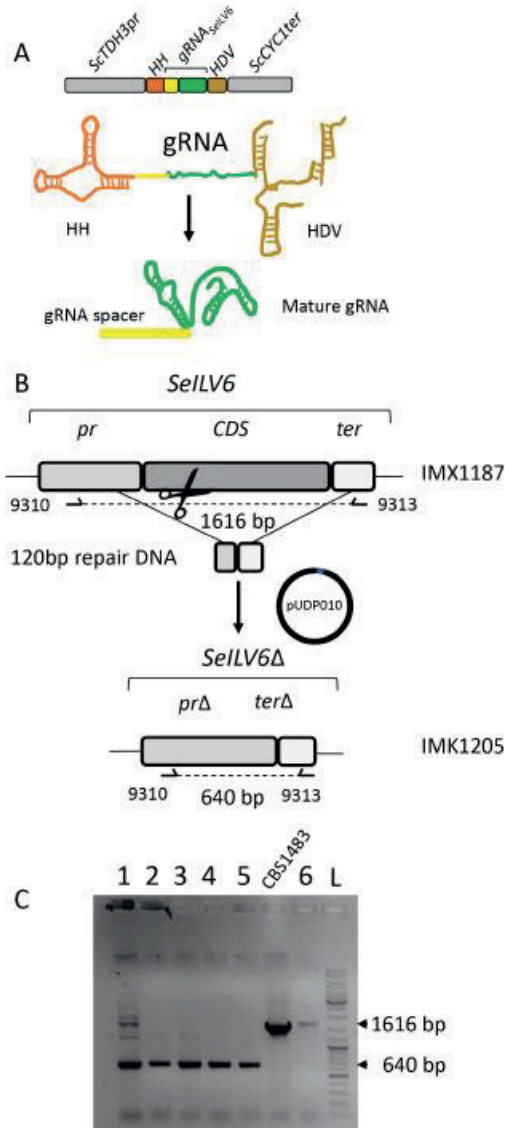
*ScURA3*). All transformants either derived from the transformation with or without supply of a repair DNA produced a band with a size compatible with the wild type allele (Figure 5B). The Sanger sequencing results of the amplified fragments showed no indels at the site of the anticipated cut. With the exception of clone #10 that showed an unresolved purine (R), all *URA3* sequences were identical to that of the reference IMX1187, confirming the absence of editing (Figure 5C). Therefore, to exclude defective expression of the gRNA, pUDR107 (gRNA<sub>URA3</sub>) was also transformed in *S. cerevisiae* IMX585 (*cas9*) (15) together with the *ScURA3* 120bp repair DNA. Out of the couple of dozens transformants, ten were randomly picked and diagnosed with by PCR. All transformants exhibited a band at 1440bp characteristic of the *URA3* deletion. The same PCR from the untransformed CEN.PK113-7D yielded a fragment of 2244 bp (Figure 5D). This result established that pUDR107 enabled functional Cas9-mediated gene editing in *S. cerevisiae* IMX585, but not in *S. pastorianus* IMX1187.



**Figure 5: Deletion of *ScURA3* in IMX1187 and IMX585 using RNA III polymerase dependent (*SNR52p*) gRNA expression.** A-representation of the native and deleted *ScURA3*. The plasmid pUDR107 carried a gRNA under the control of the *SNR52p*. Primers used for validation of the deletion are indicated. B- Validation of transformants of the *S. pastorianus* IMX1187 strain with pUDR107 in presence or not of a 120 bp repair DNA. The PCR reactions were performed with the primers 9314 and 4728. All lanes (1 to 20) showed a PCR product of 1698 bp corresponding to the wildtype allele. The lane labelled with L designated the position of the DNA ladder (Gene ruler DNA ladder Mix (ThermoFischer Scientific #SM0332)). C- Sanger sequencing results of purified PCR fragments of ten transformants derived from the transformation of IMX1187 with pUD107 (gRNA<sub>URA3</sub>). The gRNA spacer used to direct Cas9 is indicated in bold and the PAM sequence is underlined. D- Validation of transformants of the *S. cerevisiae* IMX585 strain with pUDR107 in presence or not of a 120 bp repair DNA. The PCR reactions were performed with the primers 4727 and 4728. The lanes (1 to 10) corresponding to transformants obtained with repair DNA showed a PCR product of 1440 bp corresponding to the deleted allele. The control lane labelled CEN.PK113-7D showed the wild type fragment at 2244bp. The lane labelled with L designated the position of the DNA ladder.

While gRNA transcript level was not measured, RNA polymerase III expression is known to be low (49), a level which might be insufficient to enable efficient Cas9-mediated introduction of a DSB". To circumvent this and to ensure high expression of the Cas9 programming RNA. In this approach, the gRNA was placed behind the control of the constitutive *ScTDH3* promoter. To prevent modifications inherent to RNA polymerase II transcribed RNA, the gRNA was flanked by a Hammerhead ribozyme (HH) and a Hepatitis Delta Virus ribozyme (HDV) on its 5' and 3' end respectively [21] (Figure 6A). After transcription and self-cleavage of both ribozymes, high transcript levels of mature gRNA should be possible. Such an expression system was constructed, resulting in plasmid pUDP003, which harbored the *S. cerevisiae* codon optimized *Aspergillus nidulans* acetamidase gene (*amdS*) (34) and enabled insertion of a specific gRNA. This strategy was tested by attempting deletion of the *SeILV6* gene in IMX1187 (*AaTEF1p-cas9*). The *S. pastorianus* strain CBS 1483 and IMX1187 harbored only one *ILV6* gene that originates from the *S. eubayanus* sub-genome (4). The *SeILV6* gene is located on the SeCHRIII, a chromosome present in four copies (4). The gRNA<sub>*SeILV6*</sub> was inserted in plasmid pUDP003 (Figure 2), resulting in plasmid pUDP010 (HH-gRNA<sub>*SeILV6*</sub>-HDV *amdS*). Despite the absence of a *S. cerevisiae* *ILV6* allele in IMX1187, the gRNA<sub>*SeILV6*</sub> was designed to target *ILV6* in *S. cerevisiae* as well. Thus, prior testing pUDP010 in *S. pastorianus*, the plasmid was transformed in *S. cerevisiae* IMX585. In the absence of a repair fragment, only 10 transformants were obtained while more than 500 were obtained when the repair fragment was included. Eventually a diagnostic PCR using specific primers confirmed successful deletion of *ILV6* in IMX585 for all tested colonies. Similarly, transformation of pUDP010 (HH-gRNA<sub>*SeILV6*</sub>-HDV *amdS*) in *S. pastorianus* IMX1187 (*AaTEF1p-cas9*) yielded 18 transformants when a 120bp repair fragment was co-transformed against just one when the repair fragment was omitted. Diagnostic PCR using primers 9310 and 9313 confirmed successful deletion of *SeILV6* in IMX1187 for all tested colonies (Figure 6C). It should be noted that the absence of bands of original size confirmed that all four copies of *SeILV6* were deleted. The PCR characterization of the unique transformant obtained in absence of repair DNA indicated that the *ILV6* locus was not deleted, since a band with a size compatible with the reference length was amplified, suggesting that the CRISPR-Cas9 induced DSB was repaired by NHEJ (Figure 6C).

The ability to obtain successful deletion of *ILV6* using the pUDP expression system indicated effective expression of the integrated *cas9* in *S. pastorianus* IMX1187, despite its imperfect integration in the *SPR3* locus. The failure to obtain deletion of *URA3* using the RNA Polymerase III dependent gRNA expression system in *S. pastorianus* IMX1187 while deletion was possible in *S. cerevisiae* IMX585 indicated that this gRNA expression system was not effective in *S. pastorianus*. Based on literature, this ineffectiveness may be caused by low gRNA transcripts levels. Regardless, the new pUDP expression system was functional in *S. pastorianus* and the deletion of *ILV6* constituted the first reported successful use of Cas9 engineering in *S. pastorianus*.



**Figure 6: Ribozymes flanked gRNA driven deletion of *SelLV6* in *S. pastorianus* IMX1187.** A- Representation of the gRNA expression cassette in pUDP010. The *gRNA<sub>SelLV6</sub>* was flanked on its 5' by a hammerhead ribozyme (HH represented in orange) and on its 3' by a hepatitis delta virus ribozyme (HDV represented in bronze). This construct was under the control of the RNA polymerase II promoter *SctDH3* and the *ScCYC1* terminator. Upon ribozyme self-cleavage, a mature gRNA comprising the *SelLV6* guiding spacer (in yellow) and the constant structural gRNA fragment (in green) is released. B-Schematic representation of the *SelLV6* editing upon transformation of IMX1187 with pUDP010. The primers for the validation of transformants are indicated. C- Validation of transformants of the *S. pastorianus* IMX1187 strain with pUDP010 in presence of a 120 bp repair DNA. The lanes (1 to 5) corresponding to the transformants obtained with repair DNA showed a PCR product of 640 bp corresponding to the deleted allele. One of the transformants exhibiting an *SelLV6* deletion was renamed IMK1205. The control lane labelled CBS 1483 and lane 6 corresponding to one transformant obtained without repair DNA showed a PCR product corresponding to the wild type fragment at 1616 bp. The lane labelled with L designated the position of the DNA ladder (Gene ruler DNA ladder Mix (ThermoFischer Scientific #SM0332)).

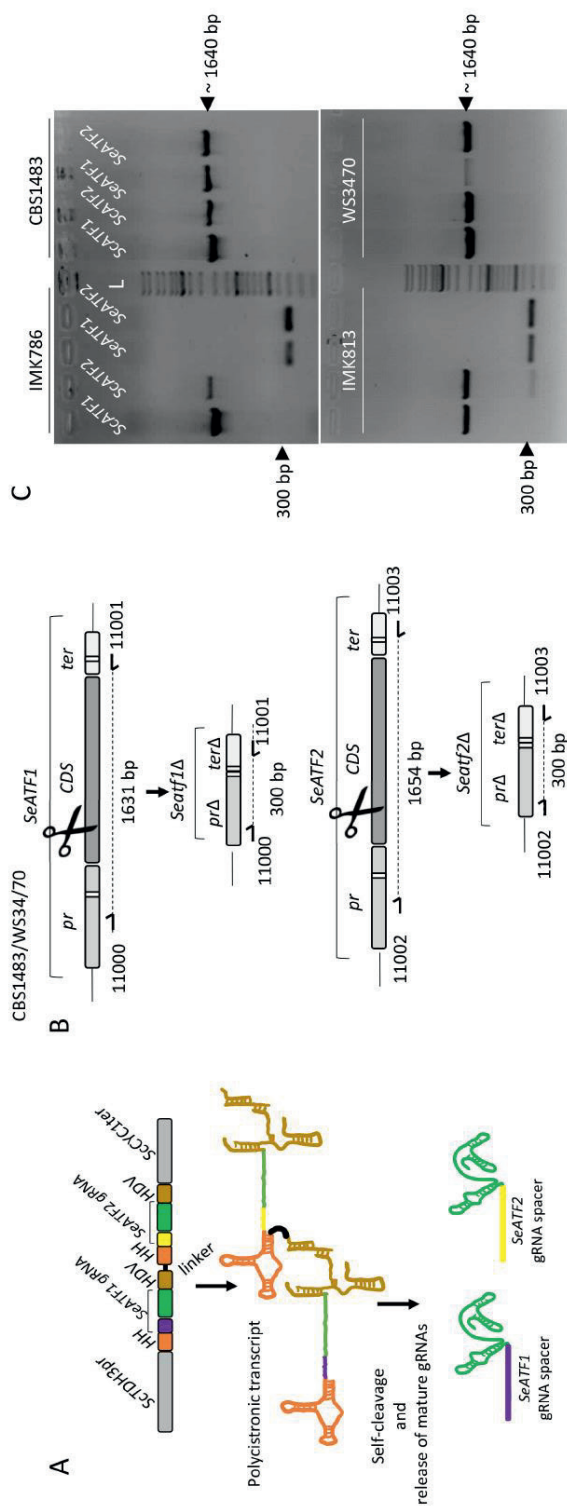
### Plasmid-based co-expression of Cas9 and gRNA in *S. pastorianus*

Given the notoriously low efficiency of gene insertion by homologous recombination in the genome of *S. pastorianus*, a plasmid was designed for co-expression of *cas9* together with the gRNA, which would render *cas9* expression more reproducible and facilitate genome editing in different *S. pastorianus* strains. The plasmid pUDP004 combined the *cas9* expression cassette previously integrated in IMX1187 and the different elements of pUDP003 including the RNA Polymerase II dependent gRNA expression cassette (Figure 2). To assess the efficacy of the pUDP004 system relative to the chromosome borne *cas9* together with the pUDP003 system, gRNA<sub>SeILV6</sub> was inserted in pUDP004 and the resulting plasmid pUDP012 was used to transform CBS 1483. In absence of a 120-bp repair DNA, a total of 14 transformants were obtained, while the number of transformants increased by 63-fold reaching a total of 884 transformants when the repair fragment was co-transformed. Diagnostic PCR using primers 9310 and 9313 confirmed successful deletion of *SeILV6* in for all tested colonies and one colony producing a fragment corresponding to effective deletion of *SeILV6* was stocked as IMK771. To eliminate any doubt, the IMK771 genome was resequenced using Illumina sequencing technology. The 150-bp pair-end reads were mapped on the CBS 1483 reference genome sequence (4) and as expected no reads mapped to the region targeted for deletion, indicating complete deletion of all four alleles of *SeILV6*. These results demonstrated that the plasmid-based co-expression of *cas9* and a gRNA was functional and could be used for effective genome editing in *S. pastorianus*.

### Multiplexing gene targeting by expression of double ribozyme flanked gRNAs array

Despite the preexisting good genetic accessibility of *S. cerevisiae* strains, CRISPR-Cas9 mediated editing greatly simplified genome engineering approaches. In particular, the ability to multiplex editing events (15, 18, 50). Therefore, the possibility of multiplexed gRNA expression was investigated in the pUDP expression system. Conveniently, the self-cleaving properties of the ribozymes might be compatible with the construction of adjacent HH-gRNA-HDV linked in a polycistronic array.

Encouraged by the successful *SeILV6* deletion using pUDP004 based gRNA expression, a tandem array of [HH-gRNA-HDV] targeting *SeATF1* and *SeATF2* in *S. pastorianus* was designed. The two HH-gRNA-HDV were spaced with a 10-bp linker. The synthesized array was placed under the control of the *ScTDH3* promoter in pUDP004 as described earlier for the *SeILV6* gene. The recombinant plasmid pUDP044 (*amdS cas9 TDH3p*-HH-gRNA<sub>SeATF1</sub>-HDV-HH-gRNA<sub>SeATF2</sub>-HDV-CYC1t) was then used to transform two *S. pastorianus* strains: CBS 1483 and Weihenstephan 34/70. (Figure 7A). CBS 1483 harboured one and three copies of *SeATF1* and *SeATF2* respectively, while Weihenstephan 34/70 missed one *SeATF2* allele relative to CBS 1483. Co-transformation of CBS 1483 and Weihenstephan 34/70 with pUDP044 and the corresponding repair fragments yielded 43 and 189 transformants per plate respectively. In the absence of repair fragments, 15 and 44 colonies were obtained in CBS 1483 and Weihenstephan 34/70, respectively. A randomly picked set of seven colonies transformed with repair fragment were verified by PCR, which confirmed that all copies of *SeATF1* and *SeATF2* were deleted. One of the CBS 1483 transformants exhibiting the correct double *SeATF1/SeATF2* deletion was named IMK786 and similarly a Weihenstephan transformant was named IMK813 (Figure 7). The designed gRNAs were also confirmed to be specific to the *S. eubayanus* genes as the *ScATF1* and *ScATF2* genes were not affected (Figure 7C). To the best of our knowledge, this represents the first application of polycistronic ribozyme flanked gRNA, as well as the first demonstration of a successful double deletion in *S. pastorianus*.



**Figure 7: Simultaneous deletion of all *SeATF1* and *SeATF2* alleles using a single ribozymes flanked gRNA array in *S. pastorianus* CBS 1483 and Weihenstephan 3470 (WS3470).** A- Representation of the gRNA array expression cassette in pUDP044. The dual gRNA array was under the control of the RNA polymerase II promoter *ScTDH3* and *ScCYC1* terminator. Each gRNA was flanked on its 5' by a hammerhead ribozyme (HH represented in orange) and on its 3' by a hepatitis delta virus (HDV represented in bronze) ribozyme and they were separated by a 10 bp linker. Upon ribozyme self-cleavage, the mature gRNAs are released. The *SeATF1* guiding spacer (in purple), the *SeATF2* guiding spacer (in yellow) and the constant structural gRNA fragment (in green) are indicated. B-Schematic representation of the *SeATF1* and *SeATF2* editing upon transformation of CBS 1483 with pUDP044. The primers for the validation of transformants are indicated. C- Validation of transformants of the *S. pastorianus* CBS 1483 strain with pUDP044 in presence of a 120 bp repair DNA. The PCR reactions were performed with the primers pairs 11000/ 11001 for *SeATF1* and 11002/11003 for *SeATF2*. The isolate renamed IMK786 exhibited bands at 300bp corresponding to the deletions of *SeATF1* and *SeATF2*. *SeATF1* and *SeATF2* were amplified using the primer pairs 10996/10997 and 10998/10999 respectively and exhibited wild type length. Similarly, transformants resulting from the transformation of pUDP044 in presence of a 120 bp repair DNA were checked with the primers pairs 11000/ 11001 for *SeATF1* and 11002/11003 for *SeATF2*. The isolate renamed IMK813 exhibited bands at 300 bp corresponding to the deletions of *SeATF1* and *SeATF2*.

## Discussion

### ***S. pastorianus* is not genetically amenable**

The results reported in this study firmly established that CRISPR-Cas9 improves the performance of homology-directed recombination in *S. pastorianus*. In contrast to *S. cerevisiae*, a species amenable to genetic modification, the interspecific hybrid *S. pastorianus* has shown higher resilience to targeted genetic alterations. This was exemplified by the attempt to integrate the *cas9* gene at a specific chromosomal site using traditional double cross over. The size of the cassette complicated the genotype characterisation, but the presence of the endonuclease gene was confirmed and although whole genome resequencing of the strain IMX1187 did not completely resolve the structure of the recombined locus, it strongly suggested that a single crossover integration event occurred, resulting in integration of *cas9* next to *SPR3* instead of replacing *SPR3* as intended. Several literature reports corroborated our unfortunate experience (7, 8). In different microbial systems, the efficiency of integration by homologous recombination was improved by impairing the Non-Homologous End-Joining (NHEJ) function (51-53). This approach, though successful, was often accompanied by side effects such as an exacerbated sensitivity to environmental stresses. In *S. cerevisiae*, inactivation of Ku70 and Ku80, two proteins involved in NHEJ, resulted in severe alterations of telomere maintenance and function as well as in deregulation of the cell cycle (54-58), which might explain why this strategy has never been attempted in *S. pastorianus*. Furthermore, the absence of improvement of the *S. pastorianus* genetic accessibility is not so surprising after all, since the brewing industry as most industries involved in fermentation of products intended for human consumption, has been reluctant to apply genetically modified organisms by fear of consumers group opinion (59), and has privileged classical strain improvement programmes.

Eventually, the results reported in this study demonstrated that the introduction of a DSB, which stimulates occurrence of homologous recombination, would represent an efficient solution to circumvent the natural resilience to targeted genetic modification in *S. pastorianus*.

### **gRNA expression in *S. pastorianus***

Editing systems developed for *S. cerevisiae* could not be directly transferred to *S. pastorianus*. Although convoluted, the functionality of Cas9 in *S. pastorianus* was eventually demonstrated. In contrast to the situation in *S. cerevisiae*, the expression of the gRNA from the *SNR52* promoter was unsuccessful. While the objectives of the study were not to fully understand the origin of the lack of functionality of the *SNR52* driven gRNA expression, we could hypothesize that this problem might arise from the hybrid genome composition of *S. pastorianus*. Their alloaneuploid genome is a source of genetic innovations, e. g. increased chromosome copy number has facilitated introduction of allelic variations and cohabitation of the two parental genomes might have stimulated the adjustment of transcription circuits which together have contributed to adaptation of lager yeast to the intensified brewing environment (4, 60, 61). Furthermore, many cellular functions are controlled by protein complexes which in hybrid strains may be formed by assemblies of subunits originating from both parental sub-genomes, thereby creating another source of variation (62). The RNA polymerase III is a complex formed of six different subunits (*TFC1*, 3, 4, 6, 7 and 8) and the strain CBS 1483 retained both parental gene sets (4). Thus, the absence of editing might reflect a modification of the RNA polymerase III transcriptional control in *S. pastorianus* relative to *S. cerevisiae*. This could also be associated with promoter sequence variations between the parents and the hybrid. The inspection of the *SNR52* promoter sequences of the *S. cerevisiae* and *S. eubayanus* parents revealed nucleotide variations with *S. pastorianus* promoters (Figure 8). The

*ScSNR52* promoter from CBS 1483 carried one mutation in position -4 (G to A), while the CBS 1483 *SeSNR52* promoter exhibited four single nucleotide variations with two located between the positions -1 and -100. In all configurations, the absence of editing points towards too low gRNA expression.

```

ScSNR52p CBS1483 -----TCTTTGAAAAGATAATGTATGATTATGCTTTCACCTCATAT
ScSNR52p S288C -----TCTTTGAAAAGATAATGTATGATTATGCTTTCACCTCATAT
ScSNR52p pROS12 -----TCTTTGAAAAGATAATGTATGATTATGCTTTCACCTCATAT
                        *****

ScSNR52p CBS1483 TTATACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGA
ScSNR52p S288C TTATACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGA
ScSNR52p pROS12 TTATACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGA
                        *****

ScSNR52p CBS1483 AGTACAACCTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAGGTGCGCATTTTTTTC
ScSNR52p S288C AGTACAACCTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAGGTGCGCATTTTTTTC
ScSNR52p pROS12 AGTACAACCTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAGGTGCGCATTTTTTTC
                        *****

ScSNR52p CBS1483 ACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTGGT
ScSNR52p S288C ACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTGGT
ScSNR52p pROS12 ACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTGGT
                        *****

ScSNR52p CBS1483 GCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATAATC-----
ScSNR52p S288C GCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATC-----
ScSNR52p pROS12 GCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATC-----
                        *****

```

**Figure 8: Sequence alignment of *ScSNR52* promoters derived from *S. cerevisiae* S288C (GenBank ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)) accession: NC\_001137), pROS12 (Euroscarf (<http://www.euroscarf.de/>) accession: P30789) (15) and *S. pastorianus* CBS 1483 (Bioproject ([www.ncbi.nlm.nih.gov/bioproject/](http://www.ncbi.nlm.nih.gov/bioproject/)) accession: PRJNA266750 (4)).** The sequences aligned using Clustal W (69) with gap open penalty and gap extension penalty parameters set to 10 and 0.05 in the multiple alignment mode.

Fortunately, the proposed alternative involving expression of a ribozyme protected gRNA system turned out to be successful. In this method already used in human cells (63), plants (21, 32) and fungi (64, 65) the gRNA construct is expressed from a RNA polymerase II promoter. All CRISPR-Cas9 assisted deletions attempted (*SeILV6*, *SeATF1* and *SeATF2*) were introduced with high fidelity. The challenging nature of genetic modification in *S. pastorianus* does not come only from the low efficiency of homologous recombination, but also from the requirement to delete multiple alleles simultaneously due to its extensive aneuploidy (61). In the case of *SeILV6*, four alleles were simultaneously deleted without introduction of any markers at the loci. The fact that all alleles were deleted at once as intended demonstrates the potency of a CRISPR-Cas9 induced DSB to stimulate targeted homology-mediated integration and circumvent unreliability of recombination in *S. pastorianus*. Remarkably, this could be achieved simultaneously at two different chromosomal loci (*SeATF1* on CHR SeVIII-SeXV and *SeATF2* CHR SeVII-SeVII) as well (4). In total, this resulted in the deletion of five different alleles, one short to the highest number of simultaneously completed deletions in *S. cerevisiae* (15). While previously suggested (21, 63), polycistronic ribozymes flanked gRNA expressed from a RNA pol II promoter had never been assayed before. Our results experimentally confirmed that 5' and 3' extension as designed at the junction of the two gRNA cassettes did not hinder self-cleavage of HH and HDV ribozymes and allow release of functional mature gRNAs. This result provided a glimpse of the potential of this mode of expression. It would



suggest that construction of polycistronic array including more than two gRNA could be contemplated.

#### **Expanding the *S. pastorianus* genetic tool box**

The present study delivered the first really efficient technical solution readily useable to perform targeted genetic modifications in *S. pastorianus*. The functionality of two modes of Cas9 expression was shown. Chromosomal integration of *cas9* (IMX1187) coupled with plasmid-based gRNA expression might be privileged when successive transformations are foreseen (7). However, plasmid-based *cas9* and gRNA co-expression proved to be as effective and presents the advantage to be easily transferable in multiple strain backgrounds. For efficient use of the provided repair fragment to recombine at the locus of the Cas9-induced DSB, Cas9 activity and presence of the repair DNA have to be synchronous. The correct integration of the repair fragment during single and double gene editing showed that the endonuclease was transcribed and translated fast enough for free linear DNA to still be available for repair of the induced DSB. These outcomes were in line with similar approaches attempted in *S. cerevisiae* or in *Aspergillus niger* (36, 47, 64). The presence of the gRNA is not constantly needed, as soon as the chromosomal double cut is inserted and preferably repaired, the plasmid has to be lost to recover a plasmid-free modified strain to either test the strain physiology or to prepare the constructed strain for a next editing round. The selection marker and replication origin used in the pUDP expression system tested in this work were designed to be broadly applicable and to facilitate rapid plasmid recycling. The dominant acetamidase marker confers the ability to use acetamide as sole nitrogen source and can be used in prototrophic strains such as lager yeasts or more generally industrial *Saccharomyces* strains. Plasmids carrying the *amdS* marker can be counter selected by growth in presence of fluoro-acetamide (34). Additionally the panARSopt replication origin (40)) derived from *K. lactis* used in the pUDP expression system was shown to be functional in a wide range of yeast species including *S. cerevisiae*. Contrarily to most replication origins such as the 2  $\mu$ m replication origin, which necessitates the presence of a wild type native 2  $\mu$ m plasmid to provide the enzymatic replicative machinery, panARSopt does not require any other genetic element. Furthermore, like *ARS-CEN*-based plasmids, panARS-based plasmids showed loss frequencies ranging between 5-10 % per generation when grown in non-selective conditions (40). These properties should permit efficient use of the pUDP expression system in various strain backgrounds, which might help to standardize a genome editing protocol starting from the design and cloning of the gRNA to the selection of correctly edited strains which have lost the pUDP plasmid.

Finally, while the scope of this work limited the tools application to single and double gene deletions, the availability of CRISPR-Cas9 editing tool makes a broad range of genetic modifications possible. Analogously to modification techniques applied in *S. cerevisiae*, the pUDP expression system might be applied for *in vivo* site directed mutagenesis and targeted introduction of multiple genes or entirely new pathways. In *S. pastorianus*, such modifications would finally allow to systematically investigate the contribution of genes involved in brewing-relevant phenotypes of *S. pastorianus*. In particular, the use of subgenome specific gRNA targets could enable targeted modification of genes from the *S. cerevisiae* and *S. eubayanus* subgenomes and thereby enable research on their interaction. For example, elucidation of the role of individual flocculation genes or implication of individual maltose and maltotriose transporter in *S. pastorianus* could now be envisaged.

## Conclusion

The gRNA and Cas9 expression system developed in this study enabled CRISPR-Cas9 engineering in *S. pastorianus*. The system was applied successfully for the deletion of all alleles of *SeILV6* and could be multiplexed successfully to obtain the simultaneous deletion of all alleles of *SeATF1* and *SeATF2*. While the system was only tested for gene deletion in this study, functional CRISPR-Cas9 engineering in *S. pastorianus* should also facilitate approaches such as gene insertions and directed mutagenesis. As *S. pastorianus* is notoriously resilient to genetic modification, these developments significantly improve its genetic accessibility and facilitate future research into the complex allo-aneuploid genome of *S. pastorianus*.

## Acknowledgements

The authors would like to thank Hannes Jürgens (Delft University of Technology) and Veronica Gast (Delft University of Technology) for their contributions to the pUDP plasmid modification system. We also would like to thank Xavier Hakkaart for kindly providing pUDR107. We are thankful to Prof. Jack T. Pronk (Delft University of Technology) and dr. Niels Kuijpers (HEINEKEN Supply Chain B.V.) for their support during this project and their critical reading of the manuscript.

## References

1. Libkind D, *et al.* (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. U. S. A.* 108(35):14539-14544.
2. Nakao Y, *et al.* (2009) Genome Sequence of the Lager Brewing Yeast, an Interspecies Hybrid. *DNA Res.* 16(2):115-129.
3. Okuno M, *et al.* (2016) Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res.* 23(1):67-80.
4. van den Broek M, *et al.* (2015) Chromosomal Copy Number Variation in *Saccharomyces pastorianus* Is Evidence for Extensive Genome Dynamics in Industrial Lager Brewing Strains. *Appl. Environ. Microbiol.* 81(18):6253-6267.
5. Walther A, Hesselbart A, & Wendland J (2014) Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. *G3 (Bethesda)* 4(5):783-793.
6. Hewitt SK, Donaldson IJ, Lovell SC, & Delneri D (2014) Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. *PLoS One* 9(3):e92203.
7. Duong CT, *et al.* (2011) Identification of Sc-type *ILV6* as a target to reduce diacetyl formation in lager brewers' yeast. *Metab. Eng.* 13(6):638-647.
8. Bolat I, Romagnoli G, Zhu F, Pronk JT, & Daran JM (2013) Functional analysis and transcriptional regulation of two orthologs of ARO10, encoding broad-substrate-specificity 2-oxo-acid decarboxylases, in the brewing yeast *Saccharomyces pastorianus* CBS1483. *FEMS Yeast Res* 13(6):505-517.
9. Murakami N, *et al.* (2012) Construction of a *URA3* deletion strain from the allotetraploid bottom-fermenting yeast *Saccharomyces pastorianus*. *Yeast* 29(5):155-165.
10. Bollag RJ, Waldman AS, & Liskay RM (1989) Homologous recombination in mammalian cells. *Annu Rev Genet* 23:199-225.
11. Choulika A, Perrin A, Dujon B, & Nicolas JF (1995) Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol* 15(4):1968-1973.
12. Kuijpers NGA, *et al.* (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb. Cell Fact.* 12.
13. Jinek M, *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816-821.
14. Cong L, *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819-823.
15. Mans R, *et al.* (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15(2):fov004.
16. DiCarlo JE, *et al.* (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41(7):4336-4343.
17. Jakociunas T, *et al.* (2015) CasEMBLR: Cas9-Facilitated Multiloci Genomic Integration of in Vivo Assembled DNA Parts in *Saccharomyces cerevisiae*. *ACS Synth Biol* 4(11):1226-1234.
18. Horwitz AA, *et al.* (2015) Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Syst* 1(1):88-96.

19. Ryan OW & Cate JHD (2014) Multiplex Engineering of Industrial Yeast Genomes Using CRISPRm. *Method Enzymol* 546:473-489.
20. Turowski TW & Tollervey D (2016) Transcription by RNA polymerase III: insights into mechanism and regulation. *Biochem Soc Trans* 44(5):1367-1375.
21. Gao Y & Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. *J Int Plant Biol* 56(4):343-349.
22. Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823-826.
23. Li W, Teng F, Li T, & Zhou Q (2013) Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol* 31(8):684-686.
24. Hwang WY, et al. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31(3):227-229.
25. Friedland AE, et al. (2013) Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods* 10(8):741-743.
26. Waaijers S, et al. (2013) CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. *Genetics* 195(3):1187-1191.
27. Gratz SJ, et al. (2013) Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194(4):1029-1035.
28. Schwartz CM, Hussain MS, Blenner M, & Wheelodon I (2016) Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. *ACS Synth Biol* 5(4):356-359.
29. Weninger A, Hatzl AM, Schmid C, Vogl T, & Glieder A (2016) Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J Biotechnol* 235:139-149.
30. Feng Z, et al. (2013) Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res* 23(10):1229-1232.
31. Zhang B, Yang X, Yang CP, Li MY, & Guo YL (2016) Exploiting the CRISPR/Cas9 System for Targeted Genome Mutagenesis in *Petunia*. *Sci Rep* 6.
32. Miao J, et al. (2013) Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res* 23(10):1233-1236.
33. Verduyn C, Postma E, Scheffers WA, & Vandijken JP (1992) Effect of Benzoic-Acid on Metabolic Fluxes in Yeasts - a Continuous-Culture Study on the Regulation of Respiration and Alcoholic Fermentation. *Yeast* 8(7):501-517.
34. Solis-Escalante D, et al. (2013) amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 13(1):126-139.
35. Thompson JR, Register E, Curotto J, Kurtz M, & Kelly R (1998) An improved protocol for the preparation of yeast cells for transformation by electroporation. *Yeast* 14(6):565-571.
36. Bao ZH, et al. (2015) Homology-Integrated CRISPR-Cas (HI-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol* 4(5):585-594.
37. Yanisch-Perron C, Vieira J, & Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33(1):103-119.
38. de Kok S, et al. (2014) Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth Biol* 3(2):97-106.
39. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343-345.
40. Liachko I & Dunham MJ (2014) An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res.* 14(2):364-367.
41. Li H & Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589-595.
42. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
43. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24-26.
44. Wartmann T, et al. (2003) High-level production and secretion of recombinant proteins by *Arxula adenivorans*. *Yeast* 20:S326-S326.
45. Goldstein AL & McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* 15(14):1541-1553.
46. Bankevich A, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455-477.
47. Generoso WC, Gottardi M, Oreb M, & Boles E (2016) Simplified CRISPR-Cas genome editing for *Saccharomyces cerevisiae*. *J Microbiol Methods* 127:203-205.
48. Guffanti E, et al. (2006) A minimal promoter for TFIIC-dependent *in vitro* transcription of snoRNA and tRNA genes by RNA polymerase III. *J Biol Chem* 281(33):23945-23957.
49. Harismendy O, et al. (2003) Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J* 22(18):4738-4747.
50. Ronda C, et al. (2015) CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 14.
51. Krappmann S, Sasse C, & Braus GH (2006) Gene targeting in *Aspergillus fumigatus* by homologous recombination is facilitated in a nonhomologous end-joining-deficient genetic background. *Eukaryot Cell* 5(1):212-215.
52. Ninomiya Y, Suzuki K, Ishii C, & Inoue H (2004) Highly efficient gene replacements in *Neurospora* strains deficient for nonhomologous end-joining. *Proc Natl Acad Sci USA* 101(33):12248-12253.

53. Kooistra R, Hooykaas PJ, & Steensma HY (2004) Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* 21(9):781-792.
54. Boulton SJ & Jackson SP (1996) *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J* 15(18):5093-5103.
55. Polotnianka RM, Li J, & Lustig AJ (1998) The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities. *Curr Biol* 8(14):831-834.
56. Laroche T, et al. (1998) Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Curr Biol* 8(11):653-656.
57. Gravel S, Larrivee M, Labrecque P, & Wellinger RJ (1998) Yeast Ku as a regulator of chromosomal DNA end structure. *Science* 280(5364):741-744.
58. Barnes G & Rio D (1997) DNA double-strand-break sensitivity, DNA replication, and cell cycle arrest phenotypes of Ku-deficient *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 94(3):867-872.
59. Varzakas TH, Arvanitoyannis IS, & Baltas H (2007) The politics and science behind GMO acceptance. *Crit. Rev. Food Sci. Nutr.* 47(4):335-361.
60. Gibson B & Liti G (2015) *Saccharomyces pastorianus*: genomic insights inspiring innovation for industry. *Yeast* 32(1):17-27.
61. Gorter de Vries AR, Pronk JT, & Daran JG (2017) Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts. *Appl. Environ. Microbiol.* 83:e03206-03216.
62. Piatkowska EM, Naseeb S, Knight D, & Delneri D (2013) Chimeric protein complexes in hybrid species generate novel phenotypes. *PLoS Genet* 9(10):e1003836.
63. Nissim L, Perli SD, Fridkin A, Perez-Pinera P, & Lu TK (2014) Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol Cell* 54(4):698-710.
64. Nodvig CS, Nielsen JB, Kogle ME, & Mortensen UH (2015) A CRISPR-Cas9 System for Genetic Engineering of Filamentous Fungi. *PLoS One* 10(7):e0133085.
65. Weber J, et al. (2017) Functional Reconstitution of a Fungal Natural Product Gene Cluster by Advanced Genome Editing. *ACS Synth Biol* 6(1):62-68.
66. Bolat I, Walsh MC, & Turtoi M (2008) Isolation and characterization of two new lager yeast strains from the WS34/70 population. *Roum Biotechnol Lett* 6:62-73.
67. Nijkamp JF, et al. (2012) *De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* 11.
68. Norrander J, Kempe T, & Messing J (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26(1):101-106.
69. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.



## Chapter 6: Allele-specific genome editing using CRISPR-Cas9 is associated with loss of heterozygosity in diploid yeast

Arthur R. Gorter de Vries, Lucas G. F. Couwenberg, Marcel van den Broek, Pilar de la Torre Cortés, Jolanda ter Horst, Jack T. Pronk and Jean-Marc G. Daran

Targeted DNA double-strand breaks (DSBs) with CRISPR-Cas9 have revolutionized genetic modification by enabling efficient genome editing in a broad range of eukaryotic systems. Accurate gene editing is possible with near-perfect efficiency in haploid or (predominantly) homozygous genomes. However, genomes exhibiting polyploidy and/or high degrees of heterozygosity are less amenable to genetic modification. Here, we report an up to 99-fold lower gene editing efficiency when editing individual heterozygous loci in the yeast genome. Moreover, Cas9-mediated introduction of a DSB resulted in large scale loss of heterozygosity affecting DNA regions up to 360 kb and up to 1700 heterozygous nucleotides, due to replacement of sequences on the targeted chromosome by corresponding sequences from its non-targeted homolog. The observed patterns of loss of heterozygosity were consistent with homology directed repair. The extent and frequency of loss of heterozygosity represent a novel mutagenic side-effect of Cas9-mediated genome editing, which would have to be taken into account in eukaryotic gene editing. In addition to contributing to the limited genetic amenability of heterozygous yeasts, Cas9-mediated loss of heterozygosity could be particularly deleterious for human gene therapy, as loss of heterozygous functional copies of anti-proliferative and pro-apoptotic genes is a known path to cancer.

Essentially as published in *Nucleic Acids Research* 2019;47(3):1362-1372

Supplementary materials available online

<https://doi.org/10.1093/nar/gky1216>

## INTRODUCTION

CRISPR-Cas9-assisted genome editing requires the simultaneous presence of the Cas9 endonuclease and a guide-RNA (gRNA) that confers target-sequence specificity (1). A gRNA consists of a structural domain and a variable sequence homologous to the targeted sequence (1-4). A Cas9-gRNA complex introduces a DSB when the gRNA binds to its reverse complement sequence on the 5' side of a PAM sequence (NGG). Imperfect gRNA complementarity and/or absence of a PAM sequence strongly reduce editing efficiencies (5). CRISPR-Cas9 enables specific editing of any sequence proximal to a PAM sequence, with minimal off-targeting effects (5). The introduction of a DSB facilitates genome editing by increasing the rate of repair by homologous recombination (6). When a repair fragment consisting of a DNA oligomer with homology to regions on both sides of the introduced DSB is added, it is integrated at the targeted locus by homologous recombination, resulting in replacement of the original sequence and repair of the DSB (2-4). In *S. cerevisiae*, double stranded DNA oligomers with 60 bp of homology are sufficient to obtain accurate gene-editing in almost 100 % of transformed cells (3). By inserting sequences between the homologous regions of the repair oligonucleotide, heterozygous sequences of up to 35 Kbp could be inserted at targeted loci (7). While such gene editing approaches have been very efficient in haploid and homozygous diploid yeasts, the accurate introduction of short DNA fragments can be tedious in heterozygous yeast. In homozygous diploid and polyploid eukaryotes, CRISPR-Cas9 introduces DSBs in all alleles of a targeted sequence (8). In heterozygous genomes, gRNAs can be designed for allele-specific targeting if heterozygous loci have different PAM motifs and/or different 5' sequences close to a PAM motif (8, 9), enabling allele-specific gene editing using Cas9. In such cases, a DSB is introduced in only one of the homologous chromosomes while the other homolog remains intact. However, the presence of intact homologous chromosomes facilitates repair of DSBs by homology-directed repair (HDR) using mechanisms such as homologous recombination (HR), or break-induced repair (BIR) (10-12). In particular, HDR of DSBs can induce chromosome recombinations and even loss of heterozygosity (LOH) in diploid genomes (9, 13-16). Therefore, the presence of an intact homologous chromosome could compete with an intended gene-editing event, resulting in reduced editing efficiency and possibly in extensive genetic changes due to LOH. So far, no systematic analysis has been performed of the efficiency of Cas9-mediated gene editing at heterozygous loci. To investigate if Cas9 gene editing works differently in heterozygous diploid yeast, we tested if allele-specific targeting of heterozygous loci using Cas9 enables accurate gene editing in an interspecies *Saccharomyces* hybrid, and investigated the resulting transformants. In addition, we systematically investigated the efficiency of Cas9-mediated genome editing when targeting various homozygous and heterozygous loci in diploid laboratory *Saccharomyces cerevisiae* strains while monitoring genetic changes.

## MATERIAL AND METHODS

### Strains, plasmids, primers and statistical analysis

*S. cerevisiae* strains used in this study are derived from the laboratory strains CEN.PK113-7D and S288C (17, 18). Yeast strains, plasmids and oligonucleotide primers used in this study are provided in Tables S3, S4 and S5. Statistical significance was determined using two-tailed unpaired Student's t-tests in GraphPad Prism 4 (GraphPad, La Jolla, CA).

### Media and growth conditions

Plasmids were propagated overnight in *Escherichia coli* XL1-Blue cells in 10 mL LB medium containing 10 g/L peptone, 5 g/L Bacto Yeast extract, 5 g/L NaCl and 100 mg/L ampicillin at 37°C. Unless indicated otherwise, yeast strains were grown at 30 °C and 200 RPM in 100 mL flat-bottom flasks containing 50 mL YPD medium, containing 10 g/L Bacto yeast extract, 20 g/L Bacto peptone, and 20 g/L glucose. Alternatively, strains were grown in synthetic medium (SM) containing 3.0 g/L  $\text{KH}_2\text{PO}_4$ , 5.0 g/L  $(\text{NH}_4)_2\text{SO}_4$ , 0.5 g/L  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mL/L trace elements, 1 mL/L vitamin solution and 20 g/L glucose (19). For uracil auxotrophic strains, SM-derived media were supplemented with 150 mg/L uracil (20). Solid media were supplemented with 20 g/L agar. Selection for the amdSYM marker was performed on SM-AC: SM medium with 0.6 g/L acetamide and 6.6 g/L  $\text{K}_2\text{SO}_4$  as nitrogen and sulfur sources instead of  $(\text{NH}_4)_2\text{SO}_4$  (21). The amdSYM marker was lost by growth on YPD and counter-selected on SM-FAC: SM supplemented with 2.3 g/L fluoroacetamide (21). Yeast strains and *E. coli* containing plasmids were stocked in 1 mL aliquots after addition of 30 % v/v glycerol to the cultures and stored at -80 °C.

### Flow cytometric analysis

Overnight aerobic cultures in 100 mL flat-bottom flasks on 20 mL YPD medium were vortexed thoroughly to disrupt cell aggregates and used for flow cytometry on a BD FACSAria™ II SORP Cell Sorter (BD Biosciences, Franklin Lakes, NJ) equipped with 355 nm, 445 nm, 488 nm, 561 nm and 640 nm lasers and a 70 µm nozzle, and operated with filtered FACSFlow™ (BD Biosciences). Cytometer performance was evaluated prior to each experiment by running a CST cycle with CS&T Beads (BD Biosciences). Drop delay for sorting was determined by running an Auto Drop Delay cycle with Accudrop Beads (BD Biosciences). Cell morphology was analysed by plotting forward scatter (FSC) against side scatter (SSC). The fluorophore mRuby2 was excited by the 561 nm laser and emission was detected through a 582 nm bandpass filter with a bandwidth of 15 nm. The fluorophore mTurquoise2 was excited by the 445 nm laser and emission was detected through a 525 nm bandpass filter with a bandwidth of 50 nm. The fluorophore Venus was excited by the 488 nm laser and emission was detected through a 545 nm bandpass filter with a bandwidth of 30 nm. For each sample, 100'000 events were analysed and the same gating strategy was applied to all samples of the same strain. First, "doublet" events were discarded on a FSC-A/FSC-H plot, resulting in at least 75'000 single cells for each sample. Of the remaining single cells, cells with and cells without fluorescence from Venus were selected in a FSC-A/Venus plot. For both these groups, cells positive for mRuby2 and mTurquoise2, cells positive for only mRuby2, cells positive for only mTurquoise2 and cells negative for mRuby2 and mTurquoise2 were gated. The same gating was used for all samples of each strain. Sorting regions ('gates') were set on these plots to determine the types of cells to be sorted. Gated single cells were sorted in 96-well microtiter plates containing YPD using a "single cell" sorting mask, corresponding to a yield mask of 0, a purity mask of 32 and a phase mask of 16. FACS data was analysed using FlowJo® software (version 3.05230, FlowJo, LLC,



Ashland, OR). Separate gating strategies were made for IMX1555, IMX1557 and IMX1585 to account for possible differences in cell size, shape and morphology.

#### **Plasmid assembly**

Plasmid pUD574 was *de novo* synthesised at GeneArt (Thermo Fisher Scientific, Waltham, MA) containing the sequence 5' GGTCTCGCAAAATTACTGATGAGTCCGTGAGGACGAAACGAGTAAAGCTCGTCTGTAATATCTTAATGCTAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTGAAAAAGTGGCACCAGTCCGGTCTTTGGCCGGCATGGTCCCAGCCTCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACACAGCGAGACC 3'.

Plasmids pUD429 was constructed in a 10 µL golden gate assembly using T4 ligase (Thermo Fisher Scientific) and BsaI (New England BioLabs, Ipswich, MA) from 10 ng of parts pYTK002, pYTK047, pYTK067, pYTK079, pYTK081 and pYTK083 of the yeast toolkit as described previously (22). Similarly, pUD430 was constructed from pYTK003, pYTK047, pYTK068, pYTK079, pYTK081 and pYTK083, and pUDP431 from pYTK004, pYTK047, pYTK072, pYTK079, pYTK081 and pYTK083. Plasmid pUDE480 expressing mRuby2 was constructed from GFP dropout plasmid pUD429 with pYTK011, pYTK034 and pYTK054 using golden gate assembly as described previously (22). Similarly, pUDE481 expressing mTurquoise2 was constructed from pUD430, pYTK009, pYTK032 and pYTK053, and pUDE482 expressing Venus from pUD431, pYTK013, pYTK033 and pYTK055.

Plasmids pUDR323, pUDR324, pUDR325, pUDR358, pUDR359, pUDR360, pUDR361 and pUDR362, expressing gRNAs targeting *SIT1*, *FAU1*, *spcas9*, *UTR2*, *FIR1*, *AIM9*, *YCK3* and intergenic region 550K respectively, were constructed using NEBuilder® HiFi DNA Assembly Master Mix by assembling the 2 µm fragment amplified from pROS11 with primers 12230, 12235, 9457, 12805, 12806, 12807, 12808, 12809 respectively, and the plasmid backbone amplified from pROS11 with primer 6005 as described previously (3, 23).

Plasmid pUDP045, expressing gRNA<sub>MAL11</sub> and *cas9*, was constructed by Golden Gate cloning by digesting pUDP004 and pUD574 using BsaI and ligating with T4 ligase (24). Correct assembly was verified by restriction analysis using PmlI.

#### **Strain construction**

Yeast strains were transformed according to the high-efficiency protocol by Gietz *et al* (25). IMX1544 was constructed by transforming IMX581 with 1 µg pUDR323 and 1 µg of a repair fragment amplified from pUD481 using primers 12233 and 12234 containing an expression cassette for mTurquoise2 and 60 bp homology arms with the *FAU1* locus. IMX1555 was constructed by transforming IMX1544 with 1 µg pUDR324 and 1 µg of repair fragment amplified from pUD480 using primers 12228 and 12229 containing an expression cassette for mRuby2 and 60 bp homology arms with the *SIT1* locus. Transformants were selected on SM-AC plates, three single colony isolates were grown overnight on YPD and streaked on SM-FAC plates. Genomic DNA of a single colony was extracted, insertion of mTurquoise2 in *FAU1* was confirmed by PCR using primers 12236 and 12237, and insertion of mRuby2 in *SIT1* was confirmed by PCR using primers 12231 and 12232 followed by digestion with PvuII and XhoI digestion. IMX1557 was constructed by adding 10 µL of stationary phase culture of IMX1555 and of IMK439 in 1 mL of SM medium, incubating overnight at 30 °C and plating on SM plates with 10 mg/L clonNAT and 100 mg/L G418. IMX1585 was constructed by adding 10 µL of stationary phase culture of IMX1555 and of S288C in 1 mL of SM medium, incubating overnight at 30 °C and plating on SM plates with 10 mg/L clonNAT without added uracil. All constructed strains

were grown overnight in YPD and fluorescence corresponding to mRuby2 and mTurquoise2 was verified by flow cytometry.

#### **Cas9 mediated targeting in *S. cerevisiae* x *S. eubayanus* hybrid IMS0408**

IMX1421, IMX1422, IMX1423 and IMX1424 were constructed by transforming IMS0408 with 1 µg pUDP045 and 1 µg of a 120 bp repair fragment constructed by annealing primers 10813 and 10814 as described previously (8). Transformants were selected on SM-AC plates, genomic DNA of 10 single colonies was extracted, but no band could be obtained when amplifying the *MAL11* locus using primer sets 1084/1470 and 1657/1148. The exact same procedure was performed without the addition of the 120 bp repair fragment. Four randomly selected colonies transformed with repair fragment were re-streaked three times on YPD agar, the plasmid was counter-selected for by plating on SM-FAC and the isolates were stocked as IMX1421, IMX1422, IMX1423 and IMX1424.

#### **Cas9 mediated introduction of DSBs in *S. cerevisiae* strains**

DSBs were introduced by transforming yeast strains using 1 µg of purified gRNA expression plasmid and 1 µg of gel-purified double stranded repair fragment. The expression of gRNAs was done with plasmids pMEL11 to target *CAN1*, pUDR325 to target *cas9*, pUDR358 to target *UTR2*, pUDR359 to target *FIR1*, pUDR360 to target *AIM9*, pUDR361 to target *YCK3* and pUDR362 to target 550K according to Mans *et al* (3, 23). Repair fragments containing Venus expression cassettes were PCR amplified from plasmid pUDE482 with primers with an overlap of about 20 bp with the nucleotides flanking the targeted open reading frame and purified on a 1 % agarose gel (Table S5). Upon transformation, the cells were transferred to 100 mL flat-bottom flasks containing 20 mL SM-AC medium and grown until stationary phase at 30°C and 200 RPM to select cells transformed with the gRNA expression plasmid. After about 72h, 0.2 mL of these cultures was transferred to fresh SM-AC and grown under the same conditions to stationary phase to dilute any remaining untransformed cells. After about 48h, 0.2 mL of these cultures was transferred to 100 mL flat-bottom flasks containing 20 mL YPD medium and grown for about 12h under the same conditions to obtain optimal fluorescence signals.

#### **DNA extraction and whole genome analysis**

IMX1557, IMX1585, IMX1596-IMX1635, IMS0408 and IMX1421-IMX1424 were incubated in 500 mL flat-bottom flasks containing 100 mL liquid YPD medium at 30 °C on an orbital shaker set at 200 RPM until the strains reached stationary phase with an OD<sub>660</sub> between 12 and 20. Genomic DNA was isolated using the Qiagen 100/G kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and quantified using a Qubit® Fluorometer 2.0 (Thermo Fisher Scientific). Between 11.5 and 54.6 µg genomic DNA was sequenced by Novogene Bioinformatics Technology Co., Ltd (Yuen Long, Hong Kong) on a HiSeq 2500 (Illumina, San Diego, CA) with 150 bp paired-end reads using TruSeq PCR-free library preparation (Illumina). For IMX1557, IMX1585 and IMX1596-IMX1635, reads were mapped onto the *S. cerevisiae* CEN.PK113-7D genome (17) using the Burrows–Wheeler Alignment tool (BWA) and further processed using SAMtools and Pilon for variant calling (26-28). Homozygous SNPs from IMX1585 were subtracted from the list of homozygous SNPs of each strain and a list of homozygous SNPs on chromosome V was compiled per strain. Based on the list of heterozygous SNPs in IMX1585, all homozygous SNPs corresponded to the nucleotide from S288C while the nucleotide from IMX1557 was lost, and regions were identified in which all contiguous heterozygous SNPs lost heterozygosity for each strain. LOH was confirmed by visualising the generated .bam files in the Integrative Genomics Viewer (IGV) software (29). Regions mapped as

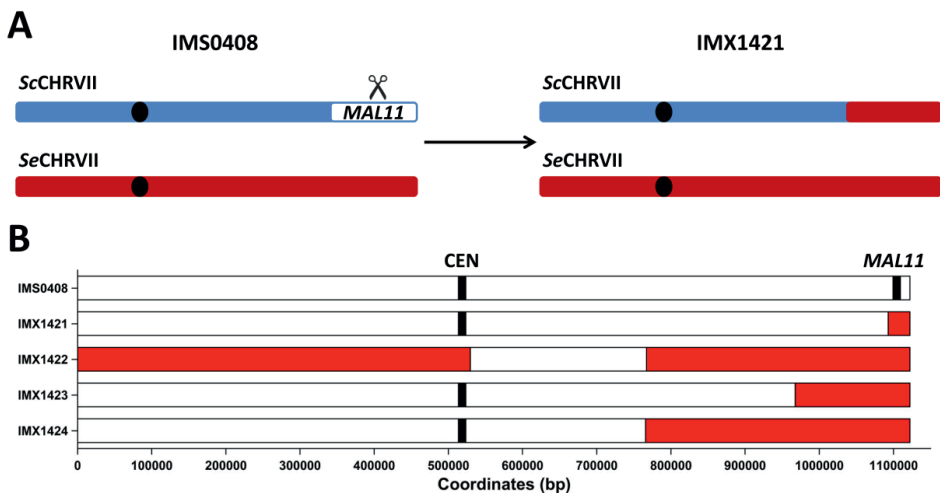
having lost heterozygosity correspond to regions between the first and last nucleotide which lost heterozygosity. For IMS0408 and IMX1421-IMX1424, reads were aligned to a reference genome obtained by combining the reference genome of CEN.PK113-7D (17) and the reference genome of *S. eubayanus* strain CBS 12357 (30) as they are closely related to the haploid parents of IMS0408. Regions affected by LOH were defined as regions in which reads did not align to the *S. cerevisiae* reference chromosome VII while reads aligned to the corresponding region of the *S. eubayanus* reference chromosome VII with approximately double the normal coverage.

## RESULTS

### Targeting of a heterozygous gene in a *S. cerevisiae* x *S. eubayanus* hybrid

To investigate Cas9 gene editing in a genetic context with extensive heterozygosity, we targeted a heterozygous locus in an interspecies *S. cerevisiae* x *S. eubayanus* hybrid. The hybrid IMS0408 was constructed previously by mating a haploid *S. cerevisiae* laboratory strain and a haploid spore from the *S. eubayanus* type strain CBS 12357, resulting in an allodiploid strain with approximately 85 % nucleotide identity between corresponding chromosomes of the two subgenomes (31). The *MAL11* gene encodes a membrane transporter located on chromosome VII in *S. cerevisiae*, which is absent in the *S. eubayanus* CBS 12357 genome. Therefore, the *S. cerevisiae* chromosome VII could be specifically targeted using Cas9 and a gRNA targeting *MAL11*. IMS0408 was transformed with plasmid pUDP045, expressing Cas9 and a gRNA targeting *MAL11*, with and without a repair fragment with 60 bp of homology to sequences adjacent to the 5' and 3' ends of the coding region of *MAL11*. Normally, selection for the presence of the Cas9/gRNA expression plasmid is sufficient to obtain accurate gene editing in almost 100 % of transformed cells without the need of a selection marker incorporated in the repair fragment in *Saccharomyces* yeast (3, 8). In common laboratory strains, replacement of a sequence with a repair DNA can be detected by diagnostic PCR. However, in the hybrid strain IMS0408, multiple attempts failed to yield the expected fragments after transformation with the gRNA targeting *MAL11* and a repair fragment. Therefore, the genomes of four random transformants, named IMX1421 to IMX1424, were sequenced using 150 bp paired-end Illumina reads and aligned to a haploid *S. cerevisiae* x *S. eubayanus* reference genome. While reads of strain IMS0408 aligned unambiguously to the *MAL11* locus on chromosome VII of the *S. cerevisiae* sub-genome, *MAL11* DNA was absent in transformants IMX1421-IMX1424. Absence of *MAL11* was associated with loss of large regions of chromosome VII, ranging from 29 to 356 kbp (Fig. 1). For IMX1422-IMX1424, the corresponding regions on the *S. eubayanus* chromosome VII devoid of *MAL11* ortholog showed double sequence coverage, indicating that targeting of *MAL11* using Cas9 resulted in replacement of varying regions of the targeted *S. cerevisiae* chromosome by regions from the corresponding *S. eubayanus* chromosome (Fig. S1). The recombination in IMX1423 occurred between the *S. cerevisiae* *HSV2* gene and its *S. eubayanus* *HSV2* ortholog. The recombination events in IMX1422 and IMX1424 both occurred between the *S. cerevisiae* *YGR125W* gene and its *S. eubayanus* *YGR125W* ortholog. The exact coordinates of the recombination within *YGR125W* were separated by more than 1000 nucleotides. For IMX1421, the loss of *S. cerevisiae* chromosome VII started at the *IMA2* gene. However, the presence of other *IMA* genes with high identity to *IMA2* prevented unique read alignment. Therefore, read pairing information did not reveal with which sequence the right arm of chromosome VII was replaced. In addition, the subtelomeric position of other *IMA* genes in the genome prevented identification of this duplicated sequence by sequencing coverage analysis, as sequencing depth is highly irregular in subtelomeric regions due to the abundance of repetitive elements (17). While the *MAL11* locus was targeted in all four strains, the recombination events

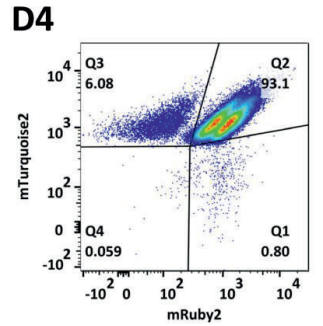
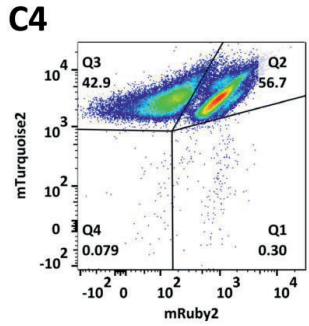
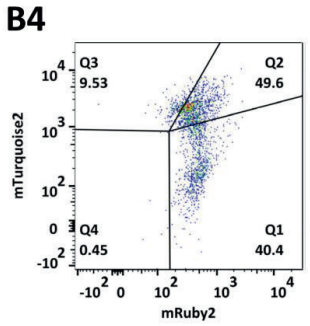
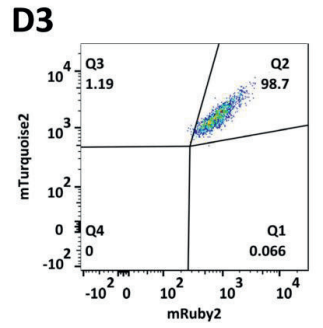
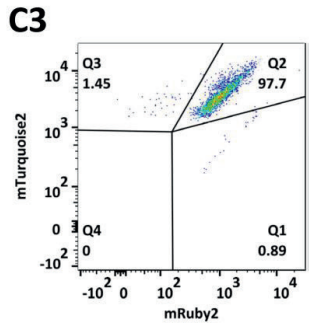
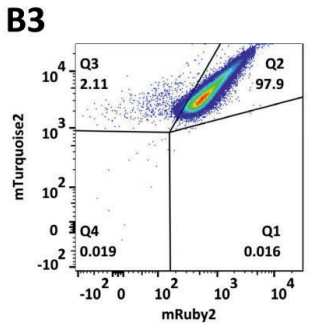
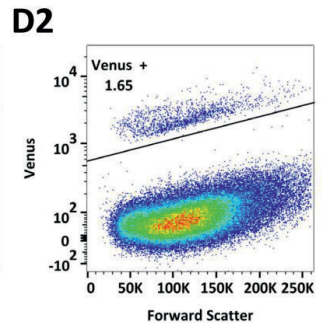
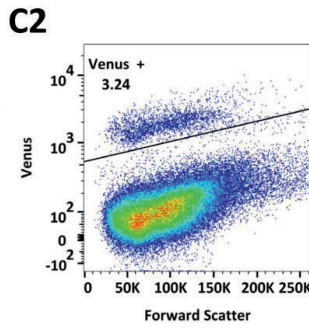
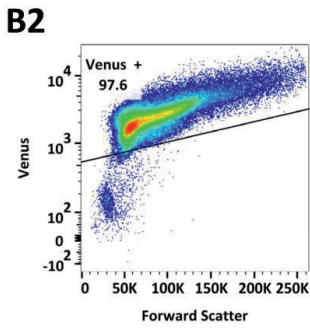
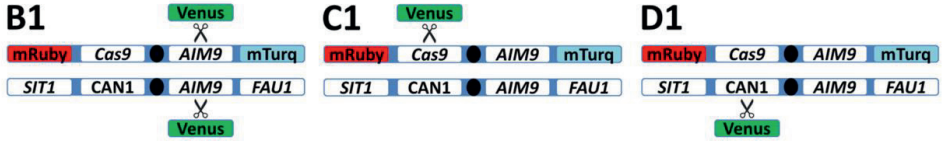
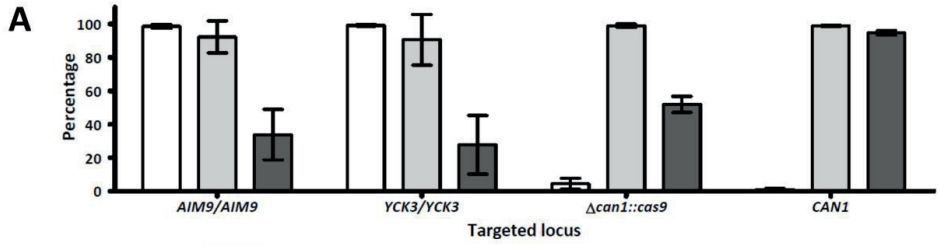
leading to LOH occurred at four unique loci. The distance of these loci to the targeted site varied between 7 and 334 kbp, possibly reflecting different degrees of DNA resection at the DSB site. The sequence similarities of the *S. cerevisiae* and *S. eubayanus* orthologs of *HSV2* and *YGR125W* were 80 % and 82 %, which is lower than the average 85 % identity between the two subgenomes. This observation indicates that recombination events did not only occur in regions with particularly high homology. It should be noted that in IMX1422, LOH did not only affect the right arm of *S. cerevisiae* chromosome VII, but also the first 530 kbp of the left arm of *S. cerevisiae* chromosome VII (Fig. 1). Since no segmental aneuploidies were observed on non-targeted chromosomes in IMX1421-IMX1424 (Fig. S1), the observed LOH is likely due to the targeting of *MAL11*. These results indicated that genome editing using Cas9 caused LOH rather than the intended gene editing when targeting a locus present on just one of two homologous chromosomes in a heterozygous yeast.



**Figure 1. Loss of heterozygosity observed by whole genome sequencing upon Cas9-targeting of *MAL11* on the *Saccharomyces cerevisiae* derived chromosome VII in the *S. cerevisiae* x *S. eubayanus* hybrid IMS0408.** IMS0408 was transformed with a 120 bp repair fragment with 60 bp flanks corresponding to the sequence upstream and downstream of the *MAL11* ORF, and with plasmid pUDP045 expressing Cas9 and a gRNA targeting the *S. cerevisiae* specific gene *MAL11* gene. Upon plating on selective medium, four randomly picked colonies were selected and sequenced using 150 bp pair-end reads and mapped against a reference genome composed of chromosome level assemblies of *S. cerevisiae* and of *S. eubayanus*. The centromere and targeted gene *MAL11* are shown at their exact coordinates, but their size is not at scale. Loss of heterozygosity is shown in red and was defined as regions in which reads did not align to the *S. cerevisiae* reference chromosome VII while reads aligned to the corresponding region on the *S. eubayanus* reference chromosome VII with approximately double the normal coverage.

### Targeting of heterozygous loci in a mostly homozygous diploid *S. cerevisiae* strain

To investigate if the observed lack of efficient gene editing was specific to this highly heterozygous *S. cerevisiae* x *S. eubayanus* hybrid, we systematically investigated the impact of target-sequence heterozygosity on the efficiency of gene editing in *S. cerevisiae* strains. To this end, DSBs were introduced at homozygous and heterozygous loci on chromosome V of several strains that carried a Cas9 expression cassette integrated at the *CAN1* locus. Plasmid-based gRNA expression was performed as described previously (3). Use of a repair fragment expressing the fluorescent protein Venus enabled analysis of editing efficiency by flow cytometry (22). To verify functional Cas9 and gRNA expression, the  $\Delta can1::Spcas9$  locus was first targeted in the haploid *S. cerevisiae* strain

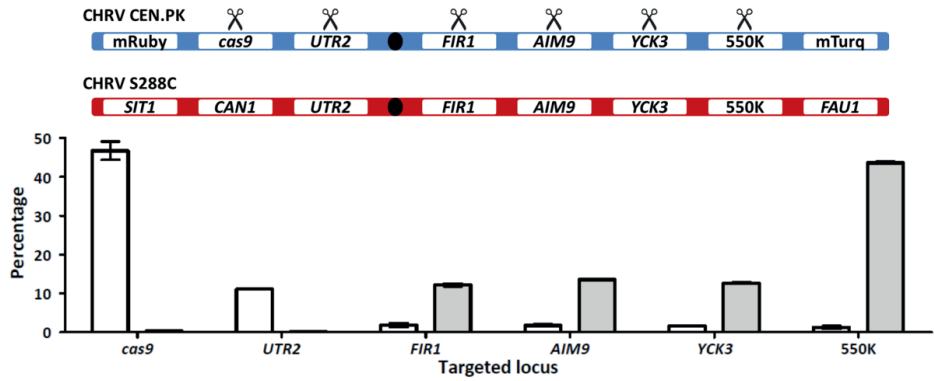


**Figure 2. Cas9-mediated gene editing of homozygous and heterozygous loci on chromosome V of *S. cerevisiae*.** (A) Average fluorescence of cell populations in which the homozygous *AIM9* and *YCK3* alleles and the heterozygous *Cas9* and *CAN1* alleles were targeted in the diploid strain IMX1557. The percentage of cells expressing Venus (white), the percentages of cells expressing both mTurquoise2 and mRuby2 in the Venus positive (light grey) and in Venus negative cells (dark grey) are shown. For each target, averages and standard deviation for biological triplicates are shown. When targeting the hemizygous *CAN1* allele, LOH manifests itself in a lower percentage of cells expressing Venus, but it does not affect mTurquoise2 and mRuby2 fluorescence, as these fluorophores are located on the non-targeted chromosome copy. (B,C and D) Fluorescence profiles obtained when targeting *AIM9*, *Cas9* and *CAN1* in IMX1557. (row 1) Schematic representation of both copies of chromosome V in IMX1557, with the alleles at the *SIT1*, *CAN1*, *AIM9* and *FAU1* loci and scissors indicating Cas9 targeting. While one chromosome copy has the wildtype alleles for all loci, the other copy has mRuby2 integrated in *SIT1*, Cas9 integrated in *CAN1* and mTurquoise2 integrated in *FAU1*. (rows 2, 3 and 4) Flow cytometry profiles of targeted cells. Each gene was targeted in three biological replicates and flow cytometric data for a representative replicate is shown. After transformation, 100,000 cells were analysed by flow cytometry and single cells were selected based on a FSC-A/FSC-H plot to avoid multicellular aggregates. For each replicate, at least 75,000 single cells remained and the fluorescence corresponding to Venus was used to determine gene-editing efficiency (row 2). For each gene, the fluorescence corresponding to mRuby2 and mTurquoise2 is plotted for the cells with expression of Venus (row 3) and for cells without expression of Venus (row 4). Fluorescence results for all samples are provided in Table S1.

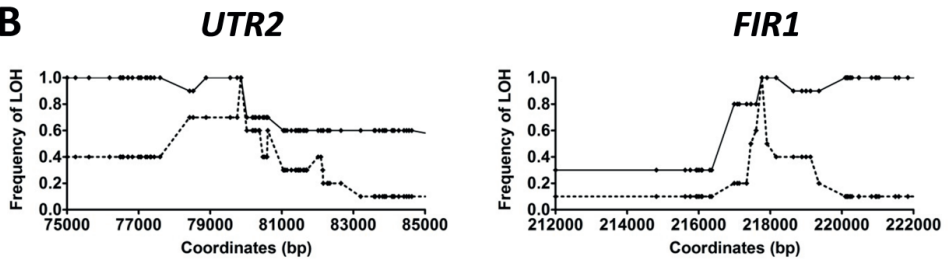
IMX1555, resulting in integration of the repair fragment in  $98.3 \pm 1.3$  % of cells (Table S1). Subsequently, the homozygous alleles of *AIM9* and *YCK3* were targeted in the congenic diploid *S. cerevisiae* strain IMX1557, resulting in integration of the repair fragment in  $98.6 \pm 0.8$  % and  $99.2 \pm 0.4$  % of cells, respectively (Fig. 2A). In contrast, when individually editing each allele of the heterozygous *CAN1*/ $\Delta can1::cas9$  locus in the diploid strain IMX1557, the repair fragment was integrated in only  $4.4 \pm 2.5$  % of cells when targeting the  $\Delta can1::cas9$  allele, and  $0.9 \pm 0.6$  % of the cells when targeting the *CAN1* allele (Fig. 2A). These results indicated that gene editing efficiencies were up to 99-fold lower for heterozygous target loci than for homozygous target loci ( $p < 10^{-4}$ ). Since IMX1557 was homozygous in most of its genome, except the targeted locus, the introduction of a DSB in only one of two homologous chromosomes rather than genome heterozygosity itself, impeded accurate and efficient gene editing using Cas9.

To further investigate if Cas9 gene editing resulted in LOH, as observed in the hybrid IMS0408, the presence of both chromosome arms of the targeted chromosome homolog was monitored by flow cytometry. IMX1557 expressed the fluorophores mRuby2 and mTurquoise2 from the *SIT1* and *FAU1* loci of the chromosome V copy harbouring the  $\Delta can1::cas9$  allele, but not from the non-modified homologous chromosome (Fig. 2, B1-D1). Loss of the left and right arms of the copy of chromosome V harbouring  $\Delta can1::cas9$  could therefore be monitored by measuring fluorescence corresponding to respectively mRuby2 and mTurquoise2 (22). For all targeted loci, when the expression of Venus indicated correct gene-editing, over 97.7 % of cells expressed both mTurquoise2 and mRuby2 (Fig. 2, panels A and B3-D3). However, when targeting the  $\Delta can1::cas9$  allele on the chromosome harbouring mRuby2, 42.9 % of cells which did not integrate Venus had lost mRuby2 fluorescence, while mTurquoise2 was still expressed (Fig. 2, quadrant Q3 in panel C4). These results indicated that targeting of the heterozygous  $\Delta can1::cas9$  allele resulted in LOH of the targeted chromosome arm harbouring mRuby2, but did not affect the opposite chromosome arm. In addition, when targeting the *CAN1* allele on the chromosome without mRuby2, an additional population expressing both mRuby2 and mTurquoise2 emerged among the cells which did not integrate Venus (Fig. 2, quadrant Q2 in panel D4). Within quadrant Q2, the two adjacent populations had the same average mTurquoise2 fluorescence, but their average mRuby2 fluorescence differed by a factor of 2. The difference in fluorescence suggested a duplication of mRuby2, consistent with replacement of the targeted non-fluorescent chromosome by an additional copy of the chromosome harbouring mRuby2. Loss of mRuby2 fluorescence upon transformation with a gRNA targeting  $\Delta can1::cas9$  and doubling of mRuby2 fluorescence when targeting *CAN1* were also observed in the absence of a

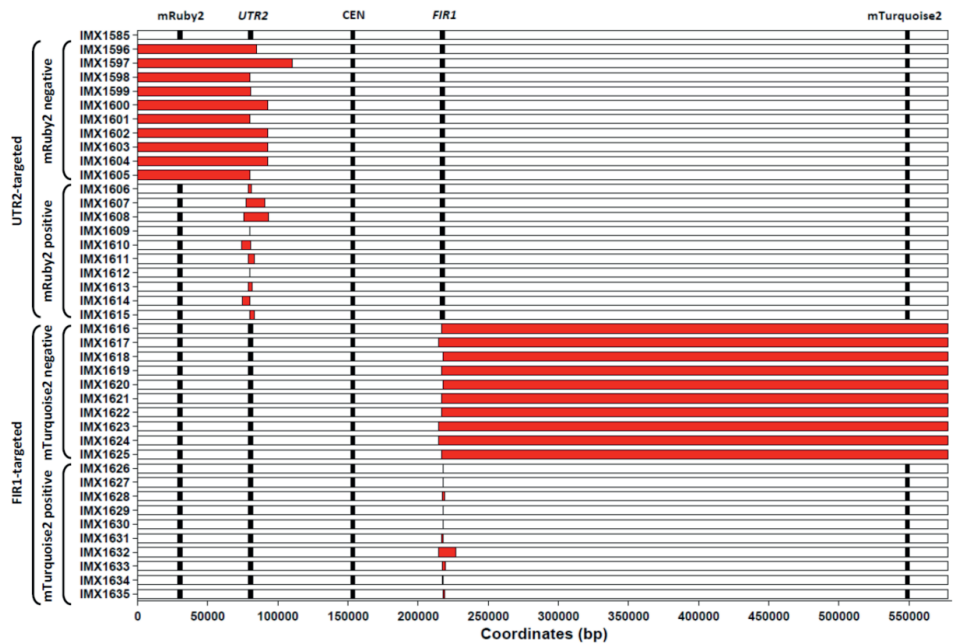
**A**



**B**



**C**



**Figure 3: Loss of heterozygosity caused by Cas9-mediated gene editing at heterozygous loci in the heterozygous *S. cerevisiae* diploid IMX1585.** Mating of the haploid *S. cerevisiae* strains IMX1555 (CEN.PK-derived) and S288C yielded the heterozygous diploid strain IMX1585 (on average 4 heterozygous nucleotides per kbp on chromosome V). The CEN.PK-derived chromosome harbours the fluorophores mRuby2 and mTurquoise2, enabling detection of the loss of each arm of the CEN.PK-derived chromosome V by flow cytometry. DSBs were introduced specifically in the CEN.PK-derived chromosome and loss of heterozygosity was monitored at the population level using flow cytometry and in single cell isolates by whole genome sequencing. (A) Population-level loss of heterozygosity after targeting *cas9*, *UTR2*, *FIR1*, *AIM9*, *YCK3* and 550K in IMX1585. The targeted loci on the CEN.PK-derived and S288C-derived chromosome V of IMX1585 are represented schematically. The *SIT1*, *CAN1*, *UTR2*, *FIR2*, *AIM9*, *YCK3*, 550K and *FAU1* loci and the centromeres are indicated. The CEN.PK-derived chromosome harbours mRuby2 at the *SIT1* locus and mTurquoise2 at the *FAU1* locus. Scissors indicate CEN.PK-derived alleles which were specifically targeted using Cas9. In the graph, the percentage of cells having lost mRuby2 fluorescence (white) and mTurquoise2 (grey) is shown for each targeted locus. Averages and standard deviations were calculated from biological triplicates. (B) Loss of heterozygosity at the nucleotide level in single isolates obtained by targeting *UTR2* and *FIR1* in IMX1585. For both targeted loci, the frequency at which LOH was observed for each heterozygous nucleotide in a 10 kbp region around the targeted site was determined by whole genome sequencing. For *UTR2*, the 79,857<sup>th</sup> nucleotide was targeted and frequencies are indicated for 10 isolates with intact fluorescence (dashed line, isolates IMX1606-IMX1615) and 10 isolates having lost fluorescence corresponding to mRuby2 (continuous line, IMX1596-IMX1605). For *FIR1*, the 217,767<sup>th</sup> nucleotide was targeted and frequencies are indicated for 10 isolates with intact fluorescence (dashed line, isolates IMX1626-IMX1635) and 10 isolates having lost fluorescence corresponding to mTurquoise2 (continuous line, IMX1616-IMX1625). (C) Loss of heterozygosity at the chromosome scale in single isolates obtained by targeting *UTR2* and *FIR1* in IMX1585. Whole genome sequencing data was used to identify regions of chromosome V affected by loss of heterozygosity in isolates after targeting of *UTR2* (IMX1596-IMX1615) and of *FIR1* (IMX1616-IMX1635). For each strain, the fluorophores mRuby2 and mTurquoise2, the targeted genes *UTR2* and *FIR1* and the centromere are shown at their exact coordinates, but their size is not at scale. Loss of heterozygosity was defined as regions in which nucleotides which were heterozygous in IMX1585 were no longer heterozygous in the isolate (in red). In isolates which lost the fluorophore mRuby2 after targeting of *UTR2* or which lost the fluorophore mTurquoise2 after targeting of *FIR1*, entire chromosome arms were affected by LOH. Exact coordinates are provided in Table S2.

co-transformed repair fragment (Table S1). These results indicated that introduction of a DSB at a heterozygous locus caused LOH through replacement of a targeted chromosome segment by duplication of the corresponding segment from its homologous chromosome, as was observed when targeting *MAL11* in the *S. cerevisiae* x *S. eubayanus* hybrid IMS0408.

### Elucidation of genetic changes caused by Cas9-targeting using whole genome sequencing

Chromosome-arm LOH has previously been reported upon introduction of a DSB in one of two homologous chromosomes, but was considered rare and has not been described as disruptive to gene-editing approaches (9, 13, 32). To investigate the extent and nature of the LOH caused by Cas9-editing of heterozygous loci, a strain with an average of four heterozygous SNPs or INDELS per kbp was generated by mating IMX1555 (CEN.PK genetic background, expressing Cas9, mRuby2 and mTurquoise2 from chromosome V) with S288C (Table S6). LOH could be monitored at the chromosome arm level by flow cytometry and at the nucleotide level by whole-genome sequencing. By using PAM sequences absent in S288C, we specifically targeted the CEN.PK-derived chromosome V, which carried expression cassettes for mRuby2 and mTurquoise2 on its left and right arms, respectively, at the *CAN1*, *UTR2*, *FIR1*, *AIM9* and *YCK3* loci and at intergenic coordinate 549603, referred to as 550K. Upon targeting of the *CAN1* and *UTR2* loci, mRuby2 fluorescence was lost in  $46.7 \pm 2.4$  and  $11.2 \pm 0.2$  % of cells, respectively, while mTurquoise2 fluorescence was unaffected in at least  $99.6 \pm 0.2$  % of the cells (Fig. 3A). Targeting of the *FIR1*, *AIM9*, *YCK3* or 550K loci caused loss of mTurquoise2 fluorescence in  $12.2 \pm 0.4$ ,  $13.6 \pm 0.1$ ,  $12.7 \pm 0.2$  and  $43.6 \pm 0.3$  % of cells, respectively, while mRuby2 fluorescence was conserved in at least  $98.1 \pm 0.5$  % of cells (Fig. 3A).

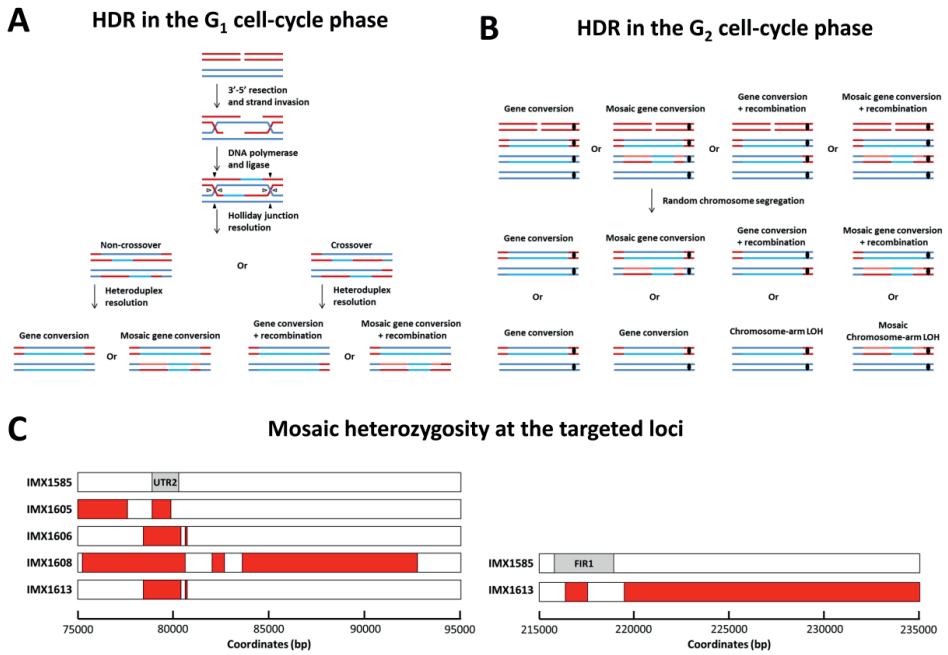


As the centromere is located between *UTR2* and *FIR1*, these results confirm that, for all investigated loci, a large fraction of cells lost the targeted chromosome arm. Fluorescence-activated cell sorting (FACS) was subsequently used to isolate 10 single cells each from the following populations: *UTR2*-targeted cells with mRuby2 fluorescence (IMX1606-IMX1615), *UTR2*-targeted cells without mRuby2 fluorescence (IMX1596-IMX1605), *FIR1*-targeted cells with mTurquoise2 fluorescence (IMX1626-IMX1635), and *FIR1*-targeted cells without mTurquoise2 fluorescence (IMX1616-IMX1625). Whole-genome sequencing and alignment of reads to the CEN.PK113-7D genome sequence (17) revealed LOH of the targeted locus in all 40 isolates (Fig. 3B). In cell lines that did not lose a fluorophore, LOH was local, affecting regions ranging from 3 to 17,495 nucleotides for *UTR2*-targeted cells and regions ranging from 1 to 11,900 nucleotides for *FIR1*-targeted cells, corresponding to up to 79 heterozygous nucleotides (Fig. 3C and Table S2). In isolates that did lose a fluorophore, LOH affected the chromosome arm harbouring the targeted locus, affecting 79,859 to 110,289 nucleotides for *UTR2*-targeted cells and 359,841 to 362,790 nucleotides for *FIR1*-targeted cells, corresponding to up to 1,697 heterozygous nucleotides (Fig. 3C and Table S2). Absence of newly introduced SNPs at targeted loci indicated that repair of DSBs did not involve non-homologous end joining (33).

#### **Identification of repair patterns corresponding to homology-directed repair**

We conclude that introduction of a DSB at a heterozygous locus results in low gene-editing efficiencies due to a competing repair mechanism that causes local or chromosome-arm LOH. In eukaryotes, repair using homologous chromosomes typically relies on BIR or HR (34), which occur by distinct mechanisms and yield different results (10-12). In the case of BIR, the entire targeted chromosome arm is lost and an additional copy of its homolog is generated from the 5' strand by replication, using the homolog as a polymerase template. Depending on the degree of strand resection prior to BIR, this mechanism results in complete loss of heterozygosity for varying portions of the targeted chromosome arm, including the locus in which a DSB was introduced (35). In the case of HR, the DSB is repaired by strand invasion, strand elongation, ligation, Holiday junction resolution and heteroduplex resolution (Fig. 4A). The Holiday junction can be resolved by crossover (CO), resulting in gene conversion with a chromosomal recombination, or by non-crossover (NCO), resulting in gene conversion only (Fig. 4A). In addition the resolution of heteroduplex DNA can result in mosaic LOH patterns due to a combination of gene conversion and some restoration (Fig. 4A). Such mosaic patterns can also result from template switching during repair synthesis. Of the strains sequenced in this study, IMX1606-IMX1615 and IMX1626-IMX1635 lost heterozygosity only in the region surrounding the targeted DSB, indicating HR had occurred (Fig. 3). In these strains, mosaic patterns resulting from heteroduplex resolution were observed in strains IMX1606, IMX1608 and IMX1613 (Fig. 4C and Table S2). Since strains IMX1596-IMX1605 and strains IMX1616-1625 lost heterozygosity of entire chromosome arms (Fig. 3), repair could have occurred by BIR. However, mosaic patterns corresponding to heteroduplex resolution were observed in strains IMX1605 and IMX1619 (Fig. 4C and Table S2). While BIR does not cause mosaic LOH, chromosome-arm LOH is not a commonly-recognized result of HR (Fig. 4A) (10-12). Therefore, we propose a repair mechanism that involves HR of at least one of the targeted chromatids at the G<sub>2</sub> stage of the cell cycle (Fig. 4B). The proposed mechanism would result in daughter cells with either local LOH or chromosome-arm LOH, with and without mosaic heterozygosity at the targeted locus. The proposed mechanism is consistent with all phenotypes and genotypes encountered in this study as well as in previous studies involving

hemizygous introduction of DSBs (9, 13-16, 32, 36). While HR at the G<sub>2</sub> stage of the cell cycle could explain all observed genotypes, HR at the G<sub>1</sub> stage and BIR could also contribute.



**Figure 4. Proposed mechanism for Cas9-mediated loss of heterozygosity in a diploid genome based on homologous recombination (HR) between homologous chromosomes.** (A) Possible outcomes of HR in cells with one chromosome complement during the G<sub>1</sub> stage of the cell cycle. (B) Possible outcomes of HR in cells with two chromosome complements during the G<sub>2</sub> stage of the cell cycle. The targeted chromosome (red), its homolog (blue) and the centromere are indicated (black, where relevant). Newly synthesized DNA is shown in a lighter shade. Heteroduplex resolution occurs prior to chromatid segregation and the strand with the targeted NGG PAM sequence is always discarded due to Cas9 activity. For HR during the G<sub>2</sub> stage of the cell cycle, HR occurs between one chromatid of the targeted and one chromatid of the non-targeted chromosome, as with HR during the G<sub>1</sub> stage. The chromatids subsequently segregate according to their centromere pairing, with one red and one blue centromere in each daughter cell. For simplicity, only repair of one targeted chromatid is shown in the figure. Repair of both targeted sister chromatids results in the same genome alterations as shown here. As indicated in the figure, HR during the G<sub>2</sub> stage of the cell cycle could yield local as well as chromosome-arm LOH by mitotic crossover, both with and without mosaic structures. (C) Mosaic loss of heterozygosity at the targeted loci in single isolates obtained by targeting *UTR2* and *FIR1* in IMX1585. Strains with mosaic loss of heterozygosity are at the *UTR2* locus (left) and at the *FIR1* locus (right) are indicated. The location of the ORFs is indicated in IMX1585 (grey). For each strain, heterozygous sequence is indicated in red. Exact coordinates of heterozygous nucleotides are indicated in Table S2.

## DISCUSSION

The efficiency of gene editing using Cas9 can decrease by almost two orders of magnitude when targeting only one of two homologous chromosomes due to a competing repair mechanism causing either local or chromosome-arm scale LOH. In previous work, cas9-mediated gene editing was reported to cause large deletions at the targeted loci (37), which sometimes resulted in loss of heterozygosity. Here, the observed LOH consisted not only of loss of genetic material from the

targeted chromosome, but also of replacement of the affected sequence by an additional copy of sequence homologous to the targeted site. While such LOH upon introduction of a hemizygous DSB has been observed in the yeasts *S. cerevisiae* and *Candida albicans* (9, 13), this study demonstrates that repair by LOH is not only possible, but occurs at rates which impede gene editing approaches based on integration of repair fragments. Gene-editing was similarly inhibited in an *S. cerevisiae* diploid with 99 % homozygosity and in an interspecies *S. cerevisiae* x *S. eubayanus* hybrid with 85 % homozygosity. In addition, the recombination events occurred at loci with homologies as low as 80 %. The lack of necessity for high identity suggests that Cas9-mediated gene editing may also cause LOH by translocations resulting from recombination events between paralogous genes. However, such translocations were not observed in this study, and Cas9-mediated gene-editing has been applied successfully to delete various paralogs without resulting in translocations (38, 39).

Cas9-mediated LOH is likely to contribute to a lesser genome accessibility of heterozygous yeasts relative to laboratory strains, which tend to be haploid or homozygous. Therefore, these results are likely to affect the genome editing of hybrids, industrial yeasts and natural isolates due to their frequent heterozygosity (40), and should be used to update guidelines for designing gene editing strategies. We strongly recommend to design gRNAs targeting homozygous nucleotides stretches when targeting heterozygous genomes. When allele-specific gene editing is required, we recommend the use of repair fragments with integration markers such as the Venus fluorophore in this study, since accurate gene editing is not impossible, simply inefficient. When the use of a marker is not permissible, extensive screening of transformants for correct gene editing may be required.

While the HDR machinery is well conserved in eukaryotes (11, 12), further research is required to determine if LOH occurs at similar rates in eukaryotes other than *S. cerevisiae*, and if it impedes gene editing. While DSB-mediated LOH was observed in *S. cerevisiae*, *C. albicans*, *Drosophila melanogaster* and *Mus musculus* (9, 13, 32, 36), relative contributions of HR, BIR and NHEJ to DSB repair vary across species. However, since integration of a repair fragment and repair by LOH both involve HR (41, 42), targeting heterozygous loci likely causes low gene-editing efficiencies and LOH in other eukaryotes as well, regardless of the efficiency of NHEJ.

Targeting of heterozygous loci is common in gene editing, for example during allele propagation of gene drives and disease allele correction in human gene therapy (41, 42). Although gene drives are based on LOH by HR (42), the extent of LOH beyond the targeted locus has not been systematically studied but could, by analogy with the present study, potentially affect entire chromosome arms. Allele-specific gene editing generally aims at repair by HR using a co-transformed repair fragment instead of a homologous chromosome. Reports of LOH after targeting a heterozygous allele in human embryos despite availability of an adequate repair fragment, are consistent with Cas9-induced LOH extending beyond the targeted locus, as described here (41). While, in the human-embryo study, repair by LOH was perceived as a success, the reported role of LOH in cancer development (43) indicates that large-scale LOH can have important phenotypic repercussions. Therefore we recommend avoiding allele-specific gene editing when possible until further research determines if it is a risk in other eukaryotes. Based on the proposed HR mechanism for CRISPR/Cas9-mediated LOH (Fig. 4B), the risk of LOH can be mitigated by designing gRNAs that cut all alleles of heterozygous loci, even if only a single allele needs to be edited. Eventually, CRISPR-Cas9 editing could become safer by favouring DSB-independent gene-editing methods such as guided nickases and base-editing strategies for preventing or reducing the incidence of LOH (44-47).

## ACKNOWLEDGEMENT

ARGdV conceived the study and designed the experiments. ARGdV and LGFC performed plasmid and strain construction. ARGdV, LGFC, PdITC and JtH performed the experimental work. ARGdV and MvdB performed bioinformatics analysis. ARGdV, JTP and JMGD supervised the study and wrote the manuscript. All authors read and approved the final manuscript.

We thank Liset Jansen for drawing our attention to the difficulty to edit a heterozygous gene, Robert Mans for his expertise with gene editing in *Saccharomyces cerevisiae*, Melanie Wijsman for constructing and Pascale Daran-Lapujade for sharing plasmids pUDE480, pUDE481 and pUDE482, Sai T. Reddy for his insights in the potential impact for human gene therapy and Nick Brouwers, Alex Salazar, Xavier D. V. Hakkaart, Ioannis Papapetridis, Niels G.A. Kuijpers, Jan-Maarten Geertman and Thomas Abeel for their critical input.

## REFERENCES

1. Sander JD & Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32(4):347.
2. Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823-826.
3. Mans R, et al. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15(2).
4. DiCarlo JE, et al. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41(7):4336-4343.
5. Klein M, Eslami-Mossallam B, Arroyo DG, & Depken M (2018) Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. *Cell Rep.* 22(6):1413-1423.
6. Jasin M & Rothstein R (2013) Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* 5(11):a012740.
7. Kuijpers NGA, et al. (2016) Pathway swapping: Toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. USA* 113(52):15060-15065.
8. Gorter de Vries AR, de Groot PA, van den Broek M, & Daran J-MG (2017) CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb. Cell. Fact.* 16(1):222.
9. Sadhu MJ, Bloom JS, Day L, & Kruglyak L (2016) CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science* 352(6289):1113-1116.
10. Li X & Heyer W-D (2008) Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.* 18(1):99.
11. Haber JE (2000) Partners and pathways: repairing a double-strand break. *Trends Genet.* 16(6):259-264.
12. Moynahan ME, Chiu JW, Koller BH, & Jasin M (1999) Brca1 controls homology-directed DNA repair. *Mol. Cell* 4(4):511-518.
13. Feri A, et al. (2016) Analysis of repair mechanisms following an induced double-strand break uncovers recessive deleterious alleles in the *Candida albicans* diploid genome. *MBio* 7(5):e01109-01116.
14. Charles JS, et al. (2012) High-resolution Genome-wide Analysis of Irradiated (UV and gamma rays) Diploid Yeast Cells Reveals a High Frequency of Genomic Loss of Heterozygosity (LOH) Events. *Genetics:genetics.* 111.137927.
15. Argueso JL, et al. (2008) Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc. Natl. Acad. Sci. USA* 105(33):11845-11850.
16. Hum YF & Jinks-Robertson S (2017) Mitotic gene conversion tracts associated with repair of a defined double-strand break in *Saccharomyces cerevisiae*. *Genetics:genetics.* 300057.302017.
17. Salazar AN, et al. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN. PK113-7D. *FEMS Yeast Res.* 17(7).
18. Goffeau A, et al. (1996) Life with 6000 genes. *Science* 274(5287):546-567.
19. Verduyn C, Postma E, Scheffers WA, & van Dijken JP (1990) Physiology of *Saccharomyces Cerevisiae* in anaerobic glucose-limited chemostat cultures. *J. Gen. Microbiol.* 136(3):395-403.
20. Pronk JT (2002) Auxotrophic yeast strains in fundamental and applied research. *Appl. Environ. Microbiol.* 68(5):2095-2100.
21. Solis-Escalante D, et al. (2013) amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 13(1):126-139.
22. Lee ME, DeLoache WC, Cervantes B, & Dueber JE (2015) A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* 4(9):975-986.
23. Mans R, Wijsman M, Daran-Lapujade P, & Daran J-MG (2018) A protocol for introduction of multiple genetic modifications in *Saccharomyces cerevisiae* using CRISPR/Cas9. *FEMS Yeast Res.*
24. Engler C, Kandzia R, & Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS one* 3(11):e3647.

25. Gietz RD & Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.*, (Elsevier), Vol 350, pp 87-96.
26. Li H & Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26(5):589-595.
27. Li H, et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
28. Walker BJ, et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS one* 9(11):e112963.
29. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat. Biotechnol.* 29(1):24.
30. Brickwedde A, et al. (2018) Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357<sup>T</sup>. *Front. Microbiol.* 9:1786.
31. Heblly M, et al. (2015) *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res.* 15(3).
32. Heinze SD, et al. (2017) CRISPR-Cas9 targeted disruption of the yellow ortholog in the housefly identifies the brown body locus. *Sci. Rep.* 7(1):4582.
33. Chu VT, et al. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33(5):543.
34. Jasin M & Haber JE (2016) The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA repair* 44:6-16.
35. Llorente B, Smith CE, & Symington LS (2008) Break-induced replication: what is it and what is it for? *Cell cycle* 7(7):859-864.
36. Henson V, Palmer L, Banks S, Nadeau JH, & Carlson GA (1991) Loss of heterozygosity and mitotic linkage maps in the mouse. *Proc. Natl. Acad. Sci. USA* 88(15):6486-6490.
37. Kosicki M, Tomberg K, & Bradley A (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*
38. Wijsman M, et al. (2018) A toolkit for rapid CRISPR-SpCas9 assisted construction of hexose-transport-deficient *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.*
39. Marques WL, et al. (2017) Elimination of sucrose transport and hydrolysis in *Saccharomyces cerevisiae*: a platform strain for engineering sucrose metabolism. *FEMS Yeast Res.* 17(1):fox006.
40. Salazar AN, et al. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res* 17(7).
41. Ma H, et al. (2017) Correction of a pathogenic gene mutation in human embryos. *Nature* 548(7668):413-419.
42. Champer J, Buchman A, & Akbari OS (2016) Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* 17(3):146.
43. Naylor SL, Johnson BE, Minna JD, & Sakaguchi AY (1987) Loss of heterozygosity of chromosome 3p markers in small-cell lung cancer. *Nature* 329(6138):451.
44. Kim K, et al. (2017) Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* 35(5):435.
45. Komor AC, Kim YB, Packer MS, Zuris JA, & Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533(7603):420.
46. Shen B, et al. (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* 11(4):399.
47. Ran FA, et al. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154(6):1380-1389.

## Chapter 7: Phenotype-independent isolation of interspecies *Saccharomyces* hybrids by dual-dye fluorescent staining and fluorescence-activated cell sorting

Arthur R. Gorter de Vries <sup>#</sup>, Charlotte C. Koster <sup>#</sup>, Susan M. Weening, Marijke A. H. Luttik, Niels G.A. Kuijpers, Jan-Maarten A. Geertman, Jack T. Pronk and Jean-Marc G. Daran

<sup>#</sup> These authors contributed equally to this publication and should be considered co-first authors.

Interspecies hybrids of *Saccharomyces* species are found in a variety of industrial environments and often outperform their parental strains in industrial fermentation processes. Interspecies hybridisation is therefore increasingly considered as an approach for improvement and diversification of yeast strains for industrial application. However, current hybridisation methods are limited by their reliance on pre-existing or introduced selectable phenotypes.

This study presents a high-throughput phenotype-independent method for isolation of interspecies *Saccharomyces* hybrids based on dual dye-staining and subsequent mating of two strains, followed by isolation of double-stained hybrid cells from a mating population by fluorescence-activated cell sorting (FACS). Pilot experiments on intra-species mating of heterothallic haploid *S. cerevisiae* strains showed that 80 % of sorted double-stained cells were hybrids. The protocol was further optimized by mating an *S. cerevisiae* haploid with homothallic *S. eubayanus* spores with complementary selectable phenotypes. In crosses without selectable phenotype, using *S. cerevisiae* and *S. eubayanus* haploids derived from laboratory as well as industrial strains, 10 to 15 % of double-stained cells isolated by FACS were hybrids. When applied to rare mating, sorting of double-stained cells consistently resulted in about 600-fold enrichment of hybrid cells.

Mating of dual-stained cells and FACS-based selection allows efficient enrichment of interspecies *Saccharomyces* hybrids within a matter of days and without requiring selectable hybrid phenotypes, both for homothallic and heterothallic *Saccharomyces* species. This strategy should accelerate generation of laboratory-made hybrids, facilitate research into hybrid heterosis and offer new opportunities for non-GMO industrial strain improvement and diversification.

Essentially as published in *Frontiers in Microbiology* 2019;10:871

Supplementary materials available online

<https://www.frontiersin.org/articles/10.3389/fmicb.2019.00871/full>

## Introduction

*Saccharomyces* yeasts are used in various biotechnological industries including beer brewing, wine making, biopharmaceutical protein synthesis and biofuels production (1-5). Nine *Saccharomyces* species have currently been described (6, 7), which are separated by a postzygotic barrier that causes interspecies hybrids to be mostly sterile (8, 9). Although interspecies hybrids occur in natural contexts such as the guts of wasps (10), *Saccharomyces* hybrids are most commonly found in domesticated environments (11, 12). For instance, lager beer is brewed by the *S. cerevisiae* x *S. eubayanus* hybrids, collectively indicated as *S. pastorianus* (13), *S. uvarum* x *S. eubayanus* hybrids called *S. bayanus* are used for cider brewing (14), and various double and triple hybrids between *S. cerevisiae*, *S. kudriavzevii* and *S. uvarum* play an important role in aroma production during wine fermentation (15). In addition, interspecies hybridisation followed by rounds of backcrossing contributed to the evolution of domesticated *Saccharomyces* strains by facilitating horizontal gene transfer (16). Such introgressions contribute to the distinct phenotypes of, for instance, cider-fermenting *S. uvarum* strains and wine-fermenting *S. cerevisiae* strains (17, 18).

Hybrid physiology largely depends on the specific parental strains (19, 20). The genomes of hybrids from different *Saccharomyces* species have been shown to act synergistically, a phenomenon called 'heterosis' or 'hybrid vigour', in which a hybrid performs better than either of its parents in specific environments (21). Heterosis is a complex phenomenon, involving copy number effects, interactions between different dominant and recessive alleles, and epistatic interactions (12, 21, 22). While some traits such as cryotolerance or flocculation appear to be completely inherited from one of the parental strains (23, 24), hybrids can also show phenotypes intermediary to their parental strains, as has been demonstrated for production of flavour compounds and other metabolites (25, 26).

Due to hybrid vigour, the generation of *Saccharomyces* hybrids can yield strains with novel or improved properties for industrial applications. For instance, laboratory-made *S. cerevisiae* x *S. eubayanus* hybrids displayed increased cold tolerance, faster oligosaccharide consumption, different flavour profiles, higher fermentation rates and higher ethanol titres than their parental strains (24, 25, 27). Pioneering studies on reconstruction of naturally occurring hybrids have inspired the generation of novel interspecies hybrids, such as *S. cerevisiae* x *S. paradoxus* hybrids (26), *S. cerevisiae* x *S. mikatae* hybrids (28, 29), *S. cerevisiae* x *S. arboricola* hybrids (29) and *S. cerevisiae* x *S. uvarum* hybrids (30, 31). Their phenotypic diversity showed promise for applications ranging from the fermented beverage industry to the production of biofuels (29, 32-34).

Analogous to intra-species mating, interspecies hybridization occurs either by mating haploid cells of opposite mating type, or by rare mating based on spontaneous mating-type switching caused by loss of heterozygosity at the *MAT* locus (32). However, interspecies hybridisation occurs at a relatively low rate; reported hybridisation frequencies range from 1.5 to 3.6 % for spore-to-spore mating (20, 22) to frequencies as low as  $1 \times 10^{-6}$  to  $1 \times 10^{-8}$  for rare mating (22, 35). While the efficiency of interspecies mating can be improved by genetic modification (GM) techniques, for example by overexpression of HO-endonuclease (36), isolation of hybrids from mating cultures remains necessary.

When parental strains have different selectable phenotypes, hybrids can be isolated by transferring the mating culture to conditions requiring both phenotypes for growth. Selectable phenotypes such as auxotrophies can either occur naturally (37, 38), or they can be obtained by mutagenesis and/or

laboratory evolution under conditions favouring auxotrophic strains (25, 39, 40). However, generation of auxotrophic mutants is time- and labour-intensive (36) and can be further complicated by the polyploid or aneuploid nature of many industrially-relevant *Saccharomyces* strains (12, 39, 41). Alternatively, selectable phenotypes such as antibiotic resistance can be introduced using GM techniques (24, 42, 43). However, industrial strains can be resilient to genetic modification, and customer acceptance and legislation issues still largely preclude use of GM technology for applications in the food and beverages industry (44, 45).

Fluorescence-activated cell sorting (FACS) can be used to isolate fluorescent cells from populations, even if they occur at extremely low frequencies (46). By labelling each parental strain with a fluorescent dye, FACS has previously been used to sort mated *Saccharomyces cerevisiae* cells from their mating culture, resulting in a threefold enrichment of mated cells (41). Although a three-fold enrichment would not be sufficient to isolate interspecies hybrids from a mating culture, this early study raised the question whether it might be possible to sufficiently modify staining, mating and FACS procedures to accomplish this goal. To address this question, we explored a method to isolate interspecies *Saccharomyces* hybrids based on dual fluorescent labelling of parental strains and subsequent FACS-based selection of double-stained cells, without any dependency on any selectable phenotypes. After reproducing the isolation of intra-species *S. cerevisiae* crosses, we optimized isolation of interspecies *S. cerevisiae* x *S. eubayanus* hybrids using strains with selectable phenotypes. The resulting method was then tested for phenotype-independent isolation of *S. cerevisiae* x *S. eubayanus* hybrids.

## Materials and Methods

### Strains, media and cultivation

*S. cerevisiae* and *S. eubayanus* strains used in this study are listed in table 1. Strains were routinely grown in complex medium (YP), containing 10 g L<sup>-1</sup> yeast extract and 20 g L<sup>-1</sup> peptone, supplemented with 20 g L<sup>-1</sup> glucose for YPD, and with 20 g L<sup>-1</sup> trehalose for YPT. Synthetic medium (SM) containing 20 g L<sup>-1</sup> glucose, 3 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 5.0 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7 H<sub>2</sub>O, 1 mL L<sup>-1</sup> of a trace element solution and 1 mL L<sup>-1</sup> of a vitamin solution, was prepared as described previously (47), and the pH was set to 6.0 using 2 M KOH. Presence of the KanMX marker cassette was selected for in SM+G418: SM supplemented with 0.2 g L<sup>-1</sup> of G418 (Invitrogen, Carlsbad, CA) in which (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was replaced by 1 g L<sup>-1</sup> monosodium glutamate (48). For solid media, 20 g L<sup>-1</sup> agar was added to media. Strains were grown in 500 mL round-bottom shake flasks with 100 mL medium at 200 RPM in an Innova 44 incubator shaker (Eppendorf, Nijmegen, the Netherlands). Cultures of *S. cerevisiae* and *S. eubayanus* were grown at 30 °C and at 20 °C, respectively. Liquid sporulation medium contained 20 g L<sup>-1</sup> potassium acetate and its pH set to 7.0 using acetic acid (49). Frozen stocks were prepared by addition of glycerol (30 % v/v) to exponentially growing shake-flask cultures, after which 1 mL aliquots were aseptically stored at -80 °C.



**Table 1. *Saccharomyces* strains used in this study.**

Name	Species	Parental strain(s)	Relevant genotype	Origin
CEN.PK113-5A	<i>S. cerevisiae</i>	-	MATa URA3 his3-Δ1 leu2-3,112 trp1-289	(59)
IMK439	<i>S. cerevisiae</i>	CEN.PK113-1A	MATa HIS3 LEU2 TRP1 ura3Δ::KanMX	(71)
IMK440	<i>S. cerevisiae</i>	CEN.PK113-7D	MATa HIS3 LEU2 TRP1 ura3Δ::KanMX	(71)
CEN.PK122	<i>S. cerevisiae</i>	-	MATa/MATa URA3/URA3	(59)
CBS 12357	<i>S. eubayanus</i>	-	MATa/MATa	(13)
CEN.PK113-7D	<i>S. cerevisiae</i>	-	MATa	(59, 72)
IMS0408	<i>S. eubayanus</i> × <i>S. cerevisiae</i>	CBS 12357 × IMK439	MATa/MATa SeubURA3/Scura3Δ::KanMX	(24)
CDFM21L.1	<i>S. eubayanus</i>	-	MATa/MATa	Kindly provided by F.-Y. Bai, Chinese Academy of Science (60)
Ale28	<i>S. cerevisiae</i>	-	MATa/MATa	Kindly donated by HEINEKEN Supply Chain, Zoeterwoude, The Netherlands
IMX1471	<i>S. cerevisiae</i>	IMK439 × IMK440	MATa/MATa ura3Δ::KanMX/ura3Δ::KanMX	This study
IMH001	<i>S. eubayanus</i> × <i>S. cerevisiae</i>	CBS 12357 × CEN.PK113-7D	MATa/MATa	This study
IMH002	<i>S. eubayanus</i> × <i>S. cerevisiae</i>	CBS 12357 × CEN.PK113-7D	MATa/MATa	This study
IMH003	<i>S. eubayanus</i> × <i>S. cerevisiae</i>	CDFM21L.1 × Ale28	MATa/MATa	This study
IMH004	<i>S. eubayanus</i> × <i>S. cerevisiae</i>	CDFM21L.1 × Ale28	MATa/MATa	This study
IMH005	<i>S. eubayanus</i> × <i>S. cerevisiae</i>	CDFM21L.1 × Ale28	MATa/MATa	This study
IMH006	<i>S. eubayanus</i> × <i>S. cerevisiae</i>	CDFM21L.1 × Ale28	MATa/MATa	This study
IMH007	<i>S. eubayanus</i> × <i>S. cerevisiae</i>	CDFM21L.1 × Ale28	MATa/MATa	This study

### **Sporulation, spore isolation and germination**

Sporulation was performed by aerobic incubation at 20 °C during at least 72 h on sporulation medium. Presence of asci was verified using phase-contrast microscopy at a magnification of 400x. Spores were isolated as described by Herman and Rine (50) with minor modifications. In short, spores were pelleted (1000 g, 5 min), resuspended in softening buffer (10 mM dithiothreitol, 100 mM Tris-SO<sub>4</sub>, pH set to 9.4 with H<sub>2</sub>SO<sub>4</sub>) and incubated at 30 °C for 10 min. Cells were then washed using demineralized water, resuspended in spheroplasting buffer (2.1 M sorbitol, 10 mM KH<sub>2</sub>PO<sub>4</sub>, pH set to 7.2 with 1M NaOH, 0.8 g L<sup>-1</sup> zymolyase 20-T (AMS Biotechnology Ltd., Abingdon, United Kingdom)) and incubated overnight at 30 °C. After incubation, the culture was pelleted (1000 g, 10 min), washed with demineralized water and resuspended in 0.5 % Triton X-100 (Sigma-Aldrich, Zwijndrecht, the Netherlands). Spores were then sonicated for 15 s at 50 Hz with an amplitude of 6 micron while kept on ice using a Soniprep 150 (MSE, London, United Kingdom). During initial optimization of the protocol, a short protocol with only the zymolyase-step was also tested. Isolation of spores was confirmed by microscopic inspection as described above. For germination, spores were washed once with YPD and subsequently resuspended in 20 mL YPD to a concentration of approximately 10<sup>6</sup> cells mL<sup>-1</sup>. The germination culture was incubated in a 100 mL round bottom flask at 30 °C and 200 RPM for 5 h. A protocol using different incubation times in 2 % glucose medium and in YPD was tested during initial optimization of the interspecies mating.

### **Staining of *Saccharomyces* cultures**

For staining, CellTrace™ Violet, CellTrace™ CFSE and CellTrace™ Far Red fluorescent dyes (Thermo Fisher Scientific, Waltham, MA) were prepared according to the manufacturers' recommendations. Cultures were stained with 2 µL CellTrace™ dye per mL culture and incubated overnight in the dark at 12 °C and 200 RPM. Stained cultures were washed twice with YP medium, as remaining unbound dye molecules would bind to the amide groups in yeast extract and peptone.

### **Intra-species mating**

Heterothallic haploid parental strains were propagated until mid-exponential phase. The cultures of two parental strains were washed and diluted in sterile Isoton II (Beckman Coulter, Woerden, NL) to a final cell density of approximately 10<sup>6</sup> cells mL<sup>-1</sup> and stained with CellTrace™ Violet and CellTrace™ CFSE. 100 µL of each stained culture was transferred to an Eppendorf tube and centrifuged briefly (2000 g, 1 min) to increase proximity of the cells for more efficient mating. Subsequently, the mating culture was statically incubated at 12 °C in the dark until FACS analysis.

### **Interspecies mating and rare mating**

Diploid parental strains were propagated until mid-exponential phase. Haploid parental cells from homothallic strains were obtained via sporulation, spore isolation and germination. Cells were washed and diluted in sterile Isoton II (Beckman Coulter) to a final cell density of approximately 10<sup>6</sup> cells mL<sup>-1</sup> and stained with CellTrace™ Violet and CellTrace™ CFSE as described. For rare mating, a final cell density of approximately 2 × 10<sup>7</sup> cells mL<sup>-1</sup> was used and cells were stained with CellTrace™ Far Red and CellTrace™ CFSE as described. 100 µL of each stained culture was transferred to an Eppendorf tube and centrifuged briefly (2000 g, 1 min) to increase proximity of the cells for more efficient mating. Subsequently, the mating culture was statically incubated at 12 °C in the dark until FACS analysis.

### **FACS analysis and sorting**

Cultures for FACS analysis and sorting were diluted in sterile Isoton II and vortexed thoroughly to disrupt cell aggregates. For rare mating, 50 mM EDTA was added to disrupt cell aggregates formed by flocculation. The cultures were analysed on a BD FACSAria™ II SORP Cell Sorter (BD Biosciences, Franklin Lakes, NJ) equipped with 355 nm, 445 nm, 488 nm, 561 nm and 640 nm lasers and a 70 µm nozzle, and FACSTream™ sheath fluid (BD Biosciences). Correct cytometer performance was evaluated prior to each experiment by running a Cytometer Setup & Tracking cycle using a CS&T bead kit (BD Biosciences) for calibration. Drop delay for sorting was determined by running an Auto Drop Delay cycle using Accudrop Beads (BD Biosciences). CellTrace™ Violet fluorescence was excited by the 355 nm laser and emission was detected through a 450 nm bandpass filter with a bandwidth of 50 nm. CellTrace™ CFSE was excited by the 488 nm laser and emission was detected through a 545 nm bandpass filter with a bandwidth of 30 nm. CellTrace™ Far Red was excited by the 640 nm laser and emission was detected through a 780 nm bandpass filter with a bandwidth of 60 nm. Morphology of the cells was analysed by plotting forward scatter (FSC) against side scatter (SSC). Fluorescence of mating cultures was analysed on either a CFSE versus Violet or a CFSE versus Far Red plot. Prior to sorting, at least 10<sup>5</sup> events were analysed. Sorting regions ('gates') were set on these plots to determine the types of cells to be sorted. Gated single cells were sorted in 96-well microtiter plates containing YPD using a "single cell" sorting mask (0/32/16), and the plates were incubated at RT for 2 days. When cells with selectable phenotypes were used, the fraction of mated cells was determined by replica-plating to 96-well plates with selective medium (SM or SM+G418), using an ethanol-flame sterilized 96-pin replicator. FACS data were analysed using FlowJo® software (version 3.05230, FlowJo, LLC, Ashland, OR).

### **Determination of the fraction of growing cells**

After FACS sorting, the fraction of growing cells was determined by counting the number of wells in which growth was observed. For populations with low viabilities, up to 1000 cells were sorted per well and Poisson statistics were used to estimate the fraction of growing cells (51). The fraction of growing cells was calculated from (P), the fraction of wells containing a colony, (W) the total number of wells and (n), the total number of cells sorted into the wells (Equation 1).

$$\text{Fraction of growing cells} = \frac{-\ln(1 - P) * W}{n}$$

Equation 1

### **Imaging**

Cells were imaged using a Zeiss Axio Imager Z1 (Carl Zeiss AG, Oberkochen, Germany). For fluorescent imaging, cells were excited with a xenon lamp. Fluorescence from CellTrace™ CFSE was imaged through a GFP filter set (Carl Zeiss AG) containing a 470 nm bandpass excitation filter with a bandwidth of 20 nm and a 540 nm emission filter with a bandwidth of 25 nm. CellTrace™ Far Red was imaged through a Cy5 filter set (Carl Zeiss AG) containing a 640 nm bandpass excitation filter with a bandwidth of 30 nm and a 690 nm emission filter with a bandwidth of 50 nm. Images were processed using AxioVision SE64 (Rel. 4.9.1. Carl Zeiss AG) and FIJI (52).

### **Ploidy determination by flow cytometry**

For ploidy determination, samples were fixed using ethanol as previously described (24). Staining of cells with SYTOX® Green Nucleic Acid Stain (Invitrogen) was performed as previously described (53)

with some minor modifications. Cells were washed in 50 mM Tris-Cl (pH 7.5) and resuspended in 100  $\mu$ L RNase solution (1 mg mL<sup>-1</sup> RNase A in 50 mM Tris-Cl). 100  $\mu$ L of cells was added to 1 mL of SYTOX<sup>®</sup> Green solution. When processing large numbers of samples, a high-throughput protocol in 96-well microtiter plates was used with a PIPETMAN<sup>®</sup> M multichannel electronic pipette (Gilson, Middleton, WI, USA). In this modified protocol, 100  $\mu$ L sample was fixated by adding 150  $\mu$ L 70 % ethanol and in the final step 20  $\mu$ L sample was added to 180  $\mu$ L SYTOX<sup>®</sup> Green solution. An unstained control was included along with every sample. Fluorescence of the samples was measured on a BD Accuri<sup>™</sup> C6 CSampler Flow Cytometer (BD Biosciences). The fluorophore was excited with the 488 nm laser of the flow cytometer and emission was detected through a 533 nm bandpass filter with a bandwidth of 30 nm. Ploidy data was analysed using FlowJo<sup>®</sup> software (version 3.05230, FlowJo).

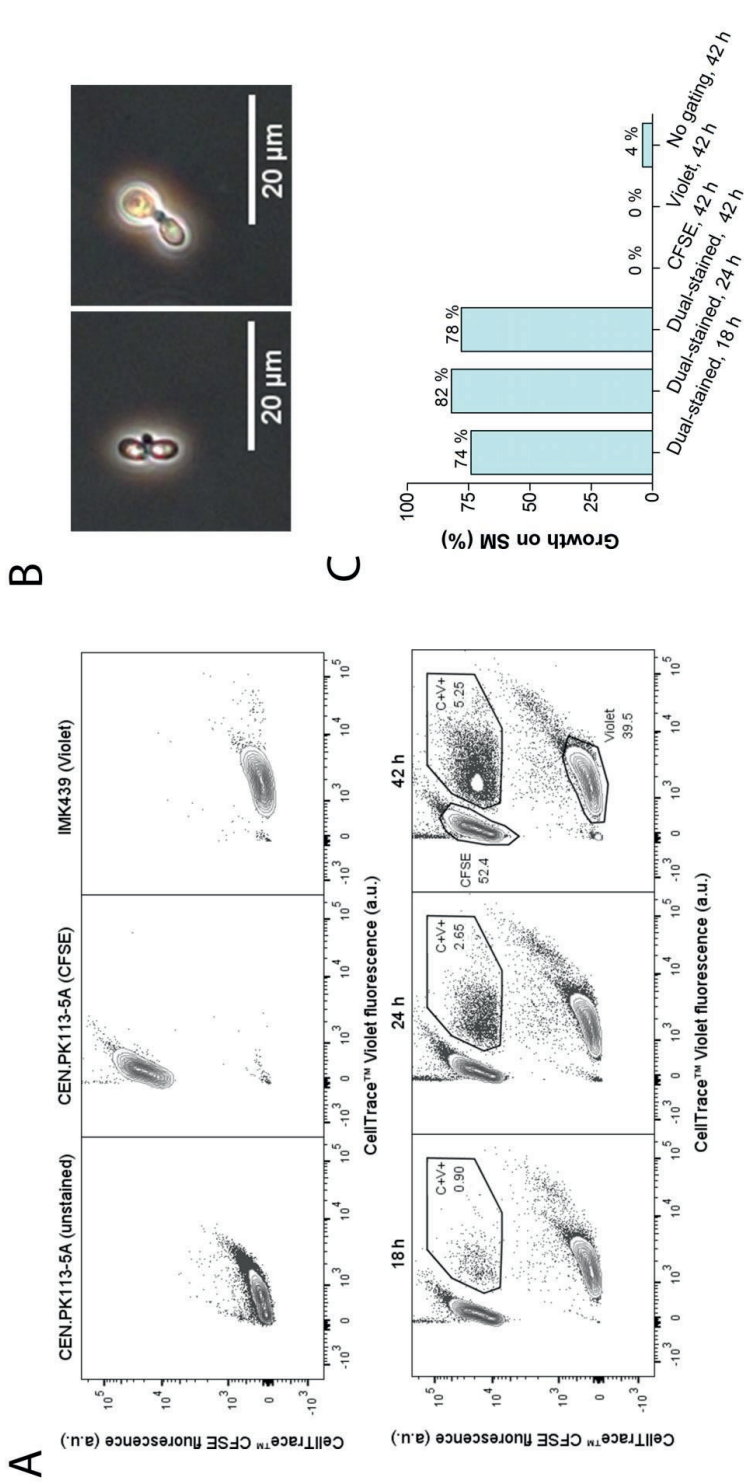
#### Identification of interspecies hybrids by PCR

The presence of genetic material from *S. cerevisiae* and from *S. eubayanus* and the mating type of potential hybrids was verified by PCR using DreamTaq PCR Mastermix (Life Technologies), as described previously (24). DNA was released by boiling 2  $\mu$ L of a liquid culture in 2  $\mu$ L of NaOH for 15 min at 99°C. The *S. cerevisiae*-specific *MEX67* gene was amplified using primers 8570 & 8571 (54Error! Reference source not found.) and the *S. eubayanus*-specific *FSY1* gene was amplified using primers 8572 & 8573 (54, 55). The mating type was determined by amplifying *MAT*-loci, using primers 11, 12 & 13 (56). PCR products were separated on a 2 % (w/v) agarose gel in 0.5X TBE buffer (45 mM Tris-borate, 1 mM EDTA, pH 8).

## Results

#### Isolating intra-species hybrids from a mating culture using FACS

A functional protocol for dual staining of parental strains, mating and FACS-based sorting of double-stained cells was developed using the heterothallic haploid *S. cerevisiae* strains CEN.PK113-5A (*MAT $\alpha$* , His<sup>-</sup>, Lys<sup>-</sup>, Trp<sup>-</sup>) and IMK439 (*MAT $\alpha$* , Ura<sup>-</sup>). Due to their complementary auxotrophies, the fraction of mated cells could easily be quantified before and after FACS-based selection of double-stained cells by measuring the ability to grow on synthetic medium without histidine, lysine, tryptophan and uracil. CEN.PK113-5A and IMK439 were stained using the commercially-available fluorescent CellTrace<sup>™</sup> dyes CFSE and Violet, respectively. These dyes covalently bind to amine groups and thereby irreversibly label the parental cells (57). Mated cells should then be identifiable by the presence of fluorescent material from both parental strains. Efficient staining of the parental strains was confirmed for both dyes using flow cytometry (Figure 1A). To minimize dilution of the dye due to cell division, stained cells were mated by co-incubation in YPT medium at 12 °C, which resulted in slow growth of *S. cerevisiae*. Flow cytometry of the mating culture indicated a progressive increase of the incidence of double-stained cells, from 0.90 % after 18 h to 5.25 % after 42 h (Figure 1A). Approximately 10 % double-stained cells exhibited a “shmoo” morphology (Figure 1B) characteristic of *Saccharomyces* zygotes (58). To determine the fraction of mated cells, cells from the total population and from the double-stained population were sorted on SM using FACS. Only 4 % of the total population was able to grow on SM, while 74 - 82 % of double-stained cells grew on this medium, indicating a 20-fold enrichment of mated cells in the double-stained population (Figure 1C).



**Figure 1. Intra-species mating of *S. cerevisiae* strains CEN.PK113-5A (*MAT $\alpha$  URA3 his3- $\Delta$ 1 leu2-3,112 trp1-289*) and IMK439 (*MAT $\alpha$  HIS3 TRP1 LEU2 ura3 $\Delta$ ::KanMX*).** (A) Fluorescence contour plots of unstained CEN.PK113-5A, CEN.PK113-5A stained with CellTrace™ CFSE, IMK439 stained with CellTrace™ Violet, and of the mating culture after 18, 24 and 42 h. The indicated gated areas were used for sorting cells, event rates of each gate are indicated as a percentage of total cell counts. (B) Microscope image (400 $\times$ ) of zygotes sorted from the double-stained population (C+V+) after 42 h of mating. (C) Percentage of cells able to grow on synthetic medium without auxotrophy-complementing supplements in different populations sorted by FACS, as indicated in panel A.

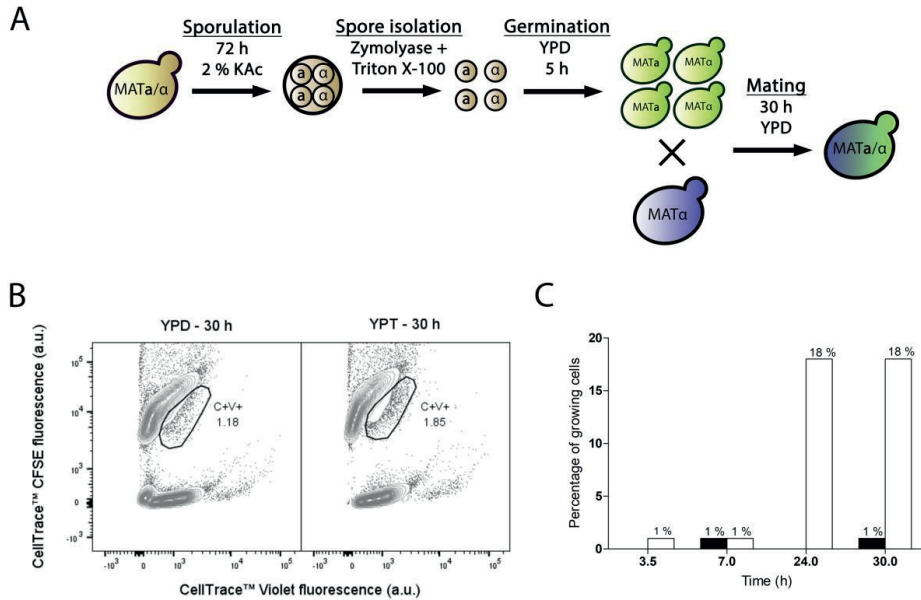
### Isolation of interspecies hybrids from a mating culture using FACS

The developed protocol was applied to obtain interspecies hybrids between *S. eubayanus* strain CBS 12357 and *S. cerevisiae* strain IMK439 (*MAT $\alpha$* , *ura3 $\Delta$ ::KanMX*). Hybrids of these strains can be easily identified due to combined uracil prototrophy and resistance to the antibiotic G418. As CBS 12357 is homothallic strain it was sporulated prior to staining and mating. Since homodiploidization by self-mating of spores could compete with interspecies mating, a protocol for digestion of the ascus sack was developed based on a combined treatment with the surfactant Triton X-100 and zymolyase (50) (Supplementary Figure 1). When staining isolated spores of *S. eubayanus* CBS 12357, approximately half of the population was not fluorescently labelled after staining and incubation (Supplementary Figure 2). As the dye may not be able to penetrate the spore cell wall, the observed loss of staining could be due to loss of bound fluorophores during germination, when the spore cell wall is lost. To allow for efficient germination of spores while minimizing cell division prior to staining, a 5 h incubation on YPD at 30 °C was implemented (Supplementary Figure 3). Using a protocol that included these optimizations, germinated spores of *S. eubayanus* CBS 12357 stained with CellTrace™ CFSE were mated with haploid cells of *S. cerevisiae* strain IMK439 stained with CellTrace™ Violet dye (Figure 2A). As the ability of *S. eubayanus* to grow on trehalose was unknown, we mated the cells in YPD as well as on YPT medium at 12 °C. The fraction of hybridized cells was monitored during mating by sorting double-stained cells onto YPD and determining the fraction of sorted cells which could grow on selective medium (Figure 2B). After 7 h, 1 % of the double-stained population of both mating cultures on YPT and YPD was hybrid (Figure 2C). In contrast to intra-species *S. cerevisiae* mating, mating on YPT yielded no increase in the fraction of hybrids upon prolonged incubation for interspecies mating. However, in YPD, the fraction of hybrids among the double-stained cells increased to 18 % after 24 h and remained stable up to 30 h (Figure C). In contrast, after 30 h incubation in YPD without sorting, only 0.3 % of the total population was able to grow on selective medium. These results indicated that FACS-based sorting of double-stained cells resulted in a 70-fold enrichment of interspecies hybrids by sorting.

### Generation of interspecies hybrids without selectable phenotypes

To test applicability of the dual fluorescent staining FACS protocol for generation of hybrids without selectable genetic markers, spores of *S. eubayanus* CBS 12357 (13) were crossed with the haploid *S. cerevisiae* strain CEN.PK113-7D (*MAT $\alpha$* ) (59). In parallel, we crossed spores of the Tibetan *S. eubayanus* isolate CDFM21L.1 (60) with spores of the ale-brewing *S. cerevisiae* isolate Ale28, which was provided by HEINEKEN Supply Chain. These diploid strains were sporulated and germinated as described previously (Figure 2A). *S. eubayanus* parents were stained with CellTrace™ CFSE, *S. cerevisiae* parents with CellTrace™ Violet, and cells were co-incubated in YPD during 30 h at 12 °C. Individual double-stained cells were sorted into 96 well plates containing 100  $\mu$ L YPD per well, and incubated at 30 °C until cultures were fully grown (Figure 3A). To eliminate false positives due to co-sorting of *S. eubayanus*/*S. cerevisiae* combinations, a single cell from each well was sorted into a second 96 well plate containing YPD. After incubation at 30 °C, the presence of genetic material from both parents was verified by PCR amplification of the *S. cerevisiae* specific *MEX67* gene and of the *S. eubayanus* specific *FSY1* gene (54, 55). For the CBS 12357  $\times$  CEN.PK113-7D cross, a band corresponding to *MEX67* and to *FSY1* was observed for 2 of 22 tested single-cell isolates. These isolates were stored as IMH001 and IMH002 (Figure 3C). For the CDFM21L.1  $\times$  Ale28 cross, a band corresponding to *MEX67* and to *FSY1* was produced for 5 of 34 tested single-cell isolates, which were stored as IMH003-IMH007 (Figure 3C). To verify if strains IMH001-IMH007 were hybrids and not mixtures of haploid *S. cerevisiae* and *S. eubayanus* cells, the ploidy of the sorted cells was determined

by DNA staining using SYTOX Green and flow cytometric analysis. The genome content of IMH001-IMH006 was diploid, whereas IMH007 was aneuploid (Figure 3B), indicating successful mating. Therefore, 9 % of tested cells from the mating between CBS 12357 and CEN.PK113-7D and 15 % of cells from the mating between CDFM21L.1 and Ale28 were hybrids. These results indicate that fluorescent staining and FACS enable a substantial enrichment of hybrid cells both for laboratory and industrial-relevant strains. A simple PCR protocol was sufficient to identify hybrids after enrichment.

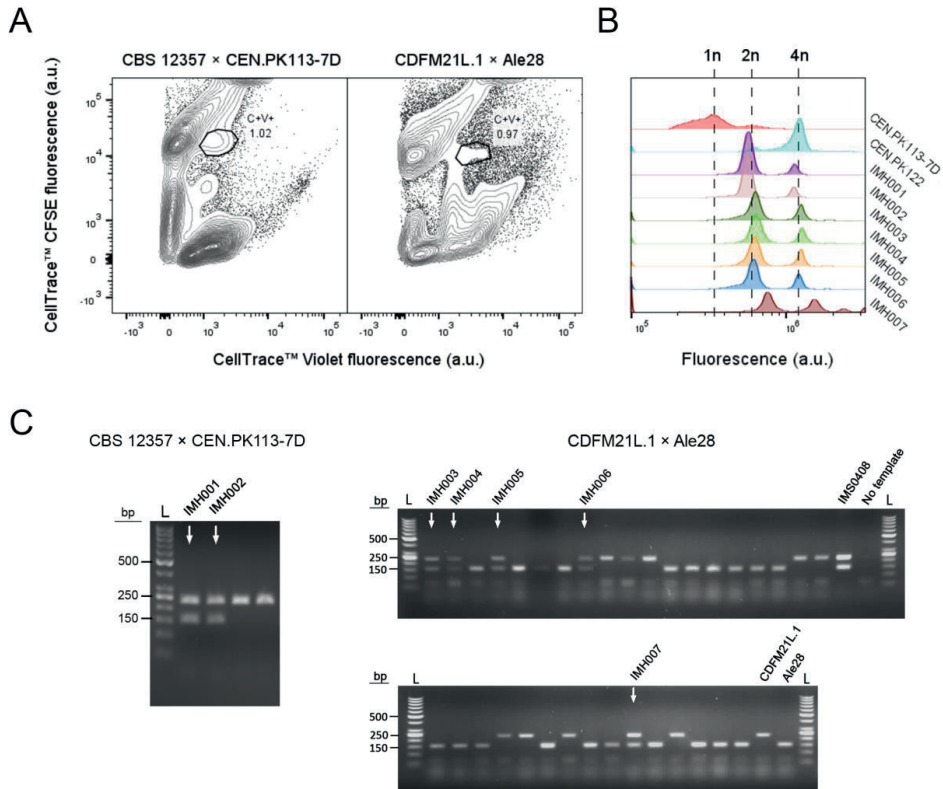


**Figure 2. Optimization of interspecies hybridization between haploid *S. eubayanus* and *S. cerevisiae* strains.** (A) Overview of the optimized protocol for interspecies spore-to-cell mating. (B) Fluorescence contour plots of mating cultures of stained CBS 12357 spores and IMK439 cells after 30 h of mating on YPD and YPT. The gated areas were used for sorting cells, event rates of each gate are indicated as a percentage. (C) Percentages of cells in the double-stained population able to grow on SM+G418 after 3.5, 7, 24 and 30 h of incubation on YPT (black) and YPD (white). Mating on YPT was only assessed at 7 and 30 h.

### Generation of interspecies hybrids by rare mating

Polyploidy and aneuploidy are commonly observed in industrial *Saccharomyces* hybrids (15, 61, 62), and chromosome copy number can play a key role in industrial performance (12, 22). The poor sporulation efficiency of many industrial strains can preclude hybridization by conventional mating (63). We explored use of the dual fluorescent staining and FACS to enrich hybrids obtained by rare mating by testing combinations of haploid and diploid *S. eubayanus* and *S. cerevisiae* strains. To evaluate low fractions of hybrid cells, we used *S. eubayanus* strains with uracil prototrophy and *S. cerevisiae* strains with uracil auxotrophy and resistance to the antibiotic G418. To obtain a diploid *S. cerevisiae* strain with uracil auxotrophy and resistance to the antibiotic G418, we first crossed IMK439 (*MATα ura3Δ::KanMX*) and IMK440 (*MATa ura3Δ::KanMX*) using fluorescent staining and FACS, resulting in IMX1471 (*MATa/MATα, ura3Δ::KanMX/ura3Δ::KanMX*). The mating types, ploidy, ability to sporulate, uracil prototrophy and G418 resistance of IMX1471 were verified (Supplementary Figure 4). Due to the anticipated low frequency of rare mating, *S. eubayanus* cells were stained with CellTrace™ CFSE and *S. cerevisiae* cells with CellTrace™ Far Red, as these dyes

have little spectral overlap (Supplementary Figure 4). In total, three different crosses were made: CBS 12357 (sporulated, 1n) × IMX1471 (2n), CBS 12357 (2n) × IMK439 (1n) and CBS 12357 (2n) × IMX1471 (2n). The frequency of hybrid cells in each mating culture was assessed by plating  $2 \times 10^8$  cells on SM+G418 plates and counting colonies. In parallel, the mating culture was analysed by FACS and double-stained cells were sorted and replica-plated to SM+G418 to determine the frequency of hybrid cells after sorting. Due to the low frequency of rare mating, wells were inoculated with 1, 10 or 100 double-stained cells and the fraction of growing cells was calculated using Poisson statistics.



**Figure 3. Generation of interspecies hybrids without selectable phenotypes from CBS 12357 (*S. eubayanus*, sporulated) and CEN.PK113-7D (*MATa*) and from CDFM21L.1 (*S. eubayanus*, sporulated) and Ale28 (*S. cerevisiae*, sporulated).** (A) Fluorescence contour plots of mating cultures between CBS 12357 (CFSE) × CEN.PK113-7D (Violet) and CDFM21L.1 (CFSE) × Ale28 (Violet). Gated areas were used for sorting cells, event rates of each gate are indicated as a percentage of the total population size. (B) Flow cytometric quantification of the genome content of constructed hybrids using SYTOX Green staining. *S. cerevisiae* strains CEN.PK113-7D and CEN.PK1122 were used as a haploid and diploid control, respectively. (C) Multiplex PCR amplification of the *S. cerevisiae* specific *MEX67* gene (150 bp) and the *S. eubayanus* specific *FSY1* gene (228 bp) in single-cell isolates of the double-stained populations from CBS 12357 × CEN.PK113-7D and CDFM21L.1 × Ale28 mating cultures. For CBS 12357 × CEN.PK113-7D, 4 of the 22 tested isolates are shown. Genomic DNA of *S. cerevisiae* Ale26, *S. eubayanus* CDFM21L.1 and *S. cerevisiae* × *S. eubayanus* IMS0408 were used as controls. Hybrid isolates are indicated by arrows. L: Generuler 50 bp DNA Ladder.



For the CBS 12357 (2n) × IMK439 (1n) cross, the fraction of hybrids in the total population varied between  $1.6 \times 10^{-6}$  and  $7.2 \times 10^{-6}$  between 24 and 168 h. After sorting, the fraction increased on average by a factor of 590 to between  $4.3 \times 10^{-4}$  and  $1.3 \times 10^{-3}$  (64). For the CBS 12357 (1n) × IMX1471 (2n) cross, the fraction of hybrids in the total population varied between  $3 \times 10^{-7}$  to  $1.5 \times 10^{-6}$  between 24 and 168h. Sorting only yielded a single hybrid after 96h, at a fraction on  $4.3 \times 10^{-4}$ , corresponding to a 540-fold enrichment. For the CBS 12357 (2n) × IMX1471 (2n) cross, a single hybrid was observed after 96h of incubation, corresponding to a rate of  $1 \times 10^{-7}$ , while no hybrids were identified after sorting. Overall, while rare mating was possible between the haploid and diploid strains, mated cells were present in very low frequencies both in the mating cultures and in the double-stained cells. In theory, fluorescent staining and FACS could be combined with high throughput PCR screening for hybrids in the sorted population. However, hundreds of cells would need to be screened for the diploid CBS 12357 x haploid IMK439 cross, and even more for the other crosses.

**Table 2: Fraction of hybrid cells after interspecies rare mating between *S. eubayanus* strain CBS 12357 and *S. cerevisiae* strains IMK439 (1n) and IMX1471 (2n) as determined by the ability to grow on SM+G418.** The fraction of hybrids in the total population was determined by plating approximately  $2 \times 10^8$  cells of the mating culture on SM+G418. For the fraction of hybrid cells in the double-stained population, 1 cell was sorted into 48 wells, 10 cells into 24 wells and 100 cells into 24 wells of a 96-well plate with YPD, which was replica-plated to SM+G418. Fractions of growing cells were calculated using with Poisson statistics. The sign “-” indicates that no hybrids were identified.

	CBS 12357 × IMK439		CBS 12357 (spores) × IMX1471		CBS 12357 × IMX1471	
	(2n × 1n)		(1n × 2n)		(2n × 2n)	
Mating time	Total population	After sorting	Total population	After sorting	Total population	After sorting
24 h	-	$4.3 \times 10^{-4}$	-	-	-	-
48 h	$4.6 \times 10^{-6}$	$2.9 \times 10^{-3}$	$8 \times 10^{-7}$	-	-	-
72 h	-	$4.1 \times 10^{-3}$	-	-	-	-
96 h	$7.2 \times 10^{-6}$	$4.3 \times 10^{-3}$	$8 \times 10^{-7}$	$4.3 \times 10^{-4}$	$1 \times 10^{-7}$	-
120 h	$1.9 \times 10^{-6}$	$1.3 \times 10^{-3}$	$9 \times 10^{-7}$	-	-	-
144 h	$1.6 \times 10^{-6}$	$4.3 \times 10^{-4}$	$3 \times 10^{-7}$	-	-	-
168 h	$4.7 \times 10^{-6}$	$3.6 \times 10^{-3}$	$1.5 \times 10^{-6}$	-	-	-

## Discussion

This study presents a new method for the generation of intra-species crosses and interspecies hybrids within the *Saccharomyces sensu stricto* complex that does not require parental strains and/or the resulting hybrids to have selectable phenotypes. By dual staining of parental cells with commercially-available fluorescent dyes prior to mating, mated cells could be enriched by up to 600-fold through sorting double-stained cells using FACS. In order to be able to mate homothallic strains, we developed a protocol for sporulation and germination prior to staining. Double-stained subpopulations selected by FACS after application of this protocol contained about 80 % mated cells for intra-species crosses and 10 – 15 % of mated cells for interspecies hybridization. By screening sorted double-stained cells using PCR, hybrids were successfully isolated from crosses of both laboratory strains and industrially relevant strains that did not have selectable phenotypes. By circumventing the need of conventional hybridization techniques for pre-existing or engineered selectable phenotypes (25, 38, 39, 43), this method enables generation of hybrids from a wide range of strains within just a few days. Interspecies hybrids have previously been obtained by fluorescent

staining and protoplast fusion. However, in contrast to protoplast fusion, hybridization by mating is not considered a GMO-technique, making it suitable for application in the food and beverage industry (64). The use of staining is not problematic for industrial application as it is rapidly lost by dilution during subsequent cell division of the hybrid cells (57).

The generation of interspecies hybrids using fluorescent labelling and FACS provides new opportunities for the use of laboratory-made hybrids for applications such as the production of fermented beverages and biofuels (20, 26, 30, 33, 37). Since hybrid physiology depends strongly on the combination of parental strains (19, 20), the possibility to mate strains without any selectable phenotype could widen the phenotypic diversity of laboratory-made hybrids. In the future, the phenotypic diversity of laboratory-hybrids could be extended by applying fluorescent labelling to mass mating approaches (32). Due to the inability of many industrial strains to sporulate (32), and due to the potential value of higher-ploidy hybrids (12, 22), the generation of interspecies hybrids by rare mating would also be valuable for industrial strain development. While fluorescent labelling and FACS did enable a 600-fold enrichment of interspecies hybrids obtained by rare mating, the isolation of hybrids would require extensive screening. However, PCR based screening of hundreds of candidates is not impossible in industrial strain improvement programmes. Moreover, the development of high throughput methods such as microfluidic lab-on-a-chip setups could further simplify screening after FACS sorting of rare hybrids (65, 66).

On a fundamental level, the ability to generate diverse interspecies hybrids using fluorescent labelling could simplify research on hybrid-specific phenomena such as heterosis (21, 67), the inheritance of mitochondrial DNA in hybrids (68), genome stability and loss of heterozygosity (69) or hybrid sterility (8, 70). Overall, the interspecies mating procedure presented here may be strongly accelerate industrial strain improvement programs and fundamental research into hybrid yeasts.

## Acknowledgments

The authors would like to thank Prof. Fen-Yang Bai (Chinese Academy of Sciences) for kindly providing us with CDFM21L.1 and Xavier Hakkaart (Delft University of Technology) for his expertise and support in fluorescence microscopy.

## References

1. Marsit S & Dequin S (2015) Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. *FEMS Yeast Res* 15(7).
2. Nielsen J (2013) Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered* 4(4):207-211.
3. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2017) Novel brewing yeast hybrids: creation and application. *Appl Microbiol Biotechnol* 101(1):65-78.
4. Balat M (2011) Production of bioethanol from lignocellulosic materials via the biochemical pathway: a review. *Energ Convers Manage* 52(2):858-875.
5. Jansen MLA, et al. (2017) *Saccharomyces cerevisiae* strains for second-generation ethanol production: from academic exploration to industrial implementation. *FEMS Yeast Res* 17(5).
6. Hittinger CT (2013) *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet* 29(5):309-317.
7. Nueno-Palop C, Bond CJ, McGhie H, Roberts IN, & Delneri D (2017) *Saccharomyces jurei* sp. nov., Isolation and genetic identification of a novel yeast species from *Quercus robur*. *Int J Syst Evol Microbiol* 67(6):2046-2052.
8. Greig D, Borts RH, & Louis EJ (2002) Epistasis and hybrid sterility in *Saccharomyces*. *Proc Biol Sci* 269(1496):1167-1171.
9. Hou J, Friedrich A, de Montigny J, & Schacherer J (2014) Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol* 24(10):1153-1159.
10. Stefanini I, et al. (2016) Social wasps are a *Saccharomyces* mating nest. *Proc Natl Acad Sci USA* 113(8):2247-2251.
11. Boynton PJ & Greig D (2014) The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast* 31(12):449-462.

12. Gorter de Vries AR, Pronk JT, & Daran J-MG (2017) Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts. *Appl Environ Microbiol*:AEM. 03206-03216.
13. Libkind D, et al. (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA* 108(35):14539-14544.
14. Naumov G, et al. (2001) Genetic identification of *Saccharomyces bayanus* var. *uvarum*, a cider-fermenting yeast. *Int J Food Microbiol* 65(3):163-171.
15. González SS, Barrio E, Gafner J, & Querol A (2006) Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res* 6(8):1221-1234.
16. Peter J, et al. (2018) Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556(7701):339.
17. Naumova ES, Naumov GI, Michailova YV, Martynenko NN, & Masneuf-Pomarède I (2011) Genetic diversity study of the yeast *Saccharomyces bayanus* var. *uvarum* reveals introgressed subtelomeric *Saccharomyces cerevisiae* genes. *Res Microbiol* 162(2):204-213.
18. Dunn B, Richter C, Kvitek DJ, Pugh T, & Sherlock G (2012) Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res* 22(5):908-924.
19. Krogerus K, Seppänen-Laakso T, Castillo S, & Gibson B (2017) Inheritance of brewing-relevant phenotypes in constructed *Saccharomyces cerevisiae* × *Saccharomyces eubayanus* hybrids. *Microb Cell Fact* 16(1):66.
20. Mertens S, et al. (2015) A large set of newly created interspecific *Saccharomyces* hybrids increases aromatic diversity in lager beers. *Appl Environ Microbiol* 81(23):8202-8214.
21. Shapira R, Levy T, Shaked S, Fridman E, & David L (2014) Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models. *Heredity (Edinb)* 113(4):316.
22. Krogerus K, et al. (2016) Ploidy influences the functional attributes of *de novo* lager yeast hybrids. *Appl Microbiol Biotechnol* 100(16):7203-7222.
23. Coloretti F, Zambonelli C, & Tini V (2006) Characterization of flocculent *Saccharomyces* interspecific hybrids for the production of sparkling wines. *Food Microbiol* 23(7):672-676.
24. Hebly M, et al. (2015) *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res* 15(3).
25. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2015) New lager yeast strains generated by interspecific hybridization. *J Ind Microbiol Biotechnol* 42(5):769-778.
26. Bellon JR, et al. (2011) Newly generated interspecific wine yeast hybrids introduce flavour and aroma diversity to wines. *Appl Microbiol Biotechnol* 91(3):603-612.
27. Steensels J, Meersman E, Snoek T, Saels V, & Verstrepen KJ (2014) Large-scale selection and breeding to generate industrial yeasts with superior aroma production. *Appl Environ Microbiol* 80(22):6965-6975.
28. Bellon JR, Schmid F, Capone DL, Dunn BL, & Chambers PJ (2013) Introducing a new breed of wine yeast: interspecific hybridisation between a commercial *Saccharomyces cerevisiae* wine yeast and *Saccharomyces mikatae*. *PLoS One* 8(4):e62053.
29. Nikulin J, Krogerus K, & Gibson B (2017) Alternative *Saccharomyces* interspecies hybrid combinations and their potential for low-temperature wort fermentation. *Yeast*.
30. Bellon JR, Yang F, Day MP, Inglis DL, & Chambers PJ (2015) Designing and creating *Saccharomyces* interspecific hybrids for improved, industry relevant, phenotypes. *Appl Microbiol Biotechnol* 99(20):8597-8609.
31. Lopandic K, et al. (2016) Genotypic and phenotypic evolution of yeast interspecies hybrids during high-sugar fermentation. *Appl Microbiol Biotechnol* 100(14):6331-6343.
32. Steensels J, et al. (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol Rev* 38(5):947-995.
33. Peris D, et al. (2017) Hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production. *Biotechnol Biofuels* 10(1):78.
34. Nikulin J, Krogerus K, & Gibson B (2018) Alternative *Saccharomyces* interspecies hybrid combinations and their potential for low-temperature wort fermentation. *Yeast* 35(1):113-127.
35. Gunge N & Nakatomi Y (1972) Genetic mechanisms of rare matings of the yeast *Saccharomyces cerevisiae* heterozygous for mating type. *Genetics* 70(1):41-58.
36. Alexander WG, et al. (2016) Efficient engineering of marker-free synthetic allotetraploids of *Saccharomyces*. *Fungal Genet Biol* 89:10-17.
37. Magalhães F, Krogerus K, Vidgren V, Sandell M, & Gibson B (2017) Improved cider fermentation performance and quality with newly generated *Saccharomyces cerevisiae* × *Saccharomyces eubayanus* hybrids. *J Ind Microbiol Biotechnol* 44(8):1203-1213.
38. Fernández-González M, Úbeda JF, & Briones AI (2015) Study of *Saccharomyces cerevisiae* Wine Strains for Breeding Through Fermentation Efficiency and Tetrad Analysis. *Curr Microbiol* 70(3):441-449.
39. Pérez-Través L, Lopes CA, Barrio E, & Querol A (2012) Evaluation of different genetic procedures for the generation of artificial hybrids in *Saccharomyces* genus for winemaking. *Int J Food Microbiol* 156(2):102-111.
40. Scannell DR, et al. (2011) The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 (Bethesda)* 1(1):11-25.
41. Bell PJ, et al. (1998) A flow cytometric method for rapid selection of novel industrial yeast hybrids. *Appl Environ Microbiol* 64(5):1669-1672.

42. Piotrowski JS, *et al.* (2012) Different selective pressures lead to different genomic outcomes as newly-formed hybrid yeasts evolve. *BMC Evol Biol* 12(1):46.
43. da Silva T, *et al.* (2015) Hybridization within *Saccharomyces* genus results in homeostasis and phenotypic novelty in winemaking conditions. *PLoS One* 10(5):e0123834.
44. Wunderlich S & Gatto KA (2015) Consumer perception of genetically modified organisms and sources of information. *Adv Nutr* 6(6):842-851.
45. Gorter de Vries AR, de Groot PA, Broek M, & Daran J-MG (2017) CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb Cell Fact* 16(1):222.
46. Cormack BP, Valdivia RH, & Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173(1):33-38.
47. Verduyn C, Postma E, Scheffers WA, & Van Dijken JP (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8(7):501-517.
48. Cheng T-H, Chang C-R, Joy P, Yablok S, & Gartenberg MR (2000) Controlling gene expression in yeast by inducible site-specific recombination. *Nucleic Acids Res* 28(24):e108-e108.
49. Bahalul M, Kaneti G, & Kashi Y (2010) Ether-zymolyase ascospore isolation procedure: an efficient protocol for ascospores isolation in *Saccharomyces cerevisiae* yeast. *Yeast* 27(12):999-1003.
50. Herman PK & Rine J (1997) Yeast spore germination: a requirement for Ras protein activity during re-entry into the cell cycle. *EMBO J* 16(20):6171-6181.
51. Dube S, Qin J, & Ramakrishnan R (2008) Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. *PLoS One* 3(8):e2876.
52. Schindelin J, *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9(7):676-682.
53. Haase SB & Reed SI (2002) Improved flow cytometric analysis of the budding yeast cell cycle. *Cell cycle* 1(2):117-121.
54. Muir A, Harrison E, & Wheals A (2011) A multiplex set of species-specific primers for rapid identification of members of the genus *Saccharomyces*. *FEMS Yeast Res* 11(7):552-563.
55. Pengelly RJ & Wheals AE (2013) Rapid identification of *Saccharomyces eubayanus* and its hybrids. *FEMS Yeast Res* 13(2):156-161.
56. Huxley C, Green ED, & Dunham I (1990) Rapid assessment of *S. cerevisiae* mating type by PCR. *Trends Genet* 6:236.
57. Filby A, Begum J, Jalal M, & Day W (2015) Appraising the suitability of succinimidyl and lipophilic fluorescent dyes to track proliferation in non-quiescent cells by dye dilution. *Methods* 82:29-37.
58. Herskowitz I (1988) Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol Rev* 52(4):536.
59. Entian K-D & Kötter P (2007) 25 Yeast genetic strain and plasmid collections. *Methods Microbiol* 36:629-666.
60. Bing J, Han P-J, Liu W-Q, Wang Q-M, & Bai F-Y (2014) Evidence for a Far East Asian origin of lager beer yeast. *Curr Biol* 24(10):R380-R381.
61. Peris D, *et al.* (2012) The molecular characterization of new types of *Saccharomyces cerevisiae* × *S. kudriavzevii* hybrid yeasts unveils a high genetic diversity. *Yeast* 29(2):81-91.
62. Querol A & Bond U (2009) The complex and dynamic genomes of industrial yeasts. *FEMS Microbiol Lett* 293(1):1-10.
63. Anderson E & Martin P (1975) The sporulation and mating of brewing yeasts. *J. Inst. Brew.* 81(3):242-247.
64. Gibson B, *et al.* (2017) New yeasts—new brews: modern approaches to brewing yeast design and development. *FEMS yeast research* 17(4).
65. Schmitz CH, Rowat AC, Köster S, & Weitz DA (2009) Dropspots: a picoliter array in a microfluidic device. *Lab Chip* 9(1):44-49.
66. Gach PC, Iwai K, Kim P, Hillson N, & Singh AK (2017) Droplet Microfluidics for Synthetic Biology. *Lab Chip* 17(17):3388-3400.
67. Bernardes JP, Stelkens RB, & Greig D (2017) Heterosis in hybrids within and between yeast species. *J Evol Biol* 30(3):538-548.
68. Hsu Y-Y & Chou J-Y (2017) Environmental factors can influence mitochondrial inheritance in the *Saccharomyces* yeast hybrids. *PLoS One* 12(1):e0169953.
69. Smukowski Heil CS, *et al.* (2017) Loss of heterozygosity drives adaptation in hybrid yeast. *Mol Biol Evol* 34(7):1596-1612.
70. Lee H-Y, *et al.* (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 135(6):1065-1073.
71. González-Ramos D, van den Broek M, van Maris AJA, Pronk JT, & Daran J-MG (2013) Genome-scale analyses of butanol tolerance in *Saccharomyces cerevisiae* reveal an essential role of protein degradation. *Biotechnol Biofuels* 6(1):48.
72. Salazar AN, *et al.* (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res* 17(7):fox074.



## Chapter 8: Laboratory evolution of a *Saccharomyces cerevisiae* x *S. eubayanus* hybrid under simulated lager-brewing conditions

Arthur R. Gorter de Vries, Maaïke A. Voskamp, Aafke C. A. van Aalst, Line H. Kristensen, Liset Jansen, Marcel van den Broek, Alex N. Salazar, Nick Brouwers, Thomas Abeel, Jack T. Pronk and Jean-Marc G. Daran

*Saccharomyces pastorianus* lager-brewing yeasts are domesticated hybrids of *S. cerevisiae* x *S. eubayanus* that display extensive inter-strain chromosome copy number variation and chromosomal recombinations. It is unclear to what extent such genome rearrangements are intrinsic to the domestication of hybrid brewing yeasts and whether they contribute to their industrial performance. Here, an allodiploid laboratory hybrid of *S. cerevisiae* and *S. eubayanus* was evolved for up to 418 generations on wort under simulated lager-brewing conditions in six independent sequential batch bioreactors. Characterization of 55 single-cell isolates from the evolved cultures showed large phenotypic diversity and whole-genome sequencing revealed a large array of mutations. Frequent loss of heterozygosity involved diverse, strain-specific chromosomal translocations, which differed from those observed in domesticated, aneuploid *S. pastorianus* brewing strains. In contrast to the extensive aneuploidy of domesticated *S. pastorianus* strains, the evolved isolates only showed limited (segmental) aneuploidy. Specific mutations could be linked to calcium-dependent flocculation, loss of maltotriose utilization and loss of mitochondrial activity, three industrially relevant traits that also occur in domesticated *S. pastorianus* strains. This study indicates that fast acquisition of extensive aneuploidy is not required for genetic adaptation of *S. cerevisiae* x *S. eubayanus* hybrids to brewing environments. In addition, this work demonstrates that, consistent with the diversity of brewing strains for maltotriose utilization, domestication under brewing conditions can result in loss of this industrially relevant trait. These observations have important implications for the design of strategies to improve industrial performance of novel laboratory-made hybrids.

Essentially as published in *Frontiers in Genetics* 2019;10:242

Supplementary materials available online

<https://doi.org/10.3389/fgene.2019.00242>

## Introduction

*Saccharomyces* yeasts are popular eukaryotic models for studying genome hybridization, chromosome (mis)segregation and aneuploidy (1, 2). The genus *Saccharomyces* arose between 10 and 20 million years ago and currently comprises eight described species, as well as interspecies hybrids (3-5). Absence of a prezygotic barrier between *Saccharomyces* species facilitates hybridization, although spore viabilities of the resulting hybrids is typically well below 10 % (3-5). Several interspecies *Saccharomyces* hybrids are tightly associated with domestication in industrial processes. *S. pastorianus* lager-brewing yeasts are domesticated *S. cerevisiae* x *S. eubayanus* hybrids (6). Double and triple hybrids between *S. cerevisiae*, *S. kudriavzevii* and *S. uvarum* are closely associated with wine fermentation (7-9). *S. bayanus* cider fermentation yeasts are domesticated *S. uvarum* X *S. eubayanus* hybrids (10). Reconstruction of the corresponding *Saccharomyces* hybrids in the laboratory showed improved performance, relative to the parental species. For example, laboratory-made *S. cerevisiae* x *S. eubayanus* hybrids combined sugar utilization characteristics of *S. cerevisiae* and the superior performance at low temperatures of *S. eubayanus* (11, 12). Similarly, hybrids of *S. cerevisiae*, *S. kudriavzevii* and *S. uvarum* combined traits of their parental species relevant to industrial wine fermentation, such as flocculence, sugar utilization kinetics, stress tolerance and aroma production (13, 14).

The relevance of laboratory hybridization of *Saccharomyces* species extends beyond reconstruction of existing, domesticated hybrids. The ability of hybridization to generate extensive phenotypic diversity has raised interest in the development of novel *Saccharomyces* hybrids for specific industrial processes (12). For example, an *S. cerevisiae* x *S. paradoxus* hybrid produced high concentrations of aromatic compounds that are of interest for wine making (15). Hybrids between *S. cerevisiae* and *S. arboricola* or *S. mikatae* were able to utilize the sugars in wort at low temperatures and produced particularly aromatic beer (16). Laboratory hybrids of *S. cerevisiae* and *S. kudriavzevii* or *S. mikatae* yielded xylose-consuming strains with high inhibitor tolerance for 2<sup>nd</sup> generation biofuel production (17).

The alloeploid genomes of laboratory hybrids of *Saccharomyces* species strongly differ from the extremely aneuploidy genomes of the domesticated strains used in traditional industrial processes. For example, the genomes of *S. pastorianus* lager-brewing yeasts contain between 45 and 79 chromosomes (18, 19), a degree of aneuploidy that is not observed elsewhere in the *Saccharomyces* genus (20). However, it remains unclear when and how domestication resulted in the extensive chromosome copy number variations and phenotypic diversity of current *S. pastorianus* strains.

Hybrid genomes have a well-documented increased tendency to become aneuploid due to an increased rate of chromosome missegregation during mitosis and/or meiosis (4, 21). Aneuploidy reduces the efficiency of sporulation and can thereby complicate genetic modification, impeding breeding and targeted strain improvement (22, 23). In evolutionary contexts, aneuploidy is generally seen as a transient adaptation mechanism, whose positive impacts are eventually taken over by more parsimonious mutations (24). When grown mitotically, sporulated hybrid strains were prone to further chromosome missegregation resulting in more extensive chromosome copy number variations (14). Even genomes of *Saccharomyces* hybrids that had not undergone meiosis displayed increased rates of mitotic chromosome missegregation during mitosis (25). Indeed, when evolved in lignocellulosic hydrolysates, cultures of *S. cerevisiae* X *S. kudriavzevii* and *S. cerevisiae* x *S. mikatae* hybrids exhibited segmental and full-chromosome aneuploidy after only 50 generations (17).

Similarly, when evolved under wine fermentation conditions, *S. cerevisiae* x *S. kudriavzevii* hybrids displayed extensive genome reorganizations that led to a significant reduction of their genome content (26).

Genetic instability of hybrid genomes could be detrimental to stable, robust industrial performance. Therefore, to assess industrial applicability of new hybrids generated in the laboratory, it is important to determine their genome stability under industrially relevant conditions. Moreover, laboratory evolution under simulated industrial conditions can increase understanding of the selective pressures that shaped the genomes of domesticated microorganisms (27-29).

The goal of the present study was to investigate how the previously constructed allodiploid *S. cerevisiae* x *S. eubayanus* hybrid IMS0408 (11) evolves under simulated lager-brewing conditions, with a specific focus on genome dynamics and on acquisition or loss of brewing-related phenotypes. To mimic successive lager beer fermentation processes, the hybrid strain was subjected to sequential batch cultivation on industrial wort, in six independent bioreactor setups. After up to 418 generations, the genotypic and phenotype diversity generated in these laboratory evolution experiments was analyzed by characterization of 55 single-cell isolates. After whole-genome resequencing of each isolate using 150 bp paired-end reads, sequence data were mapped to high-quality reference genomes of the parental strains to identify genomic changes. Phenotypic analysis of the isolates focused on the ability to utilize maltotriose, flocculation and the respiratory capacity. We interpreted these results in the context of the domestication history of *S. pastorianus* brewing strains as well as in relation to genome stability and industrial application of newly generated *Saccharomyces* hybrids.

## Materials and Methods

### Yeast strains and media

*Saccharomyces* strains used in this study are listed in Supplementary Table 1. Yeast strains and *E. coli* strains containing plasmids were stocked in 1 mL aliquots after addition of 30 % v/v glycerol to the cultures and stored at -80 °C. For preparation of stock cultures and inocula of bioreactors, yeast strains were routinely propagated in shake flasks containing 100 mL YPD (10 g.L<sup>-1</sup> yeast extract, 20 g.L<sup>-1</sup> yeast peptone and 20 g.L<sup>-1</sup> glucose) at 30 °C and 200 RPM in an Brunswick Innova43/43R shaker (Eppendorf Nederland B.V., Nijmegen, The Netherlands). For cultivation on solid media, YPD medium was supplemented with 20 g.L<sup>-1</sup> Bacto agar (Becton Dickinson, Breda, The Netherlands) and incubation was done at 30 °C. Synthetic medium (SM), containing 3 g.L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g.L<sup>-1</sup> MgSO<sub>4</sub>.7H<sub>2</sub>O, 5 g.L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mL.L<sup>-1</sup> of a trace element solution and 1 mL.L<sup>-1</sup> of a vitamin solution, was prepared as previously described (30). SM maltotriose was supplemented with 20 g.L<sup>-1</sup> of maltotriose and SM ethanol with 20 mL.L<sup>-1</sup> of ethanol. Selection for the *amdS* marker was performed on SM-AC: SM with 0.6 g.L<sup>-1</sup> acetamide and 6.6 g.L<sup>-1</sup> K<sub>2</sub>SO<sub>4</sub> instead of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> as nitrogen source (31). For counter selection of the *amdS* marker, strains were first grown on YPD and then on SM-FAC: SM supplemented with 2.3 g.L<sup>-1</sup> fluoroacetamide (31). Industrial wort was provided by HEINEKEN Supply Chain B.V., Zoeterwoude, the Netherlands, and contained 14.4 g.L<sup>-1</sup> glucose, 2.3 g.L<sup>-1</sup> fructose, 85.9 g.L<sup>-1</sup> maltose, 26.8 g.L<sup>-1</sup> maltotriose and 269 mg.L<sup>-1</sup> free amino nitrogen. The wort was supplemented with 1.5 mg.L<sup>-1</sup> Zn<sup>2+</sup> by addition of ZnSO<sub>4</sub>.7H<sub>2</sub>O, then autoclaved for 30 min at 121 °C and, prior to use, filtered through Nalgene 0.2 µm SFCA bottle-top filters (ThermoFisher Scientific, Waltham, MA). For experiments performed with diluted wort, two volumes of sterile demineralized water were added per volume of wort. To prevent excessive foaming during the aeration phase of



the bioreactor experiments, (un)diluted wort was supplemented with 0.2 mL.L<sup>-1</sup> of sterile Pluronic PE 6100 antifoam (Sigma-Aldrich, Zwijndrecht, the Netherlands).

#### **Analytical methods and statistics**

Optical density at 660 nm was measured with a Libra S11 spectrophotometer (Biochrom, Cambridge, UK). HPLC analysis of sugar and metabolite concentrations was performed with an Agilent Infinity 1260 chromatography system (Agilent Technologies, Santa Clara, CA) with an Aminex HPX-87 column (Bio-Rad, Lunteren, The Netherlands) at 65 °C, eluted with 5 mM H<sub>2</sub>SO<sub>4</sub>. Significance of data was assessed by an unpaired two-tailed Student's t-test with a 95 % confidence interval.

#### **Laboratory evolution and single colony isolation**

The hybrid yeast strain IMS0408 was evolved under three different conditions in duplicate in Minifors 2 bioreactors (INFORS HT, Velp, the Netherlands) with a working volume of 100 mL: on diluted wort at 30 °C (LG30.1 and LG30.2), on diluted wort at 12 °C (LG12.1 and LG12.2) and on full-strength wort at 12 °C (HG12.1 and HG12.2). Sequential batch cultivation was performed with 10 and 30 mL.min<sup>-1</sup> of headspace N<sub>2</sub> flushing at 12 and 30 °C, respectively. The percentage of CO<sub>2</sub> in the outlet gas stream, the culture pH and the dissolved oxygen concentration in the broth were continuously monitored. The end of a batch cultivation cycle was automatically triggered when the percentage of CO<sub>2</sub> in the offgas decreased below 75 % and 10 % of the maximum value reached during that cycle for growth on diluted wort and full-strength wort, respectively. These CO<sub>2</sub> percentages correspond to the moment at which sugar utilization was complete in the first batch cycle for each condition, as determined by HPLC measurements. When the CO<sub>2</sub> threshold was reached, the reactor was emptied while stirring at 1200 RPM leaving about 7 mL to inoculate the next batch. Upon addition of fresh medium, the broth was stirred at 500 RPM and sparged with 500 mL.min<sup>-1</sup> of pressurized air during 5 min for diluted wort or 12 h for wort. During the remainder of each batch cultivation cycle, the medium was not sparged or stirred and the pH was not adjusted. LG30.1 and LG30.2 were carried out for 116 and 117 cycles respectively, LG12.1 and LG12.2 were carried out for 29 cycles and HG12.1 and HG12.2 were carried out for 13 and 16 cycles, respectively. Culture samples from all six reactors were then streaked on YPD plates and after three subsequent restreaks, frozen stock cultures of single colony isolates were prepared. By default, five isolates were obtained for each culture. For LG12.1 and LG30.1, two different colony morphologies were observed, therefore five elevated and conically-shaped colonies and five regular flat colonies were stocked. The experiments at 12 °C on diluted wort were continued for four months until a total of 58 and 57 cycles was reached for LG12.1 and LG12.2, respectively, and five single-cell isolates were obtained for each reactor as described above.

#### **Genomic DNA extraction and whole genome sequencing**

Yeast cultures were incubated in 500-mL shake-flasks containing 100 mL YPD at 30°C on an orbital shaker set at 200 RPM until the strains reached stationary phase at an OD<sub>660</sub> between 12 and 20. Genomic DNA was isolated using the Qiagen 100/G kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and quantified using a Qubit® Fluorometer 2.0 (ThermoFisher Scientific). For IMS0408 and the evolved isolates, genomic DNA was sequenced at Novogene Bioinformatics Technology Co., Ltd (Yuen Long, Hong Kong) on a HiSeq2500 sequencer (Illumina, San Diego, CA) with 150 bp paired-end reads using PCR-free library preparation.

### Genome analysis

A high quality reference genome was constructed by combining near-complete assemblies of *S. cerevisiae* CEN.PK113-7D (32) and *S. eubayanus* CBS 12357<sup>T</sup> (33). The kanMX marker present in IMS0408 was inserted as an additional contig (34). For each evolved strain, raw Illumina reads were aligned against the reference genome using the Burrows–Wheeler Alignment tool (BWA, version 0.7.15-r1142) and further processed using SAMtools (version 1.3.1) and Pilon (version 1.18) for variant calling (35-37). SNPs and INDELS that were also called or which were ambiguous in IMS0408, were disregarded. Copy number was determined based on read coverage analysis. Chromosomal translocations were detected using Breakdancer (version 1.3.6) (38). Only translocations which were supported by at least 10 % of the reads aligned at that locus and which were absent in strain IMS0408 were considered. All SNPs, INDELS, recombinations and copy number changes were manually confirmed by visualising the generated .bam files in the Integrative Genomics Viewer (IGV) software (39). A complete list of identified mutations is provided in Supplementary Data File 1. For chimeric open-reading-frame reconstruction, reads aligning within 3 kbp of an identified recombination site and their paired reads were extracted using Python and were assembled using SPAdes (40). The resulting contigs were aligned against ORFs of genes the genes affected by the recombination to identify the recombination point, and the complete recombined ORF was reconstructed. Original and recombined ORFs were then aligned and translated using CloneManager (version 9.51, Sci-Ed Software, Denver, CO) to determine whether the translocation had introduced frameshifts or premature stop codons.

### DNA content determination by flow cytometric analysis.

Exponential-phase shake flask cultures on YPD were diluted to an OD<sub>660</sub> of 1. A 1 mL sample (approximately 10<sup>7</sup> cells) was then washed in cold demineralized water and resuspended 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. DNA was then stained with SYTOX Green as described previously (41). Samples were analyzed on a Accuri C6 flow cytometer (BD Biosciences, Franklin Lakes, NJ) equipped with a 488-nm laser and the median fluorescence of cells in the 1n and 2n phases of the cell cycle was determined using FlowJo (BD Biosciences). The 1n and 2n medians of strains CEN.PK113-7D (n), CEN.PK122 (2n) and FRY153 (3n) were used to create a standard curve of fluorescence versus genome size with a linear curve fit, as performed previously (18). The genome size of each tested strain was estimated by averaging predicted genome sizes of the 1n and 2n population in assays on three independent cultures.

### Identification of strains with respiratory deficiency

Respiratory competence was assessed through their ability to grow on ethanol. Samples from 24 h shake-flask cultures on YPD (30 °C, 200 RPM ) were washed twice with demineralized water and used to inoculate duplicate aerobic shake flasks containing 100 mL of SM with 2 % ethanol to an OD<sub>660</sub> of 0.2. After 72 h incubation at 30 °C and 200 RPM, OD<sub>660</sub> was measured.

### Assay for calcium-dependence of flocculation

Two 100 µL aliquots from overnight cultures on YPD were washed with sterile demineralized water. One aliquot was resuspended in demineralized water and the other in 50 mM EDTA (pH 7.0). Both samples were imaged at 100 x magnification under a Z1 microscope (Carl Zeiss BV, Breda, the Netherlands) to assess flocculence and its reversal by EDTA chelation of calcium ions.

### Plasmid construction

All plasmids were propagated in *E. coli* DH5 $\alpha$  (Table 1). The gRNAs to target *ScSFL1* and *SeSFL1* (Supplementary Table 2) were designed as previously described (23) and ordered as *de novo* synthesized plasmids pUD711 (*ScSFL1*) and pUD712 (*SeSFL1*) at GeneArt (ThermoFisher Scientific). Plasmid pUDP104, expressing gRNA<sub>*ScSFL1*</sub> and *cas9*, was constructed by Golden Gate cloning by digesting pUDP004 and pUD711 using BsaI and ligating with T4 ligase (42). Similarly, plasmid pUDP105, expressing gRNA<sub>*SeSFL1*</sub> and *cas9*, was constructed from pUDP004 and pUD712. Correct assembly was verified by restriction analysis using PdmI.

**Table 1: Plasmids used throughout this study.**

Name	Relevant genotype	Origin
pUD711	ori <i>bla</i> gRNA- <i>ScSFL1</i>	GeneArt™
pUD712	ori <i>bla</i> gRNA- <i>SeSFL1</i>	GeneArt™
pUDE481	ori <i>bla</i> ARS4/CEN6 hyg <sup>R</sup> <i>ScTDH3p-mTurquoise2-ScADH1t</i>	(43)
pUDE482	ori <i>bla</i> ARS4/CEN6 hyg <sup>R</sup> <i>ScTEF1p-Venus-ScENO2t</i>	(43)
pUDP004	ori <i>bla</i> panARSopt amdSYM <i>Spcas9</i>	(23)
pUDP045	ori <i>bla</i> panARSopt amdSYM <i>Spcas9</i> gRNA- <i>ScMAL11</i>	(43)
pUDP104	ori <i>bla</i> panARSopt amdSYM <i>Spcas9</i> gRNA- <i>ScSFL1</i>	This study
pUDP105	ori <i>bla</i> panARSopt amdSYM <i>Spcas9</i> gRNA- <i>SeSFL1</i>	This study

### Strain construction

The *ScTEF1p-Venus-ScENO2t* repair fragment with flanks for homologous recombination in the *ScMAL11* locus was PCR amplified from plasmid pUDE481 using primers 12989 and 12990 (Supplementary Table 2). The *ScTDH3p-mTurquoise2-ScADH1t* repair fragment with flanks for homologous recombination in the *ScSFL1* locus was PCR amplified from plasmid pUDE482 using primers 13564 and 13565. The *ScTEF1p-Venus-ScENO2t* repair fragment with flanks for homologous recombination in the *SeSFL1* locus was PCR amplified from plasmid pUDE481 using primers 13566 and 13567.

All strains were transformed by electroporation as described previously, with 300 ng of gRNA/Cas9 expression plasmid and 1  $\mu$ g of repair fragment (23). Strains IMX1698 (mVenus:: $\Delta$ *ScMAL11*), IMX1824 (mTurquoise2:: $\Delta$ *ScSFL1*) and IMX1825 (Venus:: $\Delta$ *SeSFL1*) were constructed by transforming IMS0408 with the appropriate repair fragments and plasmids pUDP045, pUDP104 and pUDP105, respectively. Strain IMX1826 (mTurquoise2:: $\Delta$ *ScSFL1* Venus:: $\Delta$ *SeSFL1*) was constructed by transforming IMX1824 using the appropriate repair fragment and plasmid pUDP105. After electroporation, cells were transferred to 20 mL SM-Ac medium to select successful transformants and incubated at 30 °C for 3 to 5 days. After growth was observed, 200  $\mu$ L of culture was transferred to 20 mL fresh SM-Ac and incubated similarly during 24h. Finally, 200  $\mu$ L from the second culture was transferred to 20 mL fresh YPD medium to maximize expression of fluorescent proteins. Successfully gene-edited cells were sorted using the BD FACSAria™ II SORP Cell Sorter (BD Biosciences) as described previously (43). The plasmids were cured from strains IMX1698, IMX1824, IMX1825 and IMX1826 by subsequent growth on YPD and plating on SM-FAC. After confirmation of the correct genotype by colony PCR, randomly picked colonies were used to prepare frozen stocks.

### Biomass sedimentation assay

IMS0408, IMS0556, IMS0558, IMS0559, IMS0617, IMX1824, IMX1825 and IMX1826 were grown in triplicate during 72h in vented 50 mL Bio-One Cellstar Cellreactor tubes (Sigma-Aldrich) on 20 mL YPD at 30 °C and 200 RPM until stationary phase. For each sample, the biomass was resuspended by

vigorous vortexing and 1 mL was sampled immediately after vortexing from right underneath the meniscus. After 60 s of stationary incubation, another sample was taken by the same procedure. Biomass sedimentation was quantified as the ratio of the OD<sub>660</sub> values of the two samples.

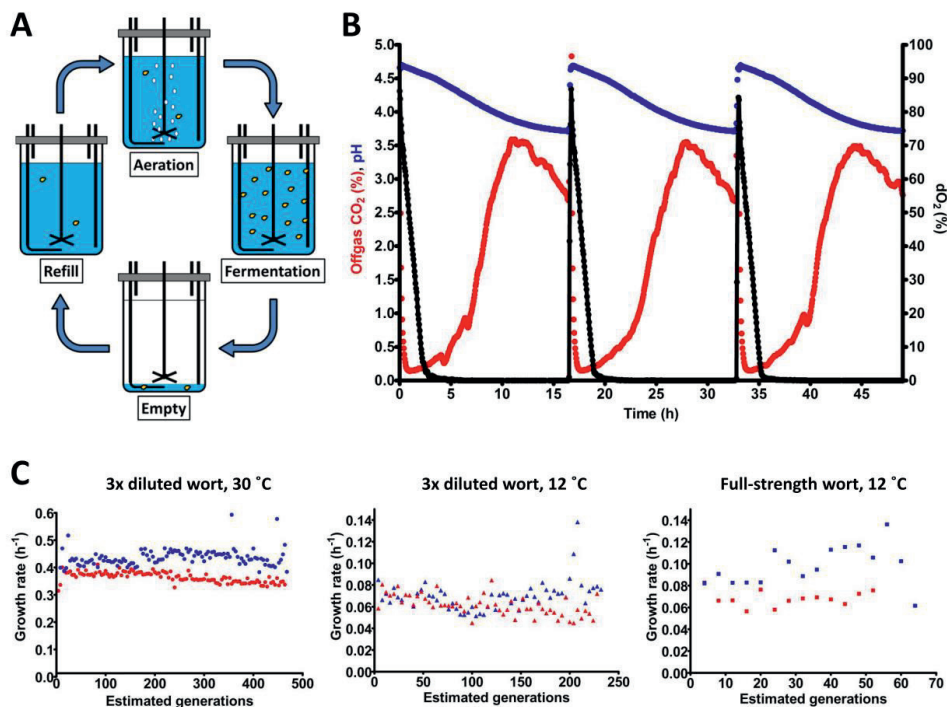
#### **Evaluation of maltotriose fermentation**

Each strain was grown microaerobically in 100 mL serum bottles containing 100 mL medium and shaken at 200 RPM. Medium (full-strength or diluted wort) and incubation temperature (12 °C or 30 °C) were the same as in the evolution experiment from which a strain had been isolated. Strain IMS0408 was included as a control for each condition. Bottles were inoculated to an OD<sub>660</sub> of 0.2 from aerobic shake-flask precultures grown on same medium and under the same conditions. During cultivation, 8 to 12 samples were taken at regular intervals for OD<sub>660</sub> measurements and metabolite analysis by HPLC. When no further sugar consumption was recorded over an interval of at least 48h, the fermentation was considered finished.

## **Results**

### **Simulating domestication under lager-brewing conditions in sequential batch bioreactors**

Industrial lager brewing involves batch cultivation of *S. pastorianus* on wort, an extract from malted barley, at temperatures between 7 and 15 °C. After a brief initial aeration phase to enable oxygen-dependent biosynthesis of unsaturated fatty acids and sterols (44, 45), brewing fermentations are not aerated or stirred, leading to anaerobic conditions during the main fermentation (46). To simulate domestication under industrial lager-brewing conditions, a laboratory evolution regime was designed in which the laboratory-made *S. cerevisiae* x *S. eubayanus* hybrid IMS0408 was grown at 12 °C in sequential batch bioreactors on industrial wort. As in industrial brewing, each cultivation cycle was preceded by an aeration phase, after which cultures were incubated without sparging or stirring until a decline of the CO<sub>2</sub> production indicated a cessation of sugar consumption. The bioreactors were then partially emptied, leaving 7 % of the culture volume as inoculum for the next aeration and fermentation cycle, which was initiated by refilling the reactor with sterile wort (Figure 1A and 1B). To mimic the low sugar concentrations during early domestication of *S. pastorianus* (47), parallel duplicate experiments at 12 °C were not only performed with full-strength 17 °Plato wort ('High Gravity', experiments HG12.1 and HG12.2), but also with three-fold diluted wort ('Low Gravity', experiments LG12.1 and LG12.2). To enable a larger number of generations during 4 months of operation, additional duplicate experiments on three-fold diluted wort were performed at 30 °C (LG30.1 and LG30.2).



**Figure 1: Laboratory evolution mimicking the domestication of lager-brewing yeast.** The *S. cerevisiae* x *S. eubayanus* laboratory hybrid IMS0408 was grown in duplicate sequential batch bioreactors in 3-fold diluted wort at 30 °C (LG30.1 and LG30.2) and at 12 °C (LG12.1 and LG12.2), and in full-strength wort at 12 °C (HG12.1 and HG12.2). (A) Experimental design for simulated sequential lager-beer brewing cycles. Each cycle consisted of four phases: (i) (re)filling of the fermenter with fresh medium up to a total volume of 100 mL, (ii) aeration at 200 mL/min while stirring at 500 RPM, (iii) a batch fermentation phase without sparging or stirring, while flushing the bioreactor headspace with N<sub>2</sub> to enable accurate analysis of CO<sub>2</sub> production and (iv) removal of broth, leaving 7 mL to inoculate the next cycle. (B) Fermentation profiles of three consecutive cycles from experiment LG30.1, performed at 30 °C in 3-fold diluted wort. Percentage of CO<sub>2</sub> in the off gas, culture pH and dissolved oxygen (dO<sub>2</sub>) concentration are indicated by red, blue and black symbols, respectively. Due to the lack of stirring and sparging, CO<sub>2</sub> was slowly released by the medium; emptying of the reactor was initiated when the offgas CO<sub>2</sub> concentration dropped to 70 % of its initial value as off-line analyses indicated that, at this point, all fermentable sugars had been consumed (C) Specific growth estimated from CO<sub>2</sub> production profiles during each cycle of the evolution lines. LG30.1 (blue circles) and LG30.2 (red circles) were grown on 3-fold diluted wort at 30 °C; LG12.1 (blue triangles) and LG12.2 (red triangles) were grown on 3-fold diluted wort at 12 °C. HG12.1 (blue squares) and HG12.2 (red squares) were evolved in full-strength wort at 12 °C. Since lack of sparging and stirring precluded exact estimates of specific growth rates, the calculated values should be taken as indicative.

Concentration of the wort and temperature strongly affected the length of the fermentation cycles, which was 17 h for LG30.1 and LG30.2, 93 h for LG12.1 and LG12.2 and 205 h for HG12.1 and HG12.2. Experiments LG30.1 and LG30.2 involved 117 and 118 batch cycles, respectively, LG12.1 and LG12.2 covered 58 and 57 cycles, respectively, and HG12.1 and HG12.2 covered 13 and 16 cycles, respectively. At the inoculum size of 7 % of the total culture volume, each cycle corresponded to approximately 4 generations. Specific growth rates, estimated from CO<sub>2</sub> production rates during the exponential growth phase of the batch cycles, were not significantly different during the first and the last five cycles of each experiment (Student's t-test,  $p > 0.05$ ). Average specific growth rates

were  $0.35 \pm 0.02 \text{ h}^{-1}$  for LG30.1,  $0.42 \pm 0.03 \text{ h}^{-1}$  for LG30.2,  $0.070 \pm 0.013 \text{ h}^{-1}$  for LG12.1,  $0.062 \pm 0.009 \text{ h}^{-1}$  for LG12.2,  $0.068 \pm 0.007 \text{ h}^{-1}$  for HG12.1 and  $0.098 \pm 0.018 \text{ h}^{-1}$  for HG12.2. While the initial specific growth rate was clearly higher at 30 °C than at 12 °C, initial growth rates on diluted and full-strength wort were not significantly different. However, CO<sub>2</sub> production from sugars continued much longer in full-strength wort. During brewing fermentation, depletion of nitrogen sources and oxygen limit biomass formation. Complete sugar conversion therefore depends on growth-independent alcoholic fermentation which, apparently, was much slower in cultures grown on full-strength wort. At the end of each evolution experiment, culture samples were streaked on YPD agar and 5 single colonies were isolated for each culture. For experiments LG12.1 and LG12.2, isolates were also made from intermediate samples after 29 cycles. Evolution lines LG30.2 and LG12.2 developed flocculence and isolates from these lines had two distinct colony morphologies: about half of the colonies were elevated and conically-shaped, while the other colonies shared the flat morphology of IMS0408 (Figure 3A). For each of these lines, five random colonies of each morphology were selected.

### **Prolonged growth under simulated brewing conditions did not cause large ploidy changes**

Six independent sequential batch fermentation experiments under simulated brewing conditions, covering 52 to 468 generations, yielded 55 isolates. Staining with the DNA-binding fluorescent dye SYTOX Green and flow cytometry indicated genome sizes of the isolates between 17.6 and 23.5 Mbp (Supplementary Table 3). These values did not differ significantly from the  $21.3 \pm 1.9 \text{ Mbp}$  genome size measured for the parental laboratory hybrid IMS0408 and therefore indicated the absence of large changes in genome content such as whole-genome duplications. For a detailed genotypic analysis, the genomes of the 55 isolates were sequenced using 150 bp pair-end reads with 101- to 189-fold coverage. A high quality-reference genome was constructed by combining the chromosome-level contigs from assemblies of CEN.PK113-7D and CBS 12357 generated with nanopore technology, including mitochondrial genome sequences (32, 33).

Copy number analysis revealed whole-chromosome aneuploidies in only 5/55 isolates (Table 2). Relative to strain IMS0408, the total chromosome number of the isolates had not changed by more than one. Isolate IMS0556 (LG30.1) had gained a copy of *ScCHRVIII*, IMS0560 (LG30.1) had gained a copy of *SeCHRX*, IMS0565 (LG30.2) had lost *ScCHRXIV* and gained a copy of *SeCHRXIV*, IMS0595 (LG12.1) had gained a copy of *SeCHRVIII* and IMS0606 (LG12.2) had lost a copy of *SeCHRVIII*.

**Table 2: Overview of phenotypic and genotypic changes in isolates obtained after laboratory evolution of the allopolyploid laboratory hybrid IMS0408 under simulated lager-brewing conditions.**

Experiment	GEN# <sup>1</sup>	Isolate	M <sup>2</sup> F <sup>3</sup>	R <sup>4</sup>	M <sub>DNK</sub> <sup>5</sup>	Aneuploidy <sup>6,7</sup>	Segmental aneuploidy and loss of heterozygosity <sup>6,7</sup>	SNPs <sup>6,8</sup>	INDELS <sup>6,8,9</sup>	
LG1.1 (3x diluted wort, 12 °C)	116	IMS0408	+	-						
		IMS0538	+	-						
		IMS0539	+	-						
		IMS0540	+	-					SeHDAZ <sup>1</sup> 1651T	SeMEDZ <sup>2</sup> 462+3N
		IMS0541	+	-	p <sup>0</sup>				ScIQGI <sup>3</sup> 2069C	SeKEXI <sup>4</sup> 1875-6N
		IMS0542	+	-						
		IMS0543	+	+				ΔSc(YKL032C-YKL054C)		
		IMS0544	+	+				ΔSc(YKL032C-YKL054C), ΔSc::Se(YLR154C-YLR <sub>end</sub> )		
		IMS0545	+	+				ΔSc(YKL032C-YKL054C)	SeSACI <sup>5</sup> 61093C	SeNAFI <sup>1</sup> 1404-30N,1436-30N
		IMS0546	+	+	p <sup>1</sup>			ΔSc(YKL032C-YKL054C)		
		IMS0547	+	+				ΔSc(YKL032C-YKL054C)		
		IMS0594	+	+						
		IMS0595	+	-			2xSe(CHRVIII)	ΔSe::Sc(YOL <sub>end</sub> -YOL072W)		SeELAI <sup>2</sup> 30+343N, SeIRA2 <sup>1</sup> 1402-1N, ScFLO9*
		IMS0596	+	-				ΔSe::Sc(YOL <sub>end</sub> -YOL057W)	SeIRA2 <sup>1</sup> C1376A, ScYER188W <sup>7</sup> 28A	
IMS0597	+	-					SeNUP1 <sup>1</sup> C1205T, ScPDC2 <sup>2</sup> G372A			
232	IMS0598	+	+					SeSRTI <sup>3</sup> C359T	SeASGI <sup>2</sup> 488+3N	
	IMS0599	+	+				ΔSc(YKL032C-YKL054C), ΔSe::Sc(YLR154C-YLR <sub>end</sub> )			
	IMS0600	+	-				ΔSe::Sc(YOL <sub>end</sub> -YOL075W)			
	IMS0601	+	-	p <sup>1</sup>			ΔSc(YKL032C-YKL054C)			
	IMS0602	+	+				ΔSc(YKL032C-YKL054C)			
	IMS0603	+	-				ΔSe::Sc(YNL <sub>end</sub> -YNL123W), ΔSe::Sc(YOL <sub>end</sub> -YOL013C), ΔSc::Se(YOL013C-YOL006C)			
LG1.2 (3x diluted wort, 12 °C)	IMS0548	+	-				ΔSe::Sc(YKL <sub>end</sub> -YKL057C), ΔSc::Se(YKL057C-YKR <sub>end</sub> )			
	IMS0549	+	-	p <sup>1</sup>						
	IMS0550	+	-	p <sup>1</sup>						
	IMS0551	+	-	p <sup>1</sup>			ΔSc(YCL <sub>end</sub> -YCL067C), ΔSc(YCR039C-YCR <sub>end</sub> )			
	IMS0552	+	-	p <sup>1</sup>			ΔSc::Se(YHL <sub>end</sub> -YHL023C)			
	IMS0604	+	-	p <sup>1</sup>			ΔSc(YKL032C-YKL054C), ΔSc::Se(YLR305C-YLR <sub>end</sub> )	SeBETI <sup>2</sup> G550A		
	IMS0605	+	-				ΔSc(YKL032C-YKL054C)			
	IMS0606	+	-	p <sup>1</sup>		ΔSe(CHRVIII)	ΔSc(YKL032C-YKL054C)	ScLRGI <sup>1</sup> C2277G	SeFLO11*	
IMS0607	+	-				ΔSc(YKL032C-YKL054C)				
IMS0608	+	-	p <sup>1</sup>			ΔSc(YKL032C-YKL054C)				

Footnotes: A complete list of mutations is provided in Supplementary Data File 1. <sup>1</sup> Estimated number of generations prior to isolation. <sup>2</sup> Ability to utilise the sugar maltotriose. <sup>3</sup> Flocculation during growth on liquid medium. <sup>4</sup> Respiratory competence. <sup>5</sup> Presence of the mitochondrial DNA. An empty field indicates presence of the *S. eubayanus* mitochondrial DNA, p<sup>0</sup> indicates complete loss of the mitochondrial genome and p<sup>0</sup> partial loss. <sup>6</sup> Sc and Se indicate sequences on the *S. cerevisiae* and *S. eubayanus* subgenomes, respectively. <sup>7</sup> Δ indicates deletion, "2x" indicates duplication and ":" indicates substitution. <sup>8</sup> SNPs and INDELS are only indicated when they affected an ORF and were non synonymous. <sup>9</sup> The number of nucleotides added (+N) or removed (-N) is indicated after the coordinate of the last unchanged nucleotide.

**Table 2 (continued): Overview of phenotypic and genotypic changes in isolates obtained after laboratory evolution of the allopolyploid laboratory hybrid IMS0408 under simulated lager-brewing conditions.**

Experiment	GEN# <sup>1</sup>	Isolate	IM <sup>2</sup>	F <sup>3</sup>	R <sup>4</sup>	M <sub>bioA</sub> <sup>5</sup>	Aneuploidy <sup>6,7</sup>	Segmental aneuploidy and loss of heterozygosity <sup>6,7</sup>	SNPs <sup>6,8</sup>	INDELs <sup>6,9</sup>
LG30.1 (3x diluted wort, 30 °C)	464	IMS0553	+	-	+					SeNHX1 <sup>1,62,2+3N</sup> SCCIS3*
		IMS0554	-	-	+			ΔSc::Se(YGR282C-YGR <sub>end</sub> )	SeBAT1 <sup>G1073A</sup> , ScMAL2 <sup>G422A</sup>	
		IMS0555	+	-	+				SeGMC1 <sup>G1579A</sup> , ScSFL1 <sup>1605A</sup>	
		IMS0556	+	-	+			2xSc(CHRVIII)	ScMAL1 <sup>G88T, A98G</sup> , ScSFL1 <sup>1605A</sup>	ScMAL1 <sup>195-1N</sup> , SeYNL247W <sup>0,62-1N</sup>
		IMS0557	-	-	+				ScMDL2 <sup>4,46,1A</sup>	
		IMS0558	-	+	+				ΔSc(YDR261C-YDR211W), ΔSc::Se(YGR218C-YGR <sub>end</sub> )	SeSFL1 <sup>1605A</sup>
		IMS0559	+	+	+				ΔSc::Sc(YOR133W-YOR <sub>end</sub> )	ScSFL1 <sup>1605A</sup>
LG30.2 (3x diluted wort, 30 °C)	468	IMS0560	+	+	+		2xSe(CHRX)	ΔSc(YDR261C-YDR211W), ΔSc::Se(YOR063W-YOR <sub>end</sub> )	ScSFL1 <sup>1605A</sup>	
		IMS0561	+	+	+			ΔSc::Sc(YBR275C-YBR <sub>end</sub> ), ΔSc::Sc(YOR133W-YOR <sub>end</sub> )	ScSFL1 <sup>1605A</sup>	
		IMS0562	+	+	+			ΔSc::Se(YNL061C-YNL055C), ΔSc::Sc(YOR133W-YOR <sub>end</sub> )	ScSFL1 <sup>1605A</sup>	
		IMS0563	-	-	+			ΔSc(YGR279C-YGR <sub>end</sub> )::Se(YMR305C-YMR <sub>end</sub> )	SeROG <sup>3G191A</sup> , SeMSS51 <sup>C48T</sup>	
		IMS0564	+	-	+				SeIZH3 <sup>A536G</sup> , ScCST6 <sup>C807A</sup>	ScRTP1 <sup>874-64N</sup>
		IMS0565	-	-	+			ΔSc::Se(CHRXIV)	ScMAL1 <sup>1A1G</sup>	
		IMS0566	+	-	+		p		ScERGG <sup>G413T</sup> , ScBUL1 <sup>G2110A</sup>	SeYBR238 <sup>C315-36N</sup> ScUUP2 <sup>14,69+1N</sup>
HG12.1 (Full-strength wort, 12 °C)	52	IMS0567	-	-	+			ΔSc::Se(YDR051C-YDR <sub>end</sub> ), ΔSc::Se(YGR271W-YGR <sub>end</sub> )		
		IMS0609	-	-	+				ScYBR259W <sup>G833A</sup>	
		IMS0610	+	-	+				ScFMP52 <sup>G406C</sup>	
		IMS0611	+	-	+					
HG12.2 (Full-strength wort, 12 °C)	64	IMS0612	+	-	+					
		IMS0613	+	-	+				SeGIC2 <sup>C346G</sup>	
		IMS0614	+	-	+				SeSFL1 <sup>C1390T</sup>	
		IMS0615	+	-	+				ScCAC2 <sup>A994T</sup>	
		IMS0616	+	-	+				ScATG1 <sup>A434C</sup>	
		IMS0617	+	-	+				SeSFL1 <sup>C1390T</sup>	
		IMS0618	+	-	+			ScALR1 <sup>G1645T</sup>	SeGCN2 <sup>227+4-18N, 2239+1-8N</sup> , SeTRAI1 <sup>4421-25N</sup>	

Footnotes: A complete list of mutations is provided in Supplementary Data File 1. <sup>1</sup> Estimated number of generations prior to isolation. <sup>2</sup> Ability to utilise the sugar maltotriose. <sup>3</sup> Flocculation during growth on liquid medium. <sup>4</sup> Respiratory competence. <sup>5</sup> Presence of the mitochondrial DNA. An empty field indicates presence of the *S. eubayanus* mitochondrial DNA, p indicates complete loss of the mitochondrial genome and p<sup>0</sup> partial loss. <sup>6</sup> Sc and Se indicate sequences on the *S. cerevisiae* and *S. eubayanus* subgenomes, respectively. <sup>7</sup> "Δ" indicates deletion, "2x" indicates duplication and ":" indicates substitution. <sup>8</sup> SNPs and INDELs are only indicated when they affected an ORF and were non synonymous. <sup>9</sup> The number of nucleotides added (+N) or removed (-N) is indicated after the coordinate of the last unchanged nucleotide. FLO9<sub>orf</sub>\* and C152<sub>orf</sub>\* indicate coverage drops within these ORFs, for which the responsible mutation could not be exactly reconstructed.



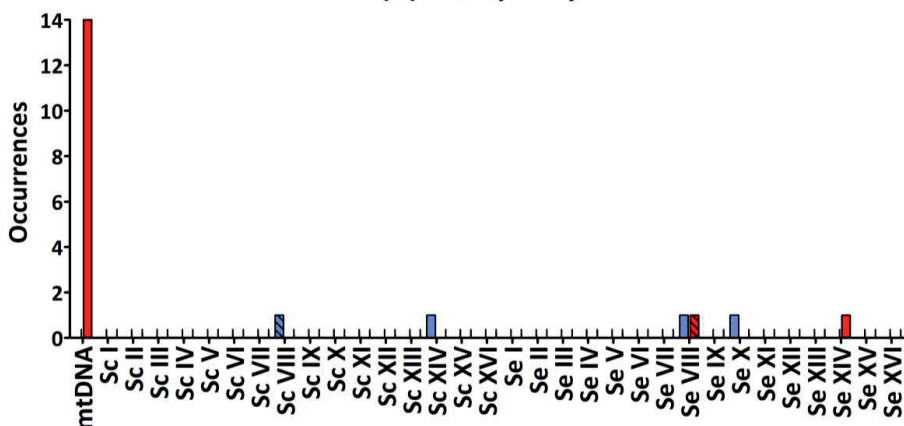


Read alignments to mitochondrial genome sequences were absent from 14/55 isolates, while 1 isolate showed only a partial alignment, indicating complete ( $p^-$ ) or partial loss ( $p^0$ ) of the mitochondrial genome in 15/55 strains. Loss of respiratory competence was confirmed by the observation that these 15 isolates, in contrast to IMS0408 and isolates containing a full mitochondrial genome, were unable to grow on YP-ethanol (Supplementary Figure S1).

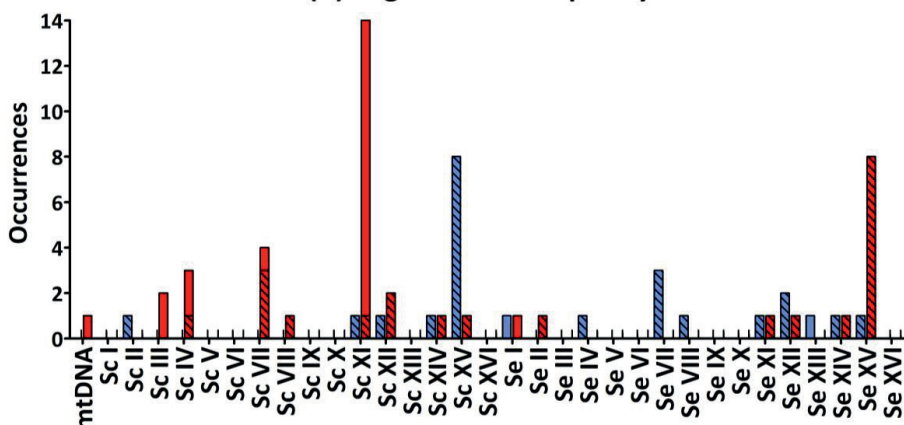
#### **Chromosomal recombinations frequently caused loss of heterozygosity**

Of the 55 evolved isolates, 29 displayed segmental copy number changes. In total, 20 of the 32 chromosomes in strain IMS0408 were affected in at least one isolate. Of the 55 evolved isolates, 24 showed chromosome segments with increased copy number and 41 showed chromosome segments with decreased copy number (Table 2, Figure 2). 17 internal recombinations resulting in deletions were observed:  $\Delta Sc(YKL032C-YKL054C)$  occurred in 13 strains,  $\Delta Sc(YDR261C-YDR211W)$  occurred in two strains, and  $\Delta Sc(YCL_{end}-YCL067C)$  and  $\Delta Sc(YCR039C-YCR_{end})$  occurred together in one strain. The internal recombinations  $\Delta Sc(YKL032C-YKL054C)$  and  $\Delta Sc(YDR261C-YDR211W)$  both resulted in loss of the sequence between the recombination sites. The recombination occurred between *IXR1* and *DEF1* in  $\Delta Sc(YKL032C-YKL054C)$  and between Ty-transposons for  $\Delta Sc(YDR261C-YDR211W)$ . Finally, the concurrent loss of  $Sc(YCL_{end}-YCL067C)$  and  $Sc(YCR039C-YCR_{end})$  indicated loss of both ends of ScCHRIII, including telomeres. This mutation is consistent with a previously-observed circularization of chromosome III by a recombination between the HMLALPHA2 (YCL067C) and MATALPHA2 (YCR039C) loci, leading to loss of both chromosome extremities (48).

### (A) Aneuploidy



### (B) Segmental aneuploidy



**Figure 2: Total number of occurrences of whole-chromosome (A) and segmental (B) aneuploidy for each chromosome of IMS0408 among 55 isolates obtained after laboratory evolution under simulated lager fermentation conditions.** For each chromosome, loss of genetic material is indicated in red and duplicated genetic material is indicated in blue. Loss or duplication of *S. cerevisiae* or *S. eubayanus* genetic material which was coupled with duplication or loss of the corresponding region of the other subgenome, is indicated by checked bars. *S. eubayanus* harbours two translocations relative to *S. cerevisiae*: between chromosomes II and IV, and between chromosomes VIII and XV. For simplicity, copy number affecting these regions were allocated based on the *S. cerevisiae* genome architecture.

The remaining 24 chromosome-segment duplications and losses reflected inter-chromosomal recombinations: one chromosomal region was replaced by an additional copy of another chromosomal region by a non-conservative recombination. The recombinations  $\Delta Sc(YGR279C-YGR_{end})::Se(YMR305C-YMR_{end})$  and  $\Delta Se(YAR050W-YAR_{end})::Se(YAL_{end}-YAL063C)$  occurred between highly similar genes; the paralogs *SCW4* and *SCW10*, and *FLO1* and *FLO9*, respectively. In the remaining 22 cases, recombination occurred between homologous genes of each subgenome. No copy-number conservative chromosome translocations were identified.



Of the 26 observed recombinations, 23 occurred inside ORFs, and thus resulted in chimeric genes (Table 3). The homology between ORFs involved in recombinations varied from <70 % to 100 %, with a median homology of 82.41 %. Chimeric ORFs were reconstructed by extracting reads from one locus affected by the recombination which were paired to the other locus affected by the recombination from the sequencing data, and using them for a local assembly. This approach allowed for identification of the recombination site at a resolution that, depending on sequence homology of the two ORFs, varied between 2 to 633 nucleotides. Due to length differences and relative INDELS between the original ORFs, recombinant ORFs differed in length. However, all recombinations occurred in frame and no premature stop codons were introduced, suggesting that these chimeric ORFs might yield functional proteins.

### ***IRA2*, *SFL1* and *MAL11* are mutated in multiple evolved isolates**

A total of 76 SNPs and 43 INDELS were identified in the genomes of the 55 isolates (Supplementary Data File 1A and 1B), of which 38 SNPs and 17 INDELS occurred in ORFs and were non-synonymous (Table 2). Gene ontology analysis of all genes affected by non-synonymous SNPs or INDELS did not yield a significant enrichment in specific biological processes, molecular functions or cellular components. However, the genes *IRA2*, *SFL1* and *MAL11* were affected in more than one strain. *IRA2* encodes a RAS GTPase-activating protein, which is disrupted in many *S. cerevisiae* genomes from the CEN.PK strain family (49-51). In strain IMS0408, the *ScIRA2* was indeed disrupted while the *SeIRA2* ORF was intact. However, *SeIRA2* was mutated in 6/10 isolates from LG12.1 after 232 generations. *SeIRA2* had a frameshift in IMS0594, a premature stop codon in IMS0596 and was completely lost in four isolates due to different loss of heterozygosities:  $\Delta Se::Sc(YOL_{end}-YOL072W)$  in IMS0595,  $\Delta Se::Sc(YOL_{end}-YOL057W)$  in IMS0597,  $\Delta Se::Sc(YOL_{end}-YOL075W)$  in IMS0600 and  $\Delta Se::Sc(YOL_{end}-YOL013C)$  in IMS0603.

*SFL1* encodes a transcriptional repressor of flocculation genes, which was present both on *ScCHR XV* and *SeCHR VIII* (52). *ScSFL1* was mutated in 6/10 isolates from LG30.1 after 464 generations, which harbored a non-conservative substitution at the 605<sup>th</sup> nucleotide, affecting its DNA binding domain (52). *SeSFL1* had a frameshift in IMS0558 (LG30.1), a single nucleotide substitution in IMS0614 and IMS0617 (HG12.2) and was completely lost in four isolates of LG30.1 due to two losses of heterozygosity:  $\Delta Se::Sc(YOR133W-YOR_{end})$  in IMS0559, IMS0561 and IMS0562, and  $\Delta Se::Sc(YOR063W-YOR_{end})$  in IMS0560 (Table 2).

*ScMAL11*, also referred to as *AGT1*, encodes the only maltose transporter of the *MAL* gene family which enables efficient uptake of maltotriose in IMS0408 (53). *ScMAL11* is located on the right arm of *ScCHR VII* and is absent in the *S. eubayanus* subgenome of IMS0408, which has no other maltotriose transporters (11, 33). *ScMAL11* had a frameshift in IMS0557 (LG30.1) and lost its start codon in IMS0565 (LG30.2) (Table 2). In addition, *ScMAL11* was completely lost due to three different losses of heterozygosity:  $\Delta Sc::Se(YGR282C-YGR_{end})$  in IMS0554 (LG30.1),  $\Delta Sc::Se(YGR218C-YGR_{end})$  in IMS0558 (LG30.1) and  $\Delta Sc::Se(YGR271W-YGR_{end})$  in IMS0567 (LG30.2), and due to the non-conservative recombination  $\Delta Sc(YGR279C-YGR_{end})::Se(YMR305C-YMR_{end})$  in IMS0563 (LG30.2).

**Table 3: Overview of all recombinations observed in 55 isolates obtained after laboratory evolution of strain IMS0408 under simulated lager fermentation conditions.**

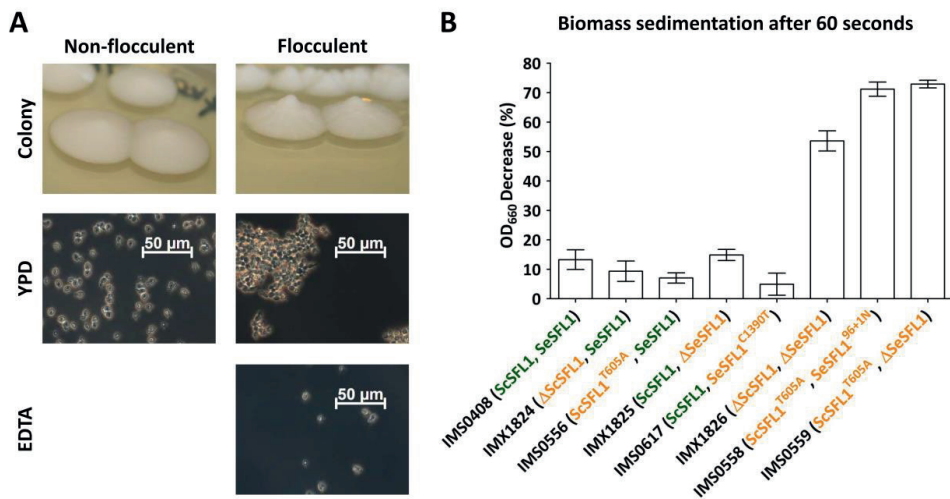
Recombination <sup>1,2</sup>	Affected isolates		Locus 1			Locus 2			Homology <sup>5</sup>		Length chimeric ORF (bp)
	Name <sup>1</sup>	Length <sup>3</sup>	Recombination <sup>4</sup>	Name <sup>1</sup>	Length <sup>3</sup>	Recombination <sup>4</sup>	Length <sup>3</sup>	Recombination <sup>4</sup>	79.23 %	4752	
$\Delta$ Se(YAR050W-YARend) ::Se(YALend-YAL063C)	SeF101	5517	7111-734	SeF109	4752	712-735	-	-	79.23 %	4752	-
$\Delta$ Se::Se(YBR275C-YBRend)	SeRIF1	5751	562-572	SeRIF1	5751	569-579	5751	569-579	77.49 %	5745	-
$\Delta$ Se::Se(YDR051C-YDRend)	SeDET1	1005	426-434	SeDET1	1005	427-435	1005	427-435	84.36 %	1005	-
$\Delta$ Se::Se(YGR218C-YGRend)	SeCRM1	3251	2626-2636	SeCRM1	3251	2627-2637	3251	2627-2637	85.84 %	3251	-
$\Delta$ Se::Se(YGR211W-YGRend)	SeSLH1	5904	1335-1349	SeSLH1	5901	1336-1350	5901	1336-1350	82.41 %	5901	-
$\Delta$ Se(YGR279C-YGRend)	SeSCW4	1146	892-899	SeSCW4	1161	908-915	1161	908-915	<70 %*	1146	-
::Se(YMR305C-YMRend)	SeBGL2	942	115-125	SeBGL2	942	116-126	942	116-126	87.75 %	942	-
$\Delta$ Se::Se(YGR282C-YGRend)	SeNPR3	3444	1551-1571	SeNPR3	3441	1561-1581	3441	1561-1581	78.80 %	3432	-
$\Delta$ Se::Se(YHLend-YHL023C)	SeNUP120	3114	1867-1868	SeNUP120	3114	1868-1869	3114	1868-1869	80.14 %	3114	-
$\Delta$ Se::Se(YK1Lend-YK1057C)	Ribosomal DNA	-	-	Ribosomal DNA	-	-	-	-	-	-	-
$\Delta$ Se::Se(YLR154C-YLRend)	Ribosomal DNA	-	-	Ribosomal DNA	-	-	-	-	-	-	-
$\Delta$ Se::Se(YLR305C-YLRend)	SeST74	5703	5013-5018	SeST74	5703	5014-5019	5703	5014-5019	81.64 %	5703	-
$\Delta$ Se::Se(YLR057C-YKReend)	SeNOP2	1857	1386-1391	SeNOP2	1860	1390-1395	1860	1390-1395	87.11 %	1857	-
$\Delta$ Se::Se(YNL061C-YNL055C)	SePOR1	852	399-401	SePOR1	852	400-402	852	400-402	85.82 %	852	-
$\Delta$ Se::Se(YN1Lend-YN1123W)	SeMMA111	2994	303-314	SeMMA111	2994	304-315	2994	304-315	83.61 %	2994	-
$\Delta$ Se::Se(YO1013C-YO1006C)	SeTOP1	2304	2010-2021	SeTOP1	2311	2017-2028	2311	2017-2028	82.64 %	2304	-
$\Delta$ Se::Se(YO1Lend-YO1013C)	SeHRD1	1644	312-329	SeHRD1	1656	313-330	1656	313-330	82.82 %	1656	-
$\Delta$ Se::Se(YO1Lend-YO1057W)	SeYOL057W	3015	99-116	SeYOL057W	2133	100-117	2133	100-117	<70 %*	2133	-
$\Delta$ Se::Se(YO1Lend-YO1072W)	SeTHP1	1368	726-731	SeTHP1	1374	733-738	1368	733-738	79.07 %	1368	-
$\Delta$ Se::Se(YO1Lend-YO1075W)	SeYOL075C	3903	2409-2414	SeYOL075C	3885	2392-2397	3885	2392-2397	81.85 %	3903	-
$\Delta$ Se::Se(YOR063W-YORend)	SeRPL3	1164	1023-1063	SeRPL3	1164	1024-1064	1164	1024-1064	94.59 %	1164	-
$\Delta$ Se::Se(YOR133W-YORend)	SeEFT1	2529	246-311	SeEFT1	2529	247-312	2529	247-312	94.94 %	2529	-
$\Delta$ Se(YCLend-YCL067C)	SCHML1ALPHA2	633	1-633	SCHML1ALPHA2	633	1-633	633	1-633	100 %	633	-
$\Delta$ Se(YCR039C-YCRend)	TY-transposon	-	-	TY-transposon	-	-	-	-	-	-	-
$\Delta$ Se(YDR261C-YDR211W)	SeDFE1	1794	944-955	SeDFE1	2217	1281-1292	2217	1281-1292	<70 %*	1881	-
$\Delta$ Se(YKLO032C-YKLO54C)	SeXRI	1794	332-341	SeDFE1	2217	1272-1281	2217	1272-1281	<70 %*	1278	-
$\Delta$ Se(YKLO032C-YKLO54C)	SeXRI	1794	332-341	SeDFE1	2217	1272-1281	2217	1272-1281	<70 %*	1278	-

Footnotes: <sup>1</sup> Sc and Se indicate sequences on the *S. cerevisiae* and *S. eubayanus* subgenomes, respectively. <sup>2</sup> "Δ" indicates deletion and "::" indicates substitution. <sup>3</sup> Length of the affected ORF in bp. <sup>4</sup> Nucleotide coordinates of the locus at which recombination could have occurred. Due to the high homology of Locus 1 and Locus 2, they shared identical sequence stretches within which the recombination could have occurred. Recombination sites are therefore indicated by a range of nucleotide coordinates within the ORFs. <sup>5</sup> Homology between Locus 1 and Locus 2 as determined by sequence alignment.



### Mutations in *SFL1* cause emergence of calcium-dependent flocculation

Eight of the 55 evolved isolates showed SNPs, INDELS or loss of heterozygosity in the flocculation inhibitor gene *SFL1*. In isolates IMS0558, IMS0559, IMS0560, IMS0561 and IMS0562 from LG30.1, both *SeSFL1* and *ScSFL1* were mutated, while in isolate IMS0556 from LG30.1 only *ScSFL1* was mutated and in isolates IMS0614 and IMS0617 from HG12.2, only *SeSFL1* was mutated. Evolved isolates that carried mutations in both *ScSFL1* and *SeSFL1* formed elevated conically-shaped colonies on YPD agar, while strain IMS0408 and evolved isolates with either an intact *SeSFL1* or *ScSFL1* did not (Figure 3A). Strains with mutations in both *ScSFL1* and *SeSFL1* also showed rapid sedimentation in micro-aerobic cultures on wort, which was not observed for the other evolved isolates or for strain IMS0408 (Figure 3B).



**Figure 3: Mutations in *ScSFL1* and *SeSFL1* correlate with flocculation in evolved isolates and reverse engineered strains.** A) Colony morphology and phase-contrast microscopy images (100x) of YPD-grown cell suspensions of the non-evolved, non-flocculent strain (IMS0408) and of a typical flocculent evolved isolate (IMS0558). Resuspension in 50 mM EDTA (pH 7.0) eliminated flocculation. B) Biomass sedimentation of evolved isolates and engineered strains with mutations in *SeSFL1* and/or *ScSFL1*. Triplicate cultures of all strains were grown on YPD and sedimentation was measured as the decrease in OD<sub>660</sub> right underneath the meniscus of a stationary cell suspensions 60 s after the suspension had been vortexed.

### Mutations in *ScMAL11* cause loss of maltotriose utilisation

All evolved isolates and the unevolved hybrid IMS0408 were grown under brewing conditions at the temperature and wort gravity used during their evolution in micro-aerobic 100 mL bottles to assess their brewing performance. Besides differences in flocculation behavior described above, only 6 isolates evolved on diluted wort at 30 °C displayed significantly different brewing performances relative to IMS0408. These six strains all harbored mutations to the *ScMAL11* gene, which encodes the sole maltotriose transporter in strain IMS0408. The unevolved IMS0408 consumed 100 ± 0 % of the maltotriose in diluted wort, and evolved strains with an intact *ScMAL11* genes consumed 98 ± 3 %. In contrast, strains IMS0554, IMS0557, IMS0558, IMS0563, IMS0565 and IMS0567, which all harbored mutations in *ScMAL11*, did not show any maltotriose consumption; instead, the concentration increased by 14 ± 3 % on average, presumably due to water evaporation. To test if the mutations affecting *ScMAL11* were responsible for the loss of maltotriose utilization, *ScMAL11* was deleted in strain IMS0408 using CRISPR-Cas9 gene editing, resulting in strain

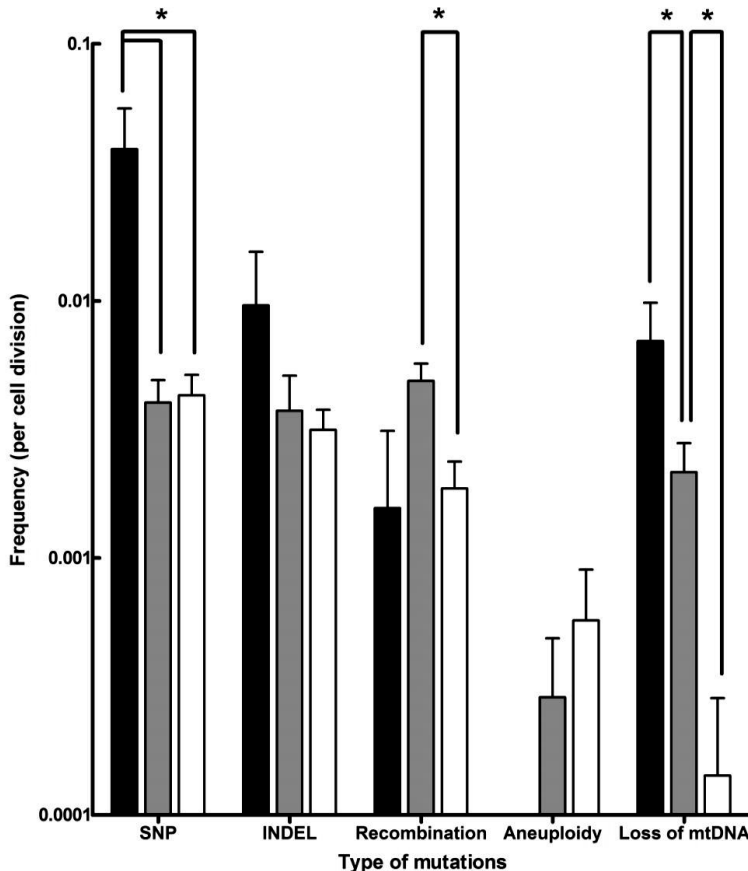
IMX1698 ( $\Delta ScMAL11::mVenus$ ). Under the same conditions used to evaluate maltotriose utilization by the evolved strains, strain IMS0408 consumed  $97 \pm 5$  % of the maltotriose while strain IMX1698 only consumed  $1 \pm 0$  % of the maltotriose. These results confirmed that loss of *ScMAL11* function was responsible for loss of maltotriose utilization.

## Discussion

Evolution of the laboratory *S. cerevisiae* x *S. eubayanus* IMS0408 under simulated lager-brewing conditions yielded a wide array of mutations, including SNPs, INDELS, chromosomal recombinations, aneuploidy and loss of mitochondrial DNA (Table 2, Additional Data File 1). SNPs were the most common type of mutation, with frequencies ranging between 0.004 and 0.039 per division (Figure 4). Based on a genome size of 24.2 Mbp, the rate at which single-nucleotide mutations occurred was between  $1.7 \cdot 10^{-10}$  and  $1.6 \cdot 10^{-9}$  per nucleotide per cell division, which is similar to a rate of  $3.3 \cdot 10^{-10}$  per site per cell division reported for *S. cerevisiae* (Lynch et al. 2008). At a frequency of 0.003 and 0.010 per cell division, INDELS occurred up to 4.1-fold less frequently than SNPs, in accordance with their twofold lower occurrence in *S. cerevisiae* (54). The higher incidence of both SNPs and INDELS in isolates evolved in full-strength wort (Figure 4) may be related to the higher concentrations of ethanol, a known mutagen, in these cultures (55). The rate of loss of mitochondrial DNA varied between 0.0001 and 0.007 per division (Figure 4), and was negatively correlated with the number of generations of selective growth, indicating loss of mitochondrial DNA is selected against. The percentage of respiratory deficient isolates, between 13 and 40 % for all evolutions, is consistent with observations during laboratory evolution under oxygen limitation, and has been associated with increased ethanol tolerance (56, 57). However, loss of respirative capacity is highly undesirable for lager brewing as it impedes biomass propagation (58).

The frequency of chromosomal recombinations was estimated between 0.002 and 0.005 per division (Figure 4), which is similar to frequencies reported for *S. cerevisiae* (59). The observed recombinations were not reciprocal translocations. Instead, in all cases, genetic material was lost due to internal deletions, or genetic material from one chromosome was replaced by an additional copy of genetic material from another chromosome. This abundant loss of heterozygosity is consistent with the evolutionary history of *S. cerevisiae* (60) and with previously observed loss of genetic material in hybrids (17, 61-63). Under selective conditions, regions from one subgenome can be preferentially affected by loss of heterozygosity due to the fitness effects of genes they harbor. Due to its irreversibility, loss of heterozygosity is a determining mechanism for the evolution of hybrids during domestication (26). For example, in an *S. cerevisiae* x *S. uvarum* hybrid, the chromosomal region harboring *ScPHO84* was preferentially retained at 30 ° C, while it was lost at the expense of its *S. uvarum* homolog at 15 ° C (62, 63). Similarly, the superior growth of *S. cerevisiae* at 30 ° C relative to *S. paradoxus* could be linked to *S. cerevisiae* alleles of 8 genes by investigating the effect of loss of heterozygosity in a laboratory hybrid (64). Moreover, the deletion of *S. cerevisiae* alleles in *S. uvarum* x *S. cerevisiae* had strong and varying impacts on fitness under glucose-, sulphate- and phosphate-limitation (65). Overall, loss of heterozygosity is an irreversible process which enables rapid adaptation of hybrid genomes to the selective pressure of their growth environment. In the present study, loss of heterozygosity notably affected *ScMAL11* and *SeSFL1*, contributing to the acquired flocculation and loss of maltotriose utilization phenotypes. Other observed losses of heterozygosity may be due to genetic drift, or they may yield a selective advantage which we did not identify. Loss of heterozygosity may be advantageous by removing unfavorable dominant alleles, by

enabling the expression of favorable recessive alleles, or by resolving redundancy, cross-talk and possible incompatibility between alleles and genes of both subgenomes (66, 67).



**Figure 4: Frequencies of different types of mutations observed in evolved isolates obtained after laboratory evolution of strain IMS0408 under simulated lager fermentation conditions.** The mutations identified in all 55 isolates evolved under brewing conditions were classified by type, and the frequency of mutation per cell division was calculated for each isolate based on its estimated number of generations of growth under simulated brewing conditions. Average frequencies of mutation types and standard deviations are shown for isolates evolved on full-strength wort at 12 °C (black), on three-fold diluted wort at 12 °C (grey) and on threefold diluted wort at 30 °C (white). The frequencies are shown on a logarithmic scale, and p-values were determined using Student's t-test.

Whole-chromosome aneuploidy was the rarest type of mutation, occurring only between 0 and 0.0003 times per division (Figure 4). The emergence of single chromosome aneuploidy in 9 % of the isolates after laboratory evolution is similar to observations during laboratory evolution of *S. cerevisiae* (20, 24, 55, 68). However, the observed small extent of aneuploidy starkly contrasts with the massive aneuploidy of *S. pastorianus* brewing strains (18, 19). These differences might of course be attributed to the difference in time scale between four months of laboratory evolution and several centuries of domestication. However, they might also be due to differences between the *S. cerevisiae* x *S. eubayanus* IMS0408 evolved in this study and the ancestral *S. pastorianus* hybrid. Firstly, IMS0408 was obtained by crossing a haploid *S. cerevisiae* strain with a haploid *S. eubayanus*

spore (11). In contrast, the progenitor of brewing strains of *S. pastorianus* may have resulted from a cross of higher ploidy strains (19). Since higher ploidy leads to higher chromosome missegregation rates, they accumulate chromosome copy number changes faster (69). In addition, the higher initial ploidy leads to a smaller relative increase of genetic material when an additional copy is gained, which may 'buffer' deleterious effects of further changes in ploidy (70). Moreover, this laboratory-made hybrid was constructed by crossing a laboratory *S. cerevisiae* strain with the first discovered *S. eubayanus* strain, which may differ considerably from the genetic background of the initial *S. pastorianus* hybrid (19). Therefore, the ancestral *S. pastorianus* hybrid may have had, or have acquired, mutations that stimulate extensive aneuploidy, such as mutations increasing the rate of chromosome missegregation or mutations increasing the tolerance against aneuploidy associated stresses (71). Finally, aneuploidy may have emerged in *S. pastorianus* by an (aborted) sporulation event, as such events can cause uneven chromosome segregation even in non-hybrid polyploid yeasts (72). Regardless of the origin of the extensive aneuploidy of *S. pastorianus*, our results show that euploid *S. cerevisiae* x *S. eubayanus* hybrids are not by definition prone to extensive aneuploidy under brewing-related experimental conditions. For industrial applications, the relative genetic stability of newly generated *Saccharomyces* hybrid strains reduces the chance of strain deterioration during the many generations involved in large-scale fermentation and/or biomass recycling (12). Due to the industry practice of limiting yeast biomass re-pitching, the genetic and phenotypic stability observed during 52 generations in 17 °P wort at 12 °C is sufficient to warrant sufficient strain stability for lager brewing applications.

In addition to showing extensive aneuploidy, *S. pastorianus* strains harbor numerous chromosomal recombinations. During the laboratory evolution experiments, two types of recombinations were observed: (i) intrachromosomal recombinations resulting in loss of chromosome segments, and (ii) interchromosomal recombinations resulting in loss of one chromosome segment and replacement by an additional copy of a segment from another chromosome, resulting in loss of heterozygosity. While in *S. cerevisiae*, chromosomal recombinations predominantly occur in repetitive regions of the genome (59, 73), here 88 % of the observed recombinations occurred within ORFs. The average homology of the recombined ORFs did not exceed the average 85 % homology of ORFs in the *S. cerevisiae* and *S. eubayanus* subgenomes. Instead, the high rate of recombinations at ORFs could reflect a correlation between transcriptional activity and recombination (74). In all cases, the reading frames were conserved, resulting in chimeric ORFs which could encode functional chimeric proteins, with altered length and sequences compared to the parental genomes. The potential selective advantage of such chimeric proteins is illustrated by recurring recombinations between ammonium permease *MEP2* alleles in a *S. cerevisiae* and *S. uvarum* hybrid during laboratory evolution under nitrogen limitation conditions (75). The formation of chimeric ORFs has even led to the emergence of novel gene functions, as illustrated by the formation of a maltotriose transporter by recombination of three non-maltotriose transporter genes in *S. eubayanus* during laboratory evolution (76).

The predominant occurrence of recombinations within ORFs in the evolved isolates has also been observed in the genomes of brewing strains of *S. pastorianus*, which all share identical recombinations at the *ZUO1*, *MAT*, *HSP82* and *XRN1/KEM1* loci (19, 77, 78). These common recombinations suggest that all *S. pastorianus* isolates descend from a common ancestor (79). However, *S. pastorianus* strains may also have emerged from two independent hybridization events, as suggested by the presence of two genetically distinct groups within *S. pastorianus* (80, 81) Indeed, since identical recombinations have been observed in independent evolutions, identical



recombinations might reflect parallel evolution due to a strong selective advantage under brewing-related conditions and/or a predisposition of specific loci for recombination (59, 75). Of the recombination loci found in the present study, only *EFT1* and *MAT* loci were associated with recombinations in *S. pastorianus*. Moreover, the recombinations at these loci in the evolved isolates were different from those in *S. pastorianus* (19, 77, 78, 82). All interchromosomal recombinations observed in this study were unique. These results, obtained under brewing-related conditions, are consistent with the notion that recombination sites are largely aleatory and that all modern *S. pastorianus* strains share the same recombinations because they descend from a single hybrid ancestor.

Different recombination events resulted in the loss of heterozygosity, in four isolates each, of the right arm of *SeCHRXV*, including *SeSFL1*, and of the right arm of *ScCHRXI*, including *ScMAL11*. These events directly affected two phenotypes relevant for brewing fermentation: calcium-dependent flocculation, which led to fast biomass sedimentation, and loss of maltotriose utilization. Biomass sedimentation can be strongly selected for in sequential batch bioreactors, as it increases the chance that cells are retained in the bioreactor during the emptying phase (83, 84). A similar selective advantage is likely to have played a role in the early domestication of *S. pastorianus*, as sedimenting yeast remaining in fermentation vessels was more likely to be used in a next fermentation. Flocculation is a key characteristic of current lager-brewing yeasts (also referred to as bottom-fermenting yeasts), as it simplifies biomass separation at the end of the fermentation (85). The present study illustrates how this aspect of brewing yeast domestication can be rapidly reproduced under simulated laboratory conditions.

At first glance, loss of the ability to utilize maltotriose, an abundant fermentable sugar in wort, appears to be undesirable from an evolutionary perspective. However, as demonstrated in studies on laboratory evolution of *S. cerevisiae* in sequential batch cultures on sugar mixtures, the selective advantage of consuming a specific sugar from a sugar mixture correlates with the number of generations of growth on that sugar during each cultivation cycle (86, 87). *Saccharomyces* yeasts, including strain IMS0408 generally prefer glucose and maltose over maltotriose (11, 53, 88). As a consequence, maltotriose consumption from wort typically only occurs when growth has already ceased due to oxygen limitation and/or nitrogen source depletion, which results in few or no generations of growth on this trisaccharide. However, loss of maltotriose utilization in six isolates in two independent evolution experiments strongly suggests that loss of *ScMAL11* expression was not merely neutral but even conferred a selective advantage. These results are consistent with the existence of many *S. pastorianus* strains with poor maltotriose utilization and with the truncation of *ScMAL11* in all *S. pastorianus* strains, including good maltotriose utilizers (89, 90). In the latter strains, maltotriose utilization depends on alternative transporters such as *Mty1* (91, 92). It is therefore unclear if a selective advantage of the loss of *ScMAL11* reflects specific properties of this gene or its encoded transporter or, alternatively, a general negative impact of maltotriose utilization under brewing-related conditions. In analogy with observations on maltose utilization by *S. cerevisiae*, unrestricted *Mal11*-mediated maltotriose-proton symport might cause maltotriose-accelerated death (93, 94). Alternatively, expression of the *Mal11* transporter might compete with superior maltose transporters for intracellular trafficking, membrane integration and/or membrane space (6, 95). Indeed, laboratory evolution to obtain improved maltotriose utilization resulted in reduced maltose uptake in *S. pastorianus* (88).

*Saccharomyces* hybrids are commonly applied for industrial applications such as lager beer brewing and wine fermentation (6, 8, 10). Recently, novel hybrids have been generated, that performed well under a broad range of industrially-relevant conditions (11, 12, 14-17). While the performance and phenotypic diversity of laboratory hybrids support their application in industrial processes, further strain development of such hybrids could improve their performance (12, 96). Especially in food and beverage fermentation processes, consumer acceptance issues largely preclude use of targeted genetic modification techniques. Laboratory evolution offers an interesting alternative strategy for 'non-GMO' strain improvement (97). However, as exemplified by the loss, in independent laboratory evolution experiments, of *MAL11* and of the mitochondrial genome, mutations that yield increased fitness under simulated industrial fermentation conditions are not necessarily advantageous for industrial performance. Therefore, instead of faithfully reconstructing industrial conditions in the laboratory, laboratory evolution experiments should be designed to specifically select for desired phenotypes. For example, a recent study illustrated how maltotriose fermentation kinetics of an *S. pastorianus* hybrid could be improved by laboratory evolution in carbon-limited chemostats grown on a maltotriose-enriched sugar mixture (88).

### Acknowledgments

We thank Erik de Hulster, Xavier Hakkaart and Robert Mans for sharing their expertise in fermentation, Marijke Luttk for her expertise in Flow cytometry and Nikola Gyurchev for assembling plasmids pUDP104 and pUDP105. We are thankful to Niels Kuijpers (Heineken Supply Chain B.V.), and Jan-Maarten Geertman (Heineken Supply Chain B.V.) for their support and for critically reading the manuscript. A pre-print version of this article has been uploaded on bioRxiv (98).

### References

1. Botstein D, Chervitz SA, & Cherry M (1997) Yeast as a model organism. *Science* 277(5330):1259-1260.
2. Sheltzer JM, et al. (2011) Aneuploidy drives genomic instability in yeast. *Science* 333(6045):1026-1030.
3. Naseeb S, et al. (2017) *Saccharomyces jurei* sp nov., isolation and genetic identification of a novel yeast species from *Quercus robur*. *Int J Syst Evol Micr* 67(6):2046-2052.
4. Liti G, Barton DB, & Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174(2):839-850.
5. Hittinger CT (2013) *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* 29(5):309-317.
6. Libkind D, et al. (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. USA* 108(35):14539-14544.
7. Marsit S & Dequin S (2015) Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. *FEMS Yeast Res.* 15(7):fov067-fov067.
8. González SS, Barrio E, Gafner J, & Querol A (2006) Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res.* 6(8):1221-1234.
9. Querol A & Bond U (2009) The complex and dynamic genomes of industrial yeasts. *FEMS Microbiol. Lett.* 293(1):1-10.
10. Naumov GI, et al. (2001) Genetic identification of *Saccharomyces bayanus* var. *uvarum*, a cider-fermenting yeast. *Int. J. Food Microbiol.* 65(3):163-171.
11. Hebly M, et al. (2015) *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res.* 15(3):fov005.
12. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2017) Novel brewing yeast hybrids: creation and application. *Appl. Microbiol. Biotechnol.* 101(1):65-78.
13. Coloretti F, Zambonelli C, & Tini V (2006) Characterization of flocculent *Saccharomyces* interspecific hybrids for the production of sparkling wines. *Food Microbiol.* 23(7):672-676.
14. Lopandic K, et al. (2016) Genotypic and phenotypic evolution of yeast interspecies hybrids during high-sugar fermentation. *Appl. Microbiol. Biotechnol.* 100(14):6331-6343.
15. Bellon JR, et al. (2011) Newly generated interspecific wine yeast hybrids introduce flavour and aroma diversity to wines. *Appl. Microbiol. Biotechnol.* 91(3):603-612.
16. Nikulin J, Krogerus K, & Gibson B (2017) Alternative *Saccharomyces* interspecies hybrid combinations and their potential for low-temperature wort fermentation. *Yeast* 35(1):113-127.

17. Peris D, *et al.* (2017) Hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production. *Biotechnol. Biofuels* 10(1):78.
18. Van den Broek M, *et al.* (2015) Chromosomal copy number variation in *Saccharomyces pastorianus* is evidence for extensive genome dynamics in industrial lager brewing strains. *Appl. Environ. Microbiol.* 81(18):6253-6267.
19. Okuno M, *et al.* (2016) Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res.* 23(1):67-80.
20. Gorter de Vries AR, Pronk JT, & Daran J-MG (2017) Industrial relevance of chromosomal copy number variation in *Saccharomyces yeasts*. *Appl. Environ. Microbiol.* 83(11):e03206-03216.
21. Chambers SR, Hunter N, Louis EJ, & Borts RH (1996) The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol. Cell Biol.* 16(11):6110-6120.
22. Santaguida S & Amon A (2015) Short-and long-term effects of chromosome mis-segregation and aneuploidy. *Nat. Rev. Mol. Cell. Biol.* 16(8):473.
23. Gorter de Vries AR, de Groot PA, van den Broek M, & Daran J-MG (2017) CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb. Cell Fact.* 16(1):222.
24. Yona AH, *et al.* (2012) Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. USA* 109(51):21010-21015.
25. Delneri D, *et al.* (2003) Engineering evolution to study speciation in yeasts. *Nature* 422(6927):68.
26. Pérez Través L, Lopes CA, Barrio Esparducer E, & Querol Simón A (2014) Stabilization process in *Saccharomyces* intra and interspecific hybrids in fermentative conditions. *Int. Microbiol.*
27. Gibbons JG & Rinker DC (2015) The genomics of microbial domestication in the fermented food environment. *Curr. Opin. Genet. Dev.* 35:1-8.
28. Gibbons JG, *et al.* (2012) The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. *Curr. Biol.* 22(15):1403-1409.
29. Bachmann H, Starrenburg MJ, Molenaar D, Kleerebezem M, & van Hylckama Vlieg JE (2012) Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. *Genome Res.* 22(1):115-124.
30. Verduyn C, Postma E, Scheffers WA, & Van Dijken JP (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8(7):501-517.
31. Solis-Escalante D, *et al.* (2013) amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 13(1):126-139.
32. Salazar AN, *et al.* (2017) Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN. PK113-7D. *FEMS Yeast Res.* 17(7).
33. Brickwedde A, *et al.* (2018) Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357<sup>T</sup>. *Front. Microbiol.* 9:1786.
34. Wach A, Brachat A, Pöhlmann R, & Philippsen P (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* 10(13):1793-1808.
35. Li H & Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26(5):589-595.
36. Li H, *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
37. Walker BJ, *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
38. Chen K, *et al.* (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* 6(9):677.
39. Robinson JT, *et al.* (2011) Integrative genomics viewer. *Nat. Biotechnol.* 29(1):24.
40. Bankevich A, *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19(5):455-477.
41. Haase SB & Reed SI (2002) Improved flow cytometric analysis of the budding yeast cell cycle. *Cell cycle* 1(2):117-121.
42. Engler C, Kandzia R, & Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS one* 3(11):e3647.
43. Gorter de Vries AR, *et al.* (2019) Allele-specific genome editing using CRISPR-Cas9 is associated with loss of heterozygosity in diploid yeast. *Nucleic Acids Res* 47(3):1362-1372.
44. Andreassen AA & Stier T (1953) Anaerobic nutrition of *Saccharomyces cerevisiae*. I. Ergosterol requirement for growth in a defined medium. *J Cell. Comp. Physiol.* 41(1):23-36.
45. Gibson BR, Lawrence SJ, Leclaire JP, Powell CD, & Smart KA (2007) Yeast responses to stresses associated with industrial brewery handling. *FEMS Microbiol. Lett.* 31(5):535-569.
46. Briggs DE, Brookes P, Stevens R, & Boulton C (2004) *Brewing: science and practice* (Elsevier).
47. Meussdoerffer FG (2009) A comprehensive history of beer brewing. *Handbook of brewing: Processes, technology, markets*:1-42.
48. Newlon C, *et al.* (1991) Analysis of a circular derivative of *Saccharomyces cerevisiae* chromosome III: a physical map and identification and location of ARS elements. *Genetics* 129(2):343-357.

49. Nijkamp JF, *et al.* (2012) *De novo* sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* 11(1):36.
50. Tanaka K, *et al.* (1990) *S. cerevisiae* genes *IRA1* and *IRA2* encode proteins that may be functionally equivalent to mammalian ras GTPase activating protein. *Cell* 60(5):803-807.
51. Tanaka K, Lin BK, Wood DR, & Tamanoi F (1991) *IRA2*, an upstream negative regulator of *RAS* in yeast, is a *RAS* GTPase-activating protein. *Proc. Natl. Acad. Sci. USA* 88(2):468-472.
52. Atsushi F, *et al.* (1989) Domains of the *SFL1* protein of yeasts are homologous to Myc oncoproteins or yeast heat-shock transcription factor. *Gene* 85(2):321-328.
53. Alves SL, *et al.* (2008) Molecular analysis of maltotriose active transport and fermentation by *Saccharomyces cerevisiae* reveals a determinant role for the *AGT1* permease. *Appl. Environ. Microbiol.* 74(5):1494-1501.
54. Lang GI & Murray AW (2008) Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* 178(1):67-82.
55. Voordeckers K, *et al.* (2015) Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genet.* 11(11):e1005635.
56. Taylor DR, Zeyl C, & Cooke E (2002) Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 99(6):3690-3694.
57. Ibeas JI & Jimenez J (1997) Mitochondrial DNA loss caused by ethanol in *Saccharomyces flor* yeasts. *Appl. Environ. Microbiol.* 63(1):7-12.
58. Gibson B, Prescott K, & Smart K (2008) Petite mutation in aged and oxidatively stressed ale and lager brewing yeast. *Lett. Appl. Microbiol.* 46(6):636-642.
59. Dunham MJ, *et al.* (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 99(25):16144-16149.
60. Magwene PM, *et al.* (2011) Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 108(5):1987-1992.
61. Sipiczki M (2008) Interspecies hybridization and recombination in *Saccharomyces* wine yeasts. *FEMS Yeast Res.* 8(7):996-1007.
62. Smukowski Heil CS, Large CR, Patterson K, & Dunham MJ (2018) Temperature preference biases parental genome retention during hybrid evolution. *bioRxiv [Preprint]:*429803.
63. Smukowski Heil CS, *et al.* (2017) Loss of heterozygosity drives adaptation in hybrid yeast. *Mol. Biol. Evol.* 34(7):1596-1612.
64. Weiss CV, *et al.* (2018) Genetic dissection of interspecific differences in yeast thermotolerance. *Nat. Genet.*:1.
65. Lancaster SM, Payen C, Heil CS, & Dunham MJ (2018) Fitness benefits of loss of heterozygosity in *Saccharomyces* hybrids. *bioRxiv [Preprint]:*452748.
66. Gibson B & Liti G (2015) *Saccharomyces pastorianus*: genomic insights inspiring innovation for industry. *Yeast* 32(1):17-27.
67. Piatkowska EM, Naseeb S, Knight D, & Delneri D (2013) Chimeric protein complexes in hybrid species generate novel phenotypes. *PLoS Genet.* 9(10):e1003836.
68. González-Ramos D, *et al.* (2016) A new laboratory evolution approach to select for constitutive acetic acid tolerance in *Saccharomyces cerevisiae* and identification of causal mutations. *Biotechnol. Biofuels* 9(1):173.
69. Storchova Z (2014) Ploidy changes and genome stability in yeast. *Yeast* 31(11):421-430.
70. Torres EM, *et al.* (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317(5840):916-924.
71. Torres EM, *et al.* (2010) Identification of aneuploidy-tolerating mutations. *Cell* 143(1):71-83.
72. Kim SR, *et al.* (2017) Metabolic engineering of a haploid strain derived from a triploid industrial yeast for producing cellulosic ethanol. *Metabolic engineering* 40:176-185.
73. Fontdevila A (2005) Hybrid genome evolution by transposition. *Cytogenet. Genome Res.* 110(1-4):49-55.
74. Thomas BJ & Rothstein R (1989) Elevated recombination rates in transcriptionally active DNA. *Cell* 56(4):619-630.
75. Dunn B, *et al.* (2013) Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet.* 9(3):e1003366.
76. Brouwers N, *et al.* (2019) In vivo recombination of *Saccharomyces eubayanus* maltose-transporter genes yields a chimeric transporter that enables maltotriose fermentation. *PLoS genetics* 15(4):e1007853.
77. Hewitt SK, Donaldson IJ, Lovell SC, & Delneri D (2014) Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. *PLoS One* 9(3):e92203.
78. Walther A, Hesselbart A, & Wendland J (2014) Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. *G3 (Bethesda)* 4(5):783-793.
79. Monerawela C & Bond U (2018) The hybrid genomes of *Saccharomyces pastorianus*: A current perspective. *Yeast* 35(1):39-50.
80. Dunn B & Sherlock G (2008) Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res.* 18(10):1610-1623.
81. Baker EC, *et al.* (2015) The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* 32(11):2818-2831.

82. Monerawela C & Bond U (2017) Recombination sites on hybrid chromosomes in *Saccharomyces pastorianus* share common sequence motifs and define a complex evolutionary relationship between group I and II lager yeasts. *FEMS Yeast Res.* 17(5):fox047-fox047.
83. Hope EA, et al. (2017) Experimental evolution reveals favored adaptive routes to cell aggregation in yeast. *Genetics:genetics.* 116.198895.
84. Oud B, et al. (2013) Genome duplication and mutations in *ACE2* cause multicellular, fast-sedimenting phenotypes in evolved *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA:*201305949.
85. Ferreira I, Pinho O, Vieira E, & Tavarella J (2010) Brewer's *Saccharomyces* yeast biomass: characteristics and potential applications. *Trends Food Sci. Technol.* 21(2):77-84.
86. Wisselink HW, Toirkens MJ, Wu Q, Pronk JT, & van Maris AJA (2009) Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered *Saccharomyces cerevisiae* strains. *Appl. Environ. Microbiol.* 75(4):907-914.
87. Verhoeven MD, de Valk SC, Daran J-MG, van Maris AJ, & Pronk JT (2018) Fermentation of glucose-xylose-arabinose mixtures by a synthetic consortium of single-sugar-fermenting *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.* 18(8):foy075.
88. Brickwedde A, et al. (2017) Evolutionary engineering in chemostat cultures for improved maltotriose fermentation kinetics in *Saccharomyces pastorianus* lager brewing yeast. *Front. Microbiol.* 8:1690.
89. Gibson BR, Storgårds E, Krogerus K, & Vidgren V (2013) Comparative physiology and fermentation performance of Saaz and Froberg lager yeast strains and the parental species *Saccharomyces eubayanus*. *Yeast* 30(7):255-266.
90. Vidgren V, Huuskonen A, Virtanen H, Ruohonen L, & Londesborough J (2009) Improved fermentation performance of a lager yeast after repair of its *AGT1* maltose and maltotriose transporter genes. *Appl. Environ. Microbiol.* 75(8):2333-2345.
91. Salema-Oom M, Pinto VV, Gonçalves P, & Spencer-Martins I (2005) Maltotriose utilization by industrial *Saccharomyces* strains: characterization of a new member of the  $\alpha$ -glucoside transporter family. *Appl. Environ. Microbiol.* 71(9):5044-5049.
92. Dietvorst J, Londesborough J, & Steensma H (2005) Maltotriose utilization in lager yeast strains: *MTT1* encodes a maltotriose transporter. *Yeast* 22(10):775-788.
93. Jansen ML, Daran-Lapujade P, de Winde JH, Piper MD, & Pronk JT (2004) Prolonged maltose-limited cultivation of *Saccharomyces cerevisiae* selects for cells with improved maltose affinity and hypersensitivity. *Appl. Environ. Microbiol.* 70(4):1956-1963.
94. Postma E, Verduyn C, Kuiper A, Scheffers WA, & Van Dijken JP (1990) Substrate-accelerated death of *Saccharomyces cerevisiae* CBS 8066 under maltose stress. *Yeast* 6(2):149-158.
95. Vidgren V (2010) Maltose and maltotriose transport into ale and lager brewer's yeast strains. *VTT publications.*
96. Steensels J, et al. (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol. Rev.* 38(5):947-995.
97. Bachmann H, Pronk JT, Kleerebezem M, & Teusink B (2015) Evolutionary engineering to enhance starter culture performance in food fermentations. *Curr. Opin. Biotechnol.* 32:1-7.
98. Gorter de Vries AR, et al. (2019) Laboratory evolution of a *Saccharomyces cerevisiae* x *S. eubayanus* hybrid under simulated lager-brewing conditions, *Frontiers in Genetics* 10:242.

## Chapter 9: *In vivo* recombination of *Saccharomyces eubayanus* maltose-transporter genes yields a chimeric transporter that enables maltotriose fermentation

Nick Brouwers <sup>#</sup>, Arthur R. Gorter de Vries <sup>#</sup>, Marcel van den Broek, Susan M. Weening, Tom D. Elink Schuurman, Niels G. A. Kuijpers, Jack T. Pronk and Jean-Marc G. Daran

<sup>#</sup> These authors contributed equally to this publication and should be considered co-first authors.

*Saccharomyces eubayanus* is the non-*S. cerevisiae* parent of the lager-brewing hybrid *S. pastorianus*. In contrast to most *S. cerevisiae* and Froberg-type *S. pastorianus* strains, *S. eubayanus* cannot utilize the  $\alpha$ -tri-glucoside maltotriose, a major carbohydrate in brewer's wort. In *Saccharomyces* yeasts, utilization of maltotriose is encoded by the subtelomeric *MAL* gene family, and requires transporters for maltotriose uptake. While *S. eubayanus* strain CBS 12357<sup>T</sup> harbors four *SeMALT* genes which enable uptake of the  $\alpha$ -di-glucoside maltose, it lacks maltotriose transporter genes. In *S. cerevisiae*, sequence identity indicates that maltotriose and maltose transporters likely evolved from a shared ancestral gene. To study the evolvability of maltotriose utilization in *S. eubayanus* CBS 12357<sup>T</sup>, maltotriose-assimilating mutants obtained after UV mutagenesis were subjected to laboratory evolution in carbon-limited chemostat cultures on maltotriose-enriched wort. An evolved strain showed improved maltose and maltotriose fermentation in 7-L fermenter experiments on industrial wort. Whole-genome sequencing revealed a novel mosaic *SeMALT413* gene, resulting from repeated gene introgressions by non-reciprocal translocation of at least three *SeMALT* genes. The predicted tertiary structure of *SeMalt413* was comparable to the original *SeMalT* transporters, but overexpression of *SeMALT413* sufficed to enable growth on maltotriose, indicating gene neofunctionalization had occurred. The mosaic structure of *SeMALT413* resembles the structure of *S. pastorianus* maltotriose-transporter gene *SpMTY1*, which has high sequence identity to alternately *S. cerevisiae* *MALx1*, *S. paradoxus* *MALx1* and *S. eubayanus* *SeMALT3*. Evolution of the maltotriose transporter landscape in hybrid *S. pastorianus* lager-brewing strains is therefore likely to have involved mechanisms similar to those observed in the present study.

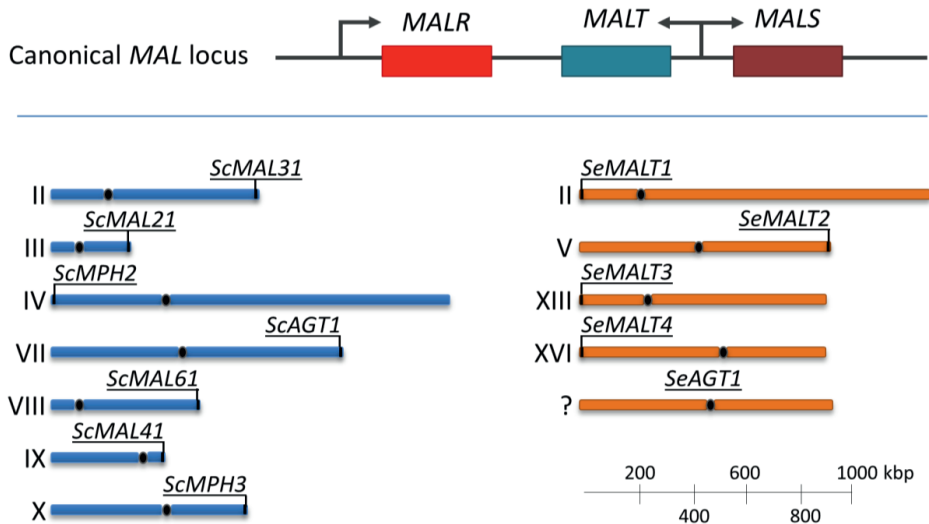
Essentially as published in PLoS Genetics 2019;15(4): e1007853

Supplementary materials available online

<https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007853>

## Introduction

*Saccharomyces eubayanus* was discovered in Patagonia and identified as the non-*S. cerevisiae* parental species of hybrid *S. pastorianus* lager-type beer brewing yeasts (1, 2). While *S. cerevisiae* is strongly associated with biotechnological processes, including dough leavening, beer brewing and wine fermentation (3), *S. eubayanus* has only been isolated from the wild (4-6). Beer brewing is performed with wort, a complex substrate containing a fermentable sugar mixture of 15 % of the monosaccharide glucose, 60 % of the  $\alpha$ -di-glucoside maltose and 25 % of the  $\alpha$ -tri-glucoside maltotriose (7). While many *S. cerevisiae* and *S. pastorianus* strains utilize all three sugars, *S. eubayanus* isolates do not utilize maltotriose (8-10). In *Saccharomyces*, the ability to utilize maltotriose requires its uptake into the cell and subsequent hydrolysis into glucose (11, 12). Maltose and maltotriose utilization are encoded by genes clustered in the *MAL* loci, which can be present on up to five different chromosomes (13). *MAL* loci typically harbor genes from up to three gene families (Fig 1): a *MALT* polysaccharide proton-symporter gene, a *MALS*  $\alpha$ -glucosidase gene which hydrolyses  $\alpha$ -oligo-glucosides into glucose, and a *MALR* regulator gene that induces the transcription of *MALT* and *MALS* genes in the presence of maltose (14). While *MALS* genes enable hydrolysis of both maltose and maltotriose, the *MALT* gene family comprises transporters with diverse substrate specificities (11, 15). In *S. cerevisiae*, most *MAL* loci harbor an *ScMALx1* transporter (Fig 1), which transports maltose and other disaccharides, such as turanose and sucrose (12, 16), but cannot import maltotriose (11). In contrast, the *MAL1* locus located on chromosome VII of *S. cerevisiae* contains *ScAGT1*, a transporter gene with only 57 % nucleotide identity with *ScMALx1* transporter genes. *ScAGT1* encodes a broad-substrate-specificity sugar-proton symporter that enables maltotriose uptake (11, 17, 18).



**Fig 1: Organization of maltose and maltotriose transporter genes in *S. cerevisiae* and *S. eubayanus*.** In *Saccharomyces* species, maltose and maltotriose utilization is encoded in the *MAL* genes, which are located in subtelomeric regions and comprise three types of genes: a *MALT*  $\alpha$ -oligo-glucoside proton-symporter gene, a *MALS*  $\alpha$ -glucosidase gene which hydrolyses  $\alpha$ -(di or tri)-glucosides into glucose, and a *MALR* regulator gene that induces the transcription of *MALT* and *MALS* genes in the presence of maltose. In canonical *MAL* loci, the *MALT* and *MALS* are expressed from a bi-directional *MALR*-dependent promoter sequence. The chromosomal location of known maltose and maltotriose transporter genes in *S. cerevisiae* and *S. eubayanus* is shown, although the presence of these genes varies among isolates. *ScMPH2* and *ScMPH3* encode  $\alpha$ -glucoside

permeases which do not enable efficient maltotriose uptake (11). *ScMAL31*, *ScMAL21*, *ScMAL61* and *ScMAL41* encode maltose transporters of the ScMalx1 family. *ScAGT1* encodes a maltotriose transporter. *SeMALT1*, *SeMALT2*, *SeMALT3* and *SeMALT4* encode maltose transporters with high sequence identity to the ScMalx1 family. *SeAGT1* is a maltotriose transporter which has recently been discovered in the north American *S. eubayanus* isolate yHRVM108 (65).

The *S. eubayanus* type strain CBS 12357<sup>T</sup> is able to utilize maltose, but cannot utilize maltotriose, suggesting that it expresses a functional maltase and a functional maltose transporter, but no maltotriose transporter (9). Indeed, the four *MAL* loci in CBS 12357<sup>T</sup> harbor a total of two *MALS* genes and four *MALT* genes with high homology to *ScMALx1*: *SeMALT1*, *SeMALT2*, *SeMALT3* and *SeMALT4* (Fig 1) (19). Deletion of these genes in *S. eubayanus* type strain CBS 12357<sup>T</sup> indicated that its growth on maltose relies on expression of *SeMALT2* and *SeMALT4* (9). *SeMALT1* and *SeMALT3* were found to be poorly expressed in the presence of maltose in this strain, supposedly due to incompleteness of the *MAL* loci which harbor them. However, no homolog of *ScAGT1* was found in the genome of CBS 12357<sup>T</sup>, and neither CBS 12357<sup>T</sup> nor its derivatives overexpressing *SeMALT* genes were able to utilize maltotriose (9).

The *MALT* transporter genes in *Saccharomyces* yeasts are localized to the subtelomeric regions (Fig 1) (9, 11, 12, 19-21), which are gene-poor and repeat-rich sequences adjacent to the telomeres (22-24). The presence of repeated sequences makes subtelomeric regions genetically unstable by promoting recombination (25, 26). As a result, subtelomeric gene families are hotspots of genetic diversity (27-30). In *S. cerevisiae*, subtelomeric gene families contain more genes than non-subtelomeric gene families, reflecting a higher incidence of gene duplications (27). As previously shown in *Candida albicans* submitted to long term laboratory evolution, the gene repertoire of the subtelomeric *TLO* family can be extensively altered due to ectopic recombinations between subtelomeric regions of different chromosomes, resulting in copy number expansion, in gene disappearance and in formation of new chimeric genes (31). Despite their common origin, genes within one family can have different functions, due to the accumulation of mutations (32, 33). *In silico* analysis of the sequences and functions of genes from the *MALT*, *MALS* and *MALR* gene families indicated functional diversification through gene duplication and mutation (27). Gene duplication is a critical step for the evolution of new gene functions (34, 35). Indeed, the presence of multiple gene copies can facilitate the emergence of advantageous mutations mainly by one of three mechanisms: (i) neofunctionalization, corresponding to the emergence of a novel function which was previously absent in the gene family (36), (ii) subfunctionalization, corresponding to the specialization of gene copies for part of the function of the parental gene (37) and (iii) altered expression due to gene dosage effects resulting from the increased copy number (38). While the different functions of *MALS* genes were assigned to subfunctionalization of the ancestral *MALS* gene (15), it is not known how the different functions of the maltotriose transporter gene *ScAGT1* and of other maltose transporter genes of the *MALT* family evolved from a common ancestor gene (27). In general, the emergence of a large array of gene functions was attributed to subfunctionalization and neofunctionalization (15, 27, 31, 39-42). However, current evidence for neofunctionalization within subtelomeric gene families is based on *a posteriori* analysis and rationalization of existing sequence diversity. While in some cases the genetic process leading to neofunctionalization could be reconstructed at the molecular level (42-44), the emergence of a completely new function within a subtelomeric gene family was never observed within the timespan of an experiment to the best of our knowledge. However, the genetic diversity within *Saccharomyces* *MALT* transporters suggests that evolution of *SeMalT* transporters could lead to the emergence of a maltotriose transporter by neofunctionalization (27).



Therefore, laboratory evolution may be sufficient to obtain maltotriose utilization in *S. eubayanus* strain CBS 12357<sup>T</sup>.

Laboratory evolution is a commonly-used method for obtaining desired properties by prolonged growth and selection under conditions favoring cells which develop the desired phenotype (45, 46). Similarly as in Darwinian natural evolution, the conditions under which laboratory evolution is conducted shape the phenotypes acquired by evolved progeny following the process of survival of the fittest (47). In *Saccharomyces* yeasts, selectable properties include complex and diverse phenotypes such as high temperature tolerance, efficient nutrient utilization and inhibitor tolerance (48-51). Laboratory evolution was successfully applied to improve sugar utilization for arabinose, galactose, glucose and xylose (49, 52-54). In *S. pastorianus*, improved maltotriose uptake was successfully selected for in a prolonged chemostat cultivation on medium enriched with maltotriose (55). Theoretically, laboratory evolution under similar conditions could select *S. eubayanus* mutants which develop the ability to utilize maltotriose.

In this study, we submitted *S. eubayanus* strain CBS 12357<sup>T</sup> to UV-mutagenesis and laboratory evolution in order to obtain maltotriose utilization under beer brewing conditions. While obtaining a non-GMO maltotriose-consuming *S. eubayanus* strain was a goal in itself for industrial beer brewing, we were particularly interested in the possible genetic mechanisms leading to the emergence of maltotriose utilization. Indeed, we hypothesized that the genetic plasticity and functional redundancy of the four subtelomeric *SeMALT* genes of CBS 12357<sup>T</sup> could facilitate the emergence of maltotriose transport by neofunctionalization. The evolution process leading to maltotriose utilization in a strain with only maltose transporters, such as CBS 12357<sup>T</sup>, may provide insight in the emergence of maltotriose utilization in general.

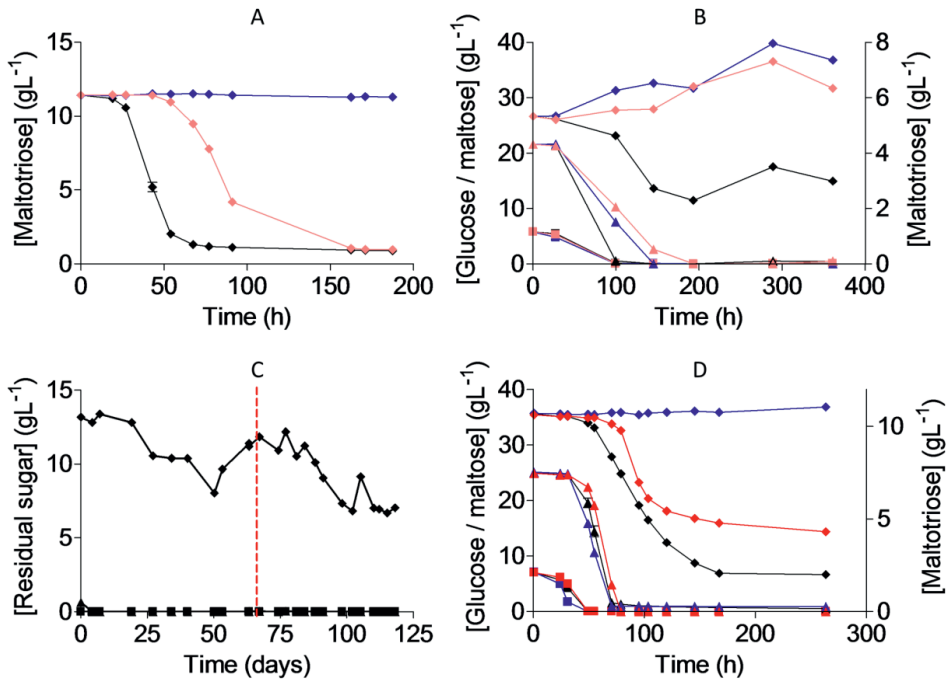
## Results

### Mutagenesis and evolution enables *S. eubayanus* to utilize maltotriose

To obtain maltotriose-consuming mutants, CBS 12357<sup>T</sup> was grown on synthetic medium containing 20 g L<sup>-1</sup> glucose (SMG) until stationary phase, and approximately 10<sup>8</sup> cells were used to inoculate synthetic medium containing 20 g L<sup>-1</sup> maltotriose (SMMt) as sole carbon source. After incubation at 20 °C for three months, neither growth nor maltotriose utilization was observed.

Exposure to UV radiation can cause DNA damage, resulting in the emergence of cells with diverse mutations due to error-prone repair. Therefore, UV-mutagenesis was applied to increase the likelihood of obtaining a mutation enabling maltotriose utilization. To this end, CBS 12357<sup>T</sup> was grown in SMG medium, sporulated, submitted to mild UV-mutagenesis (46 % survival rate) and approximately 10<sup>8</sup> cells of the mutagenized population were used to inoculate SMMt containing 20 g L<sup>-1</sup> maltotriose. After two weeks at 20 °C, growth was observed and, after 3 weeks, the maltotriose concentration had decreased to 10.5 g L<sup>-1</sup>. After two subsequent transfers in fresh SMMt, 96 single cells were sorted into a 96 well plate containing YPD medium by fluorescence-activated cell sorting (FACS). The resulting single-cell cultures were transferred to a 96 well plate containing SMMt, in which growth was monitored by OD<sub>660</sub> measurements. The seven single-cell isolates with the highest final OD<sub>660</sub> were selected and named IMS0637-IMS0643. To characterize growth on maltotriose, the strain CBS 12357<sup>T</sup>, the single-cell isolates IMS0637-IMS0643 and the maltotriose-consuming *S. pastorianus* strain CBS 1483 were grown in shake flasks on SMMt (Fig 2A and Supplementary Fig S1). After 187 h, *S. eubayanus* CBS 12357<sup>T</sup> did not show any maltotriose

consumption. Conversely, isolates IMS0637-IMS0643, all showed over 50 % maltotriose consumption after 91 h (as compared to 43 h for CBS 1483). Upon reaching stationary phase, isolates IMS0637-IMS0643 had consumed  $93 \pm 2\%$  of the initial maltotriose concentration, which was similar to the 92 % conversion reached by *S. pastorianus* CBS 1483. While these results indicated that the single cell isolates IMS0637-IMS0643 utilized maltotriose in synthetic medium, they did not consume maltotriose in presence of glucose and maltose after 145 h of incubation (Fig 2B). Under the same conditions, *S. pastorianus* CBS 1483 consumed 50 % of the maltotriose after 145 h (Fig 2B).



**Fig 2: Mutagenesis and evolution to obtain maltotriose consuming *S. eubayanus*.** (A) Characterization of *S. pastorianus* CBS 1483 (black), *S. eubayanus* CBS 12357 (blue) and IMS0637 (light red) on SMMt at 20 °C. The data for IMS0637 is representative for the other mutants IMS0638-IMS0643 (Supplementary Fig S1). The average concentration of maltotriose (◆) and average deviation were determined from two replicates (Supplementary data file 2). (B) Characterization of *S. pastorianus* CBS 1483 (black), *S. eubayanus* CBS 12357 (blue) and IMS0637 (light red) on wort at 20 °C. The concentrations of (■) glucose, (▲) maltose and (◆) maltotriose were measured from single biological measurements (Supplementary data file 3). (C) Residual maltotriose concentration in the outflow during laboratory evolution of strains IMS0637-IMS0643 in an anaerobic chemostat at 20 °C on maltotriose enriched wort. The concentrations of (■) glucose, (▲) maltose and (◆) maltotriose were measured by HPLC. The chemostat was restarted after a technical failure (red dotted line, Supplementary data file 4). (D) Characterization of *S. pastorianus* CBS 1483 (black), *S. eubayanus* CBS 12357 (blue) and IMS0750 (red) on wort at 12 °C in 250 mL micro-aerobic Neubor infusion bottles. The average concentration and standard deviation of (■) glucose, (▲) maltose and (◆) maltotriose were determined from three biological replicates. The data for IMS0751 and IMS0752 are shown in Supplementary data file 5 and Fig S2.

Nutrient-limited growth confers a selective advantage to spontaneous mutants with a higher nutrient affinity (45, 55). Therefore, to improve maltotriose utilization under industrially relevant conditions, the pooled isolates IMS0637-IMS0643 were subjected to laboratory evolution in a chemostat culture on modified brewer's wort. To avoid the presence of residual maltose, which would prevent

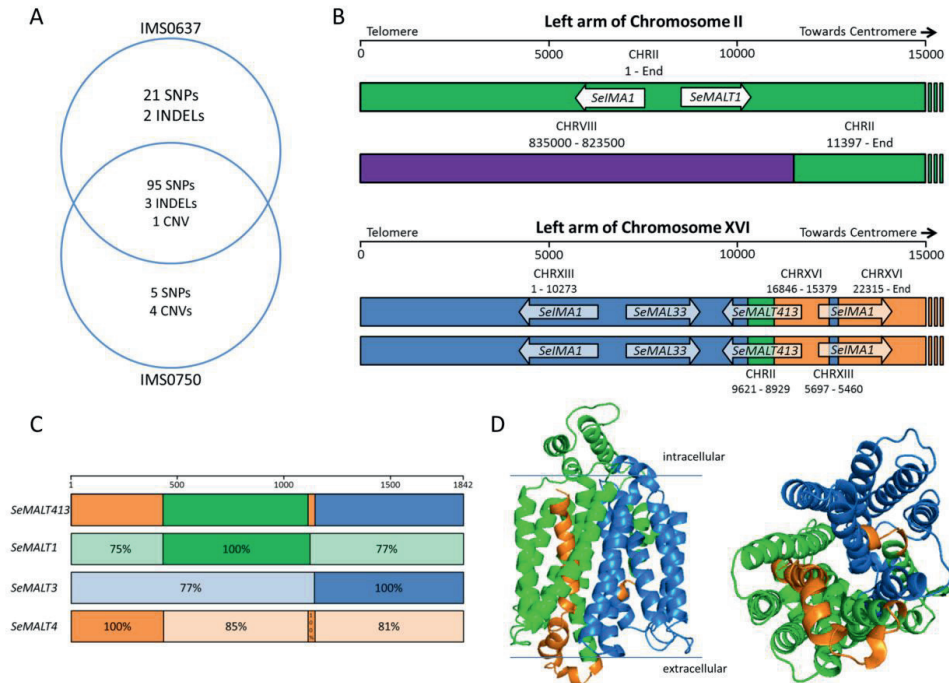
selection for maltotriose utilization, brewer's wort was diluted 6-fold. To strengthen the selective advantage for maltotriose-consuming cells, the diluted wort was complemented with 10 g L<sup>-1</sup> maltotriose, yielding concentrations of 2 g L<sup>-1</sup> glucose, 15 g L<sup>-1</sup> maltose and 15 g L<sup>-1</sup> maltotriose in the medium feed. To prevent growth limitation due to the availability of limited oxygen or nitrogen, the medium was supplemented with 10 mg L<sup>-1</sup> ergosterol, 420 mg L<sup>-1</sup> Tween 80 and 5 g L<sup>-1</sup> ammonium sulfate (56). During the batch cultivation phase that preceded continuous chemostat cultivation, glucose and maltose were completely consumed, leaving maltotriose as the only carbon source. After initiation of continuous cultivation at a dilution rate of 0.03 h<sup>-1</sup>, the medium outflow initially contained 13.2 g L<sup>-1</sup> of maltotriose. After 121 days of chemostat cultivation, the maltotriose concentration had progressively decreased to 7.0 g L<sup>-1</sup> (Fig 2C). At that point, which corresponded to *ca.* 125 generations, 10 single colonies were isolated from the culture on SMMt agar plates and incubated at 20 °C. Three single-cell lines were named IMS0750, IMS0751 and IMS0752 and selected for further characterization in micro-aerobic cultures, grown at 12 °C on 3-fold diluted wort, along with *S. eubayanus* CBS 12357<sup>T</sup> and *S. pastorianus* CBS 1483 (Fig 2D). In these cultures, strains CBS 12357<sup>T</sup> and IMS0751 only consumed glucose and maltose, while *S. pastorianus* CBS 1483, as well as the evolved isolates IMS0750 and IMS0752, also consumed maltotriose (Supplementary Fig S2). After 263 h, maltotriose concentrations in cultures of strains IMS0750 and IMS0752 had decreased from 20 to 4.3 g L<sup>-1</sup> maltotriose as compared to 2.0 g L<sup>-1</sup> in cultures of strain CBS 1483.

#### **Whole genome sequencing reveals a new recombined chimeric *SeMALT* gene**

We sequenced the genomes of the *S. eubayanus* strain CBS 12357<sup>T</sup>, of the UV-mutagenized isolates IMS0637-IMS0643 and of the strains isolated after subsequent chemostat evolution IMS0750-IMS0752 using paired-end Illumina sequencing. Sequencing data were mapped to a chromosome-level assembly of strain CBS 12357<sup>T</sup> (9) to identify SNPs, INDELS and copy number changes. The genomes of the UV-mutants IMS0637, IMS0640, IMS0641 and IMS0642 shared a set of 116 SNPs, 5 INDELS and 1 copy number variation (Supplementary data file 1). In addition to these shared mutations, isolates IMS0638, IMS0639 and IMS0643 carried three identical SNPs. Overall, 97 % of SNPs and INDELS of IMS0637-IMS0643 were heterozygous, indicating that the haploid spores of CBS 12357<sup>T</sup> diploidized by mating after mutagenesis (Supplementary data file 1). Of the mutations present in all isolates, 34 SNPs and all 5 INDELS affected intergenic regions, 30 SNPs were synonymous, 48 SNPs resulted in amino acid substitutions and 4 SNPs resulted in a premature stop codon (Supplementary data file 1). None of the 52 non-synonymous SNPs affected genes previously linked to maltotriose utilization. The only copy number variation concerned a duplication of the right subtelomeric region of CHRVIII. Read mate-pairing indicated that the duplicated region was attached to the left arm of CHRII, causing the replacement of left subtelomeric region of CHRII by a non-reciprocal translocation. The recombination resulted in loss of one of the *SeMALT1* allele, which is not expressed in CBS 12357<sup>T</sup> (9).

Since the ability to utilize maltotriose in wort emerged only after laboratory evolution during chemostat cultivation, mutations present in the chemostat-evolved strains IMS0750 and IMS0752 were studied in more detail. With the exception of one synonymous SNP, IMS0750 and IMS0752 were identical and shared 100 SNPs, 3 INDELS and 5 copy number changes (Supplementary data file 1). The non-maltotriose utilizing strain IMS0751 shared only 63 SNPs and 3 INDELS with IMS0750 and IMS0752, of which 98 % were homozygous, indicating a recent loss of heterozygosity event affected its entire genome. Of the mutations in maltotriose-utilizing strains IMS0750 and IMS0752, only 5 SNPs and 4 copy number changes were absent in IMS0637-IMS0643, and could therefore

explain the ability to utilize maltotriose in wort (Fig 3A). The 5 SNPs consisted of two intergenic SNPs and three non-synonymous SNPs in genes with no link to maltotriose. However, the changes in copy number affected several regions harboring *SeMALT* genes: a duplication of 550 bp of CHR II including *SeMALT1* (coordinates 8,950 to 9,500), a duplication of the left arm of CHR XIII including *SeMALT3* (coordinates 1-10,275), loss of the left arm of CHR XVI (coordinates 1-15,350), and loss of 5.5 kb of CHR XVI including *SeMALT4* (coordinates 16,850-22,300). Analysis of read mate pairing indicated that the copy number variation resulted from a complex set of recombinations between chromosomes II, XIII and XVI.



**Fig 3: Identification of mutations in the mutagenized strain IMS0637 and the evolved strain IMS0750.** (A) Venn diagram of the mutations found in UV-mutagenized IMS0637 and evolved IMS0750 relative to wildtype CBS 12357<sup>T</sup>. Single nucleotide polymorphisms (SNPs), small insertions and deletions (INDELS) and copy number variation (CNV) are indicated as detected by Pilon. (B) Recombined chromosome structures in IMS0637 and IMS0750 as detected by whole genome sequencing using MinION nanopore technology and *de novo* genome assembly. The first 15,000 nucleotides of the left arm of CHR II and CHR XVI are represented schematically. The origin of the sequence is indicated in green for CHR II, purple for CHR VIII, blue for CHR XIII and orange for CHR XVI. The ORFs of *SeMALT* transporter genes, *SeIMA* isomaltase genes and the MALR-type regulator *SeMAL33* which were affected by the recombinations are indicated by arrows. While the recombination of CHR II and CHR VIII was present in IMS0637 and IMS0750, the recombination of both copies of CHR XVI was found only in IMS0750 but not in IMS0637. The recombination on CHR XVI created the chimeric *SeMALT413* transporter gene. (C) Overview of the sequence identity of the 1,842 nucleotides of *SeMALT413* relative to *SeMALT1*, *SeMALT3* and *SeMALT4*. The open reading frames of the genes were aligned (Supplementary Fig S3) and regions with 100 % sequence identity were identified. For regions in which the sequence identity was lower than 100 %, the actual sequence identity is indicated for each *SeMALT* gene. The origin of the sequence is indicated in green for CHR II, red for CHR VIII, blue for CHR XIII and orange for CHR XVI. (D) Prediction of the protein structure of *SeMalt13* with on the left side a transmembrane view and on the right a transport channel view. Domains originated from *S. eubayanus* *SeMalt* transporters are indicated by the colors orange (*SeMalt4* chromosome XVI), green (*SeMalt1* chromosome II) and blue (*SeMalt3* chromosome XIII).

The high degree of sequence identity of the affected *MAL* loci and their localization in the subtelomeric regions made exact reconstruction of the mutations difficult. Therefore, IMS0637 and IMS0750 were sequenced using long-read sequencing on ONT's MinION platform, and a *de novo* genome assembly was made for each strain. Comparison of the resulting assemblies to the chromosome-level assembly of CBS 12357<sup>T</sup> indicated that two recombinations had occurred. Both in IMS0637 and IMS0750, an additional copy of the terminal 11.5 kb of the right arm of chromosome VIII had replaced the terminal 11.4 kb of one of the two copies of the left arm of chromosome II (Fig 3B). This recombination was consistent with the copy number changes of the affected regions in IMS0637-IMS0643, IMS0750 and IMS0752 and resulted in the loss of one copy of the *MAL* locus harboring *SeMALT1*. In addition, the genome assembly of IMS0750 indicated the replacement of both copies of the first 22.3 kb of CHR XVI by complexly rearranged sequences from CHR II, CHR XVIII and CHR XVI. The recombinant region comprised the terminal 10,273 nucleotides of the left arm of CHR III, followed by 693 nucleotides from CHR II, 1,468 nucleotides from CHR XVI and 237 nucleotides from CHR XIII (Fig 3B). The recombinations were non reciprocal, as the regions present on the recombinant chromosome showed increased sequencing coverage while surrounding regions were unaltered. This recombination resulted in the loss of the canonical *MAL* locus harboring *SeMALT4* on chromosome XVI. However, the recombinant sequence contained a chimeric open reading frame consisting of the 5' part of *SeMALT4* from CHR XVI, the middle of *SeMALT1* from CHR II and the 3' part of *SeMALT3* from CHR XIII (Fig 3C, Supplementary Fig S3). To verify this recombination, the ORF was PCR amplified using primers binding on the promoter of *SeMALT4* and the terminator of *SeMALT3*, yielding a fragment for strain IMS0750, but not for CBS 12357<sup>T</sup> and IMS0637. Sanger sequencing of the fragment amplified from strain IMS0750 confirmed the chimeric organization of the ORF, which we named *SeMALT413*. The sequence of *SeMALT413* showed 100 % identity to *SeMALT4* for nucleotides 1-434 and 1113-1145, 100 % identity to *SeMALT1* for nucleotides 430-1122 and 100 % identity to *SeMALT3* for nucleotides 1141-1842 (Fig 3C). Nucleotides 1123-1140, which showed only 72 % identity with *SeMALT1* and 61 % identity with *SeMALT3*, were found to represent an additional introgression (Fig 3B). While the first 434 nucleotides can be unequivocally attributed to *SeMALT4* due to a nucleotide difference with *SeMALT2*, the nucleotides 1123-1140 are identical in *SeMALT2* and *SeMALT4*. Therefore, this part of the sequence of *SeMALT413* might have come from *SeMALT2* on CHR V or from *SeMALT4* on CHR XVI. Overall, *SeMALT413* showed a sequence identity of only 85 to 87 % with the original *SeMALT* genes, with the corresponding protein sequence exhibiting between 52 and 88 % similarity. We therefore hypothesized that the recombinant *SeMalT413* transporter might have an altered substrate specificity and thereby enable maltotriose utilization.

The tertiary protein structure of the chimeric *SeMALT413* gene was predicted with SWISS-MODEL (<https://swissmodel.expasy.org/>), based on structural homology with the *Escherichia coli* xylose-proton symporter XylE (57), which has previously been used as a reference to model the structure of the maltotriose transporter *ScAgt1* (58). Similarly to the maltose transporters in *Saccharomyces*, XylE is a proton symporter belonging to the major facilitator superfamily with a transmembrane domain composed of 12  $\alpha$ -helices (Supplementary Fig S4). The same structure was predicted for *SeMalT413*, with 1  $\alpha$ -helix formed exclusively by residues from *SeMalT4*, 4  $\alpha$ -helices formed exclusively by residues from *SeMalT1* and 5  $\alpha$ -helices formed exclusively by residues from *SeMalT3* (Fig 3D). The remaining two  $\alpha$ -helices were composed of residues from more than one transporter. Since the first 100 amino acids were excluded from the model due to absence of similar residues in the xylose symporter reference model, the structure prediction underestimated the contribution of *SeMalT4*.

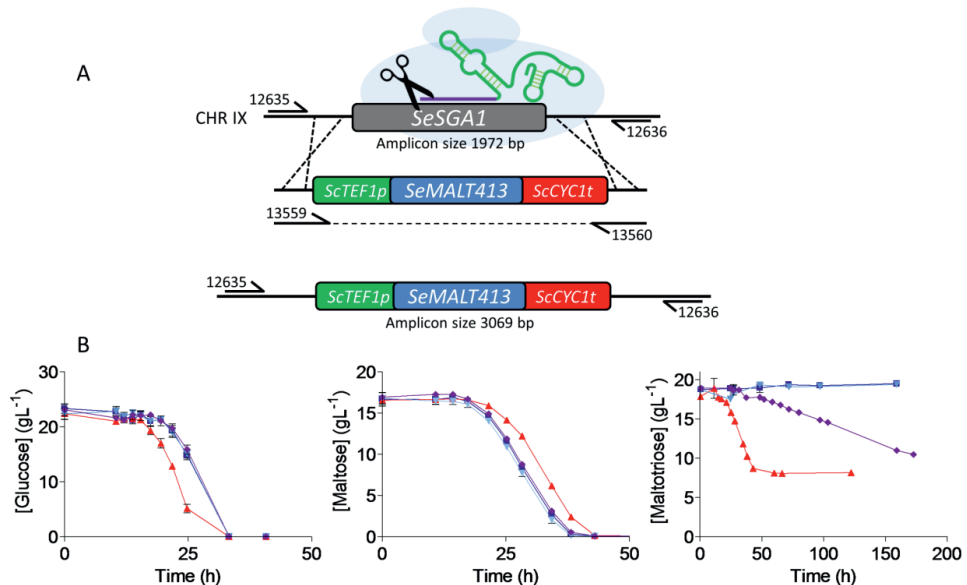
The three-dimensional arrangement of the  $\alpha$ -helices of SeMalT413 was almost identical to SeMalT1, SeMalT3 and SeMalT4, indicating that it retained the general structure of a functional maltose transporter (Supplementary Fig S5).

#### **Introduction of the *SeMALT413* gene in wildtype CBS 12357<sup>T</sup> enables maltotriose utilization**

The small structural differences identified between SeMalT413 and the wild-type *S. eubayanus* SeMalT transporters could not be used to predict the ability of SeMalT413 to transport maltotriose (58). Therefore, to investigate its role in maltotriose transport, *SeMALT413* and, as a control, *SeMALT2* were overexpressed in the wild-type strain *S. eubayanus* CBS 12357<sup>T</sup> (Fig 4A and Supplementary Fig S6). Consistently with previous gene editing in CBS 12357<sup>T</sup> (9), the expression cassettes were inserted at the *SeSGA1* locus, encoding an intracellular sporulation-specific glucoamylase which is not expressed during vegetative growth (59, 60). Growth of the resulting *S. eubayanus* strains IMX1941 (*SeSGA1* $\Delta$ ::*ScTEF1*<sub>pr</sub>-*SeMALT2*-*ScCYC1*<sub>ter</sub>) and IMX1942 (*SeSGA1* $\Delta$ ::*ScTEF1*<sub>pr</sub>-*SeMALT413*-*ScCYC1*<sub>ter</sub>), as well as the wild-type strain CBS 12357<sup>T</sup> and the evolved isolate IMS0750 was tested on SM supplemented with different carbon sources (Supplementary Fig S7). On glucose, strains IMX1941 and IMX1942 exhibited the same specific growth rate of  $0.25 \pm 0.01 \text{ h}^{-1}$  as CBS 12357<sup>T</sup>, while IMS0750 grew faster with a growth rate of  $0.28 \pm 0.01 \text{ h}^{-1}$ . Glucose was completely consumed after 33 h (Fig 4B). On maltose, the specific growth rates of CBS 12357<sup>T</sup>, IMX1941, IMX1942 and IMS0750 ranged between 0.17 and 0.19  $\text{h}^{-1}$  and did not differ significantly. Maltose was completely consumed after 43 h (Fig 4C). On maltotriose, only the evolved mutant IMS0750 and reverse engineered strain IMX1942 (*ScTEF1*<sub>pr</sub>-*SeMALT413*-*ScCYC1*<sub>ter</sub>) showed growth. IMS0750 grew with a specific growth rate of  $0.19 \pm 0.01 \text{ h}^{-1}$  and consumed 55 % of maltotriose within 172 h. Over the same period, IMX1942 grew at  $0.03 \pm 0.00 \text{ h}^{-1}$  and consumed 45 % of the maltotriose after 172 h (Fig 4D), demonstrating the capacity of *SeMALT413* to transport maltotriose.

#### **The *SpMTY1* maltotriose transporter gene displays a similar chimeric structure as *SeMALT413***

The emergence of the maltotriose transporter SeMalT413 by recombination between different *MALT* genes during laboratory evolution demonstrates that *MALT* gene neofunctionalization can contribute to the emergence of maltotriose utilization. To investigate if such neofunctionalization could have played a role in the emergence of maltotriose transporter genes in *Saccharomyces* yeasts, we analyzed the sequences of existing maltotriose transporter genes in *S. cerevisiae* and *S. pastorianus* genomes. In *S. cerevisiae* strains, maltotriose transport is encoded by the *ScAGT1* gene (11, 17, 18). However, *ScAGT1* is truncated and non-functional in *S. pastorianus* (61) and *ScAGT1* was lost during laboratory evolution of *S. cerevisiae* x *S. eubayanus* hybrids under lager brewing conditions (62). Instead, maltotriose utilization has been attributed to two *S. pastorianus*-specific genes: *LgAGT1* and *SpMTY1*. The maltotriose transporter gene *LgAGT1* was identified on *S. eubayanus* chromosome XV of *S. pastorianus* and shares 85 % sequence identity with *ScAGT1* (63, 64). Although it is absent in CBS 12357<sup>T</sup> (19), *LgAGT1* was found to enable maltotriose transport in the north-American *S. eubayanus* isolate yHRVM108 (19, 65). The *SpMTY1* gene, also referred to as *MTT1*, was found in the *MAL1* locus of *S. pastorianus* (20, 21). In addition to *S. cerevisiae* chromosome VII, *SpMTY1* was also found on *S. eubayanus* chromosome VII of *S. pastorianus*, of which the right arm originates from *S. cerevisiae* due to a recombination (63). *SpMTY1* shows 90 % sequence identity with *ScMALx1* genes (20, 21), but also displays segmental sequence identity with *SeMALT* genes (66, 67).



**Fig 4: Reverse engineering of *SeMALT413* in CBS 12357<sup>T</sup> and characterization of transporter functionality in SM.** (A) Representation of the CRISPR-Cas9 gRNA complex (after self-cleavage of the 5' hammerhead ribozyme and a 3' hepatitis- $\delta$  virus ribozyme from the expressed gRNA) bound to the *SeSGA1* locus in CBS 12357<sup>T</sup>. Repair fragment with transporter cassette *ScTEF1p-ScMALT413-ScCYC1t* was amplified from pUD814(*SeMALT413*) with primers 13559/13560 and contains overhangs with the *SeSGA1* locus for recombination. *SeSGA1* was replaced by the *ScTEF1p-ScMALT413-ScCYC1t* cassette. Correct transformants were checked using primers 12635/12636 upstream and downstream of the *SeSGA1* locus (Supplementary Fig S6). Strains were validated using Sanger sequencing. (B) Characterization of (■) CBS 12357, (▲) IMS0750, (▼) IMX1941, (◆) IMX1942 on SM glucose, maltose and maltotriose. Strains were cultivated at 20 °C and culture supernatant was measured by HPLC. Data represent average and standard deviation of three biological replicates (Supplementary data file 6).

The relatively low homology of *ScAGT1* and *LgAGT1* genes indicates that they are less related to maltose transporter genes such as *ScMALx1* and *SeMALT* than *SpMTY1*. Moreover, their sequence identity to maltose transporters from the *MALT* family such as *ScMAL31* is roughly homogenous over their coding region. Therefore there is no evidence that they resulted from recombinations between other *MALT* genes. In contrast, the identity of some segments of *SpMTY1* relative to *ScMAL31* deviates strongly from the average identity of 89 % (20). Indeed, sequence identity with *ScMAL31* of *S. cerevisiae* S288C (68) is above 98 % for nucleotides 1-439, 627-776, 796-845, 860-968 and 1,640-1,844, while it is only 79 % for nucleotides 440-626, 65 % for nucleotides 777-795, 50 % for nucleotides 846-859 and 82 % for nucleotides 969-1,639 (Supplementary Fig S8). Alignment of the sequences of *S. eubayanus* CBS 12357<sup>T</sup> *SeMALT* genes (9) to *SpMTY1* showed high sequence identity with *SeMALT3* across several regions that showed significant divergence from the corresponding *ScMAL31* sequences: 91 % identity for nucleotides 478-533, 94 % identity for nucleotides 577-626 and 94 % identity for nucleotides 778-794 (Supplementary Fig S8). These observations would indicate that the evolution of *SpMTY1* might have involved introgression events similar to those responsible for the *SeMALT413* neofunctionalization described in the present study. However, introgressions from *SeMALT* genes cannot explain the entire *SpMTY1* gene structure. Its evolution may therefore

have involved multiple introgressions, similarly as for *SeMALT413*. While most regions with low identity to *ScMAL31* and *SeMALT3* were too short to identify their provenance, the sequence corresponding to the 969<sup>th</sup> to 1,639<sup>th</sup> nucleotide of *SpMTY1* could be blasted on NCBI. In the S288C genome, *ScMAL31* was the closest hit with 82 % identity. However, when blasting the sequence against the full repository excluding *S. pastorianus* genomes, the closest hit was the orthologue of *ScMAL31* on chromosome VII of *S. paradoxus* strain YPS138. In addition to an 89 % identity to nucleotides 969-1,639 of *SpMTY1*, *SparMAL31* had an identity of 94 % for nucleotides 544-575 and of 93 % for nucleotides 846-859 (Supplementary Fig S8). Therefore, the sequence of *SpMTY1* may have resulted from recombination between different *MALT* genes, involving *ScMALx1* and other *MALT* genes such as *SeMALT3* and *SparMAL31*. While the chimeric *SeMALT413* ORF can be fully explained by recombination between *SeMALT* genes, *SpMTY1* probably accumulated additional mutations during its evolution.

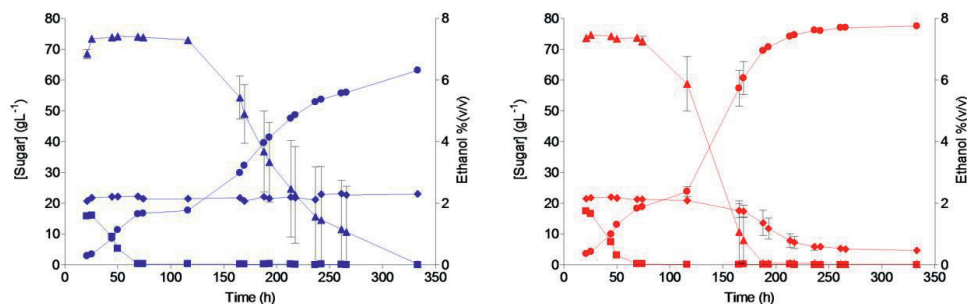
#### **Applicability of a maltotriose-consuming *S. eubayanus* strain for lager beer brewing**

*S. eubayanus* strains are currently used for industrial lager beer brewing (9). To test the evolved strain IMS0750 under laboratory-scale brewing conditions, its performance was compared with that of its parental strain CBS 12357<sup>T</sup> in 7-L cultures grown on high-gravity (16.6 ° Plato) wort (Fig 5). After 333 h, IMS0750 had completely consumed all glucose and maltose, and the concentration of maltotriose had dropped from 19.3 to 4.7 g L<sup>-1</sup> (Fig 5). In contrast, CBS 12357<sup>T</sup> did not utilize any maltotriose. In addition to its improved maltotriose utilization, IMS0750 also showed improved maltose consumption: maltose was completely consumed within 200 h, while complete maltose consumption by strain CBS 12357<sup>T</sup> took ca. 330 h (Fig 5). Consistent with its improved sugar utilization, the final ethanol concentration in cultures of strain IMS0750 was 18.5 % higher than in corresponding cultures of strain CBS 12357<sup>T</sup> (Fig 5). Brewing-related characteristics of IMS0750 were further explored by analyzing production of aroma-defining esters, higher alcohols and diacetyl. Final concentrations of esters and higher alcohols were not significantly different in cultures of the two strains, with the exception of isoamylacetate, which showed a 240 % higher concentration in strain IMS0750 (Table 1). In addition, while the concentration of the off-flavour diacetyl remained above its taste threshold of 25 µg L<sup>-1</sup> after 333h for CBS 12357<sup>T</sup>, it dropped below 10 µg L<sup>-1</sup> for IMS0750 (Table 1).

**Table 1: Concentrations of alcohols, esters and diacetyl after fermentation of wort with a gravity of 16.6 °P by *S. eubayanus* strains CBS 12357<sup>T</sup> and IMS0750.** The data correspond to the last time point (330 h) of the fermentations shown in Fig 5. The average and average deviation of duplicate fermentations are shown for each strain.

Compound	Unit	CBS 12357 <sup>T</sup>	IMS0750
Methanol	mg L <sup>-1</sup>	3.3 ± 0.3	3.7 ± 0.3
Propanol	mg L <sup>-1</sup>	23.7 ± 2.1	24.1 ± 0.9
Isobutanol	mg L <sup>-1</sup>	48.5 ± 2.4	42.9 ± 7.2
Amyl alcohol	mg L <sup>-1</sup>	138.5 ± 9.0	155.9 ± 6.4
Diacetyl	µg L <sup>-1</sup>	43.8 ± 22.9	7.5 ± 0.2
Ethylacetate	mg L <sup>-1</sup>	24.5 ± 5.5	26.1 ± 0.8
Isoamylacetate	mg L <sup>-1</sup>	1.4 ± 0.6	3.1 ± 0.3





**Fig 5: Extracellular metabolite profiles of *S. eubayanus* strains CBS 12357<sup>T</sup> and IMS0750 in high-gravity wort at 7-L pilot scale.** Fermentations were performed on wort with a gravity of 16.6 °Plato. The average concentrations of glucose (■), maltose (▲), maltotriose (◆) and ethanol (●) are shown for duplicate fermentations of CBS 12357<sup>T</sup> (blue) and IMS0750 (red). The average deviations are indicated (Supplementary data file 7).

## Discussion

UV mutagenesis and subsequent laboratory evolution yielded mutants which were able to utilize maltotriose in synthetic medium and in brewer's wort. In the resulting isolates IMS0750 and IMS0752, several recombinations affecting subtelomeric regions were identified. All four maltose transporter genes in *S. eubayanus* CBS 12357<sup>T</sup> are localized in subtelomeric *MAL* loci: *SeMALT1* on chromosome II, *SeMALT2* on chromosome V, *SeMALT3* on chromosome XIII and *SeMALT4* on chromosome XVI (9, 19). In the evolved strain IMS0750, a complex recombination between the subtelomeric regions of chromosomes II, XIII and XVI involved at least three of these *MAL* loci. Long-read nanopore sequencing enabled complete reconstruction of the recombined left arm of chromosome XVI, revealing recombinations between the ORFs of at least *SeMALT1*, *SeMALT3* and *SeMALT4*. These recombinations occurred within the open reading frame of *SeMALT4* and the newly-formed chimeric ORF *SeMALT413* encoded a full length protein with a structure comparable to that of *SeMalT* transporters. In contrast to the original *SeMALT* genes, overexpression of *SeMALT413* enabled growth on maltotriose, indicating that *SeMalT413* acquired the ability to import maltotriose.

The predicted structure of *SeMalT413* was highly comparable to the structure of other transporters from the major facilitator superfamily (69) and to the structure of *SeMalT1*, *SeMalT3* and *SeMalT4*. While nothing is known about the amino acid residues responsible for substrate specificity in *SeMalT* transporters, the threonine and serine residues at the 505<sup>th</sup> and 557<sup>th</sup> position respectively of *ScAgt1* were identified as critical for maltotriose transport (70). In *SeMalT413*, the corresponding amino acids originate from *SeMALT3*. However, since *SeMALT3* itself is unable to utilize maltotriose, the ability of *SeMalT413* to transport maltotriose likely depends on the interaction of residues from the different parental transporters, rather than from the residues of one of the transporters. Interestingly, CBS 12357<sup>T</sup> was recently evolved for maltotriose utilization in another study, resulting in a chimeric *SeMALT434* transporter which enabled maltotriose uptake (65). In this study, a 230-bp introgression of *SeMALT3* into the ORF of *SeMALT4* was found, including the 505<sup>th</sup> and 557<sup>th</sup> residues. While the shorter  $\alpha$ -helices of *SeMALT434* could lead to broader substrate-specificity by increasing structural flexibility, the length of these helices was not affected in *SeMALT413*. As a result, we hypothesize that the acquired maltotriose utilization does not depend solely on specific residues, but rather on the interaction of the residues from the different parental transporters, either by increasing structural flexibility, or by the properties of several critical residues from different

$\alpha$ -helices. The 230-bp from *SeMALT3* which were present in *SeMALT413* and in *SeMALT434* may be of particular importance. However, the specific combination of sequences from *SeMALT4*, *SeMALT1* and *SeMALT3* in *SeMALT413* may further contribute to the maltotriose specificity.

Recombinations are an important driver of evolution, as illustrated by the emergence of aerobic growth on citrate during laboratory evolution of *Escherichia coli* (71). Indeed, a tandem repeat of the citrate/succinate antiporter *citT* placed under the constitutive *rna* promoter enabled aerobic growth on citrate. Moreover, the emergence of a new ORF by recombination has been observed previously between the *TLO* genes of *C. albicans*, although it was not associated with a new gene function (31). In contrast, the emergence of *SeMALT413* is an example of gene neofunctionalization, which occurred by recombination within genes of the subtelomeric *MALT* family. Neofunctionalization by *in vivo* formation of chimeric sequences is reminiscent of the mechanism used by the pathogen *Trypanosoma brucei* to evade its host's immune system (72). *T. brucei* expresses a single variant surface glycoprotein (VSG) gene from a subtelomeric location and its genome contains many VSG pseudogenes (73). Due to 70 bp repetitive elements, the actively expressed VSG gene can be altered by gene conversion from pseudogenes, resulting in a chimeric VSG gene (74-76). While antigen switching may not qualify as neofunctionalization, it demonstrates the ability of recombinations to diversify gene functions by creating chimeric ORFs. This ability has also been exploited for *in vitro* protein engineering, a strategy known as gene shuffling or gene fusion (77, 78). Gene shuffling involves randomized assembly of diverse DNA sequences into chimeric genes, followed by screening for novel or improved functions. Analogously to *in vitro* gene shuffling, the complex protein remodeling caused by *in vivo* formation of chimeric sequences may be particularly potent for protein neofunctionalization (79). The demonstration of neofunctionalization of a sugar transporter in *S. eubayanus* by *in vivo* gene shuffling supports the notion that gene fusion is an essential driver of evolution by accelerating the emergence of new enzymatic functions (80). Moreover, analysis of the *SpMTY1* maltotriose transporter gene revealed a chimeric structure similar to that of *SeMALT413*, albeit with alternating sequence identity with *ScMAL31*, *SeMALT* and *SparMAL31*. While sequences from *S. cerevisiae* and *S. eubayanus* were already present in the genome of *S. pastorianus*, the presence of sequences from *S. paradoxus* is plausible as introgressions from *S. paradoxus* were commonly found in a wide array of *S. cerevisiae* strains (30). Therefore, the sequence of *SpMTY1* could have resulted from *in vivo* gene shuffling between genes from the *MALT* family, followed by accumulation of mutations. The emergence of *SeMALT413* could therefore be representative of the emergence of maltotriose utilization during the evolution of *S. pastorianus*. Moreover, the emergence of a maltotriose transporter after laboratory evolution of CBS 12357<sup>T</sup>, which was discovered at the same time as *SeMALT413* provides further credibility to the evolutionary importance of *in vivo* gene shuffling for gene neofunctionalization (65).

No evidence of reciprocal translocations between *SeMALT1*, *SeMALT3* and *SeMALT4* was found in the genome of IMS0750, indicating genetic introgression via non-conservative recombinations. Such introgressions can occur during repair of double strand breaks by strand invasion of a homologous sequence provided by another chromosome and resection (81), leading to localized gene conversion and loss of heterozygosity. This model, which was proposed to explain local loss of heterozygosity of two orthologous genes in an *S. cerevisiae* x *S. uvarum* hybrid (81), provides a plausible explanation of the emergence of *SeMALT413* through non-reciprocal recombination between paralogous *SeMALT* genes in *S. eubayanus*. The mosaic sequence composition of the resulting transporter gene suggests that neofunctionalization required multiple successive introgression events. As a result of these

genetic introgressions, the *SeMALT4* gene was lost. The fact that IMS0750 harbored two copies of *SeMALT413* and no copy of *SeMALT4* indicates a duplication of the newly-formed ORF at the expense of *SeMALT4* via loss of heterozygosity. As functional-redundancy enables the accumulation of mutation without losing original functions (27, 31, 32, 82), the loss of *SeMALT4* was likely facilitated by the presence of the functionally-redundant maltose transporter *SeMALT2* (9). The observation that introgressions were only found at *SeMALT4* may be due to the low number of tested mutants. However, it should be noted that introgressions in the *SeMALT1* and *SeMALT3* ORF's would have been unlikely to be beneficial, since these genes are not expressed in CBS 12357<sup>T</sup> (9).

This study illustrates the role of the rapid evolution of subtelomeric genes in adaptation to environmental changes. In *Saccharomyces* yeasts, subtelomeric regions contain a large number of gene families encoding functions critical to the interaction of a cell with its environment, such as nutrient uptake, sugar utilization, inhibitor tolerance and flocculation (27, 83-88). The high number of genes within subtelomeric families results in functional redundancy and therefore in mutational freedom (27, 31, 32, 82). In *Saccharomyces* species, many industrially-relevant brewing traits are encoded by subtelomeric gene families, such as the *MAL* genes encoding maltose utilization and the *FLO* genes encoding flocculation (89). While subtelomeric regions are difficult to reconstruct due to their repetitive nature, they encode much of the genetic diversity between genomes (90, 91). *A posteriori* sequence analysis of existing gene families can elucidate their evolutionary history. For example, the  $\alpha$ -glucosidase genes from the *MALS* family emerged by expansion of an ancestral pre-duplication gene with maltose-hydrolase activity and trace isomaltose-hydrolase activity (15). The evolution of *MALS* isomaltase genes from this ancestral gene is an example of subfunctionalization: the divergent evolution of two gene copies culminating in their specialization for distinct functions which were previously present to a lesser extent in the ancestral gene. The generation of functional redundancy by gene duplication is critical to this process as it enables mutations to occur which result in loss of the original gene function without engendering a selective disadvantage (27, 31, 32, 36, 37, 82). In contrast to subfunctionalization, neofunctionalization consists of the emergence of a function which was completely absent in the ancestral gene (40). While the emergence of many genes from a large array of organisms has been ascribed to subfunctionalization and to neofunctionalization, these conclusions were based on *a posteriori* analysis of processes which had already occurred, and not on their experimental observation (15, 27, 31, 39-42). *Ex-vivo* engineering of the subtelomeric *FLO* genes demonstrated that recombinations within subtelomeric gene families can alter their function (39). However, *in vivo* neofunctionalization within a subtelomeric gene family was never observed in real time. Here we present clear experimental evidence of neofunctionalization within a laboratory evolution experiment. The ability of *SeMALT413* to transport maltotriose proves that such *in vivo* gene shuffling is relevant for evolutionary biology. Given their high genetic redundancy of subtelomeric gene families, and the large body of evidence of gene sub- and neofunctionalization in their evolutionary history, it is likely that subtelomeric localization of genes facilitates the emergence of new functions. As a result, subtelomeric regions would not only be a hotspot of genetic diversity between different genomes, but also a preferred location for the birth of new genes and new gene functions.

While *SeMALT413* was shown to enable maltotriose utilization, it remained unclear how the UV-mutagenized cells acquired the ability to utilize maltotriose and why these mutations were insufficient to enable maltotriose utilization in wort. Since maltotriose-consuming mutants did not arise in the absence of UV-mutagenesis, the ability to utilize maltotriose likely emerged as a result of

genetic evolution rather than due to epigenetic adaptation. Moreover, the introduction of *SeMALT413* in CBS 12357<sup>T</sup> resulted in slower maltotriose utilization than IMS0750, suggesting that other mutations may contribute to the maltotriose-utilization phenotype of IMS0750. While whole genome sequencing of IMS0637-IMS0643 revealed a wide array of mutations, none affected genes which were previously linked to maltotriose utilization. The fact that *SeMALT1*, *SeMALT2*, *SeMALT3* and *SeMALT4* could be PCR amplified from the IMS0637 genome while *SeMALT413* could not, indicates that maltotriose utilization was not due to an undetected *SeMALT413* gene. In addition, alignment of short-read data to the reference genome and *de novo* genome assembly based on long-read data did not reveal any mutations affecting MAL genes, except a recombination between CHR11 and CHR7, which resulted in the loss of one of the two copies of *SeMALT1*. Since deletion of *SeMALT1* does not enable maltotriose utilization in CBS 12357<sup>T</sup> (9), this mutation is unlikely to be causal. While one of the 122 mutations affecting the UV-mutagenized strains or additional undetected mutations may have contributed to maltotriose utilization, their elucidation is beyond the scope of this study. Moreover, while overexpression of *SeMALT413* in CBS 12357<sup>T</sup> resulted in lesser maltotriose utilization than the evolved strain IMS0750, the maltotriose transporter *SeMalT413* is not necessarily suboptimal. Indeed, when overexpressing transporters suboptimal growth is commonly observed and has been attributed to imbalances between transporter activity and the subsequent metabolic steps (92, 93). Overall, regardless of the presence of other mutations contributing to maltotriose utilization, the emergence of the maltotriose transporter gene *SeMAL413* from parental genes which do not enable maltotriose transport demonstrates that gene neofunctionalization occurred.

While the introduction of *SeMALT413* in CBS 12357<sup>T</sup> via genetic engineering demonstrated its neofunctionalization, the use of GMO-strains is limited in the brewing industry by customer acceptance issues (94). However, the non-GMO evolved *S. eubayanus* isolate IMS0750 could be tested on industrial brewing wort at 7 L scale. In addition to near-complete maltotriose conversion, the maltose consumption, isoamylacetate production and diacetyl degradation of IMS0750 were superior to CBS 12357<sup>T</sup>. While the increased maltotriose consumption could be at least partially attributed to the emergence of the *SeMALT413*, it remained unclear if and what mutations could underlay the other changes. However, efficient maltose and maltotriose consumption, as well as the concomitantly increased ethanol production, are important factors determining the economic profitability of beer brewing processes (95). In addition, low residual sugar concentration, low concentrations of diacetyl and high concentrations of Isoamylacetate are desirable for the flavor profile of beer (96, 97). *S. eubayanus* strains typically generate high concentrations of 4-vinyl guaiacol, a clove-like off-flavor (98, 99), but strategy to eliminate this production in *S. eubayanus* have recently been described (98). Therefore, expansion of phenotypic landscape of *S. eubayanus* might be accelerated by combining these domesticated traits. In terms of application, the laboratory evolution approach for conferring maltotriose utilization into *S. eubayanus* presented in this paper is highly relevant in view of the recent introduction of this species in industrial-scale brewing processes (9, 100). The ability to ferment maltotriose can be introduced into other natural isolates of *S. eubayanus*, either by laboratory evolution or by crossing with evolved strains such as *S. eubayanus* IMS0750. Besides their direct application for brewing, maltotriose-consuming *S. eubayanus* strains are of value for the generation of laboratory-made hybrid *Saccharomyces* strains for brewing and other industrial applications (8, 101-103).

## Materials and methods

### Strains and maintenance

All yeast strains used and generated in this study are listed in Table 2. *S. eubayanus* type strain CBS 12357<sup>T</sup> (1) and *S. pastorianus* strain CBS 1483 (55, 104) were obtained from the Westerdijk Fungal Biodiversity Institute (Utrecht, the Netherlands). Stock cultures were grown in YPD, containing 10 g L<sup>-1</sup> yeast extract, 20 g L<sup>-1</sup> peptone and 20 g L<sup>-1</sup> glucose, at 20 °C until late exponential phase, complemented with sterile glycerol to a final concentration of 30 % (v/v) and stored at -80 °C until further use.

**Table 2: *Saccharomyces* strains used during this study**

Name	Species	Relevant genotype	Origin
CBS 12357	<i>S. eubayanus</i>	Wild-type diploid	(1)
IMS0637	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0638	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0639	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0640	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0641	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0642	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0643	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0750	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0751	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMS0752	<i>S. eubayanus</i>	Evolved strain derived from CBS 12357	This study
IMX1941	<i>S. eubayanus</i>	$\Delta$ <i>Sesga1::ScTEF1p-SeMALT2-ScCYC1t</i>	This study
IMX1942	<i>S. eubayanus</i>	$\Delta$ <i>Sesga1::ScTEF1p-SeMALT413-ScCYC1t</i>	This study
CBS 1483	<i>S. pastorianus</i>	Group II brewer's yeast, Brewery Heineken, bottom yeast, July 1927	(104)

### Media and cultivation

Plasmids were propagated overnight in *Escherichia coli* XL1-Blue cells in 10 mL LB medium containing 10 g L<sup>-1</sup> peptone, 5 g L<sup>-1</sup> Bacto Yeast extract, 5 g L<sup>-1</sup> NaCl and 100 mg L<sup>-1</sup> ampicillin at 37 °C. YPD medium was prepared using 10 g L<sup>-1</sup> yeast extract, 20 g L<sup>-1</sup> peptone and 20 g L<sup>-1</sup> glucose. Synthetic medium (SM) contained 3.0 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 5.0 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>, 7 H<sub>2</sub>O, 1 mL L<sup>-1</sup> trace element solution, and 1 mL L<sup>-1</sup> vitamin solution (56), and was supplemented with 20 g L<sup>-1</sup> glucose (SMG), maltose (SMM) or maltotriose (SMMt) by addition of autoclaved 50 % w/v sugar solutions. Maltotriose (95.8 % purity) was obtained from Glentham Life Sciences, Corsham, United Kingdom. Industrial wort was provided by HEINEKEN Supply Chain B.V., Zoeterwoude, the Netherlands. The wort was supplemented with 1.5 mg L<sup>-1</sup> of Zn<sup>2+</sup> by addition of ZnSO<sub>4</sub>·7H<sub>2</sub>O, autoclaved for 30 min at 121°C and filtered using Nalgene 0.2 µm SFCA bottle top filters (Thermo Scientific, Waltham, MA) prior to use. Where indicated, filtered wort was diluted with sterile demineralized water. Solid media were supplemented with 20 g L<sup>-1</sup> of Bacto agar (Becton Dickinson, Breda, The Netherlands). *S. eubayanus* strains transformed with plasmids pUDP052 (gRNA<sub>SesGA1</sub>) were selected on medium in which (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was replaced by 5 g L<sup>-1</sup> K<sub>2</sub>SO<sub>4</sub> and 10 mM acetamide (SM<sub>AcG</sub>: SMG) (105).

### Shake-flask cultivation

Shake-flask cultures were grown in 500 mL shake flasks containing 100 mL medium and inoculated from stationary-phase aerobic precultures to an initial OD<sub>660</sub> of 0.1. Inocula for growth experiments on SMMt were grown on SMM. In other cases, media for growth experiments and inoculum preparation were the same. Shake flasks were incubated at 20 °C and 200 RPM in a New Brunswick Innova43/43R shaker (Eppendorf Nederland B.V., Nijmegen, The Netherlands). Samples were taken at regular intervals to determine OD<sub>660</sub> and extracellular metabolite concentrations. OD<sub>660</sub>

measurements were taken with a Jenway 7200 spectrometer (Cole-Parmer, Staffordshire, UK) unless indicated otherwise.

### **Microaerobic growth experiments**

Microaerobic cultivation was performed in 250 mL airlock-capped Neubor infusion bottles (38 mm neck, Dijkstra, Lelystad, Netherlands) containing 200 mL 3-fold diluted wort supplemented with 0.4 mL L<sup>-1</sup> Pluronic antifoam (Sigma-Aldrich). Bottle caps were equipped with a 0.5 mm x 16 mm Microlance needle (BD Biosciences) sealed with cotton to prevent pressure build-up. Sampling was performed aseptically with 3.5 mL syringes using a 0.8 mm x 50 mm Microlance needle (BD Biosciences). Microaerobic cultures were inoculated at an OD<sub>660</sub> of 0.1 from stationary-phase precultures in 50 mL Bio-One Cellstar Cellreactor tubes (Sigma-Aldrich) containing 30 mL of the same medium, grown for 4 days at 12 °C. Bottles were incubated at 12 °C and shaken at 200 RPM in a New Brunswick Innova43/43R shaker. At regular intervals, 3.5 mL samples were collected in 24 deep-well plates (EnzyScreen BV, Heemstede, Netherlands) using a LiHa liquid handler (Tecan, Männedorf, Switzerland) to measure OD<sub>660</sub> and external metabolites. 30 µL of each sample was diluted 5 fold in demineralized water in a 96 well plate and OD<sub>660</sub> was measured with a Magellan Infinite 200 PRO spectrophotometer (Tecan, Männedorf, Switzerland). From the remaining sample, 150 µL was vacuum filter sterilized using 0.2 µm Multiscreen filter plates (Merck, Darmstadt, Germany) for HPLC measurements.

### **7-L wort fermentation cultivations**

Batch cultivations under industrial conditions were performed in 10 L stirred stainless-steel fermenters containing 7 L of 16.6 °Plato wort. Fermentations were inoculated to a density of 5 x 10<sup>6</sup> cells mL<sup>-1</sup> at 8 °C. The temperature was raised during 48 hours to 11 °C and increased to 14 °C as soon as the gravity was reduced to 6.5 °Plato. Samples were taken daily during weekdays and the specific gravity and alcohol content were measured using an Anton Paar density meter (Anton Paar GmbH, Graz, Austria).

### **Adaptive Laboratory Evolution**

#### **UV mutagenesis and selection**

First, we attempted to obtain maltotriose-consuming mutants without UV-mutagenesis. To this end, *S. eubayanus* CBS 12357<sup>T</sup> was grown in a 500 mL shake flask containing 100 mL SMG at 20 °C until stationary phase. Cells were washed twice with demineralized water and used to inoculate a 50 mL shake flask containing 10 mL SMMt to an OD<sub>660</sub> of 2, corresponding to an initial inoculum of approximately 10<sup>8</sup> cells. The SMMt culture was incubated at 20 °C and 200 RPM during three months. During this period, no growth was observed and HPLC measurements did not show any maltotriose consumption after three months.

In parallel, we mutagenized spores of *S. eubayanus* CBS 12357<sup>T</sup> to increase the likelihood of beneficial mutations. To this end, *S. eubayanus* CBS 12357<sup>T</sup> was grown in a 500 mL shake flask containing 100 mL SMG at 20 °C until stationary phase. The resulting cells were washed twice with demineralized water and transferred to a 500 mL shake flask containing 100 mL of 20 g.L<sup>-1</sup> potassium acetate at pH 7.0 to sporulated. After three days, the presence of ascospores was verified by optic microscopy and diluted to an OD<sub>660</sub> of 1. Of the resulting suspension, 50 mL was spun down at 4816 g for 5 min and washed twice with demineralized water. 25 mL of washed cells was poured into a 100 mm x 15 mm petri dish (Sigma-Aldrich) without lid and irradiated with a UV lamp (TUV 30 W T8, Philips, Eindhoven, The Netherlands) at a radiation peak of 253.7 nm. 25 mL of non-mutagenized and

5 mL of mutagenized cells were kept to determine survival rate. From both samples, a 100-fold dilution was made, from which successive 10-fold dilutions were made down to a 100,000-fold dilution. Then, 100  $\mu$ L of each dilution was plated on YPD agar and the number of colonies were counted after incubation during 48h at room temperatures. After 10,000-fold dilution, 182 colonies formed from the non-mutagenized cells against 84 colonies for the mutagenized cells, indicating a survival rate of 46 %. The remaining 20 mL of mutagenized cells, corresponding to about  $10^8$  cells, was spun down at 4816 g for 5 min and resuspended in 1 mL demineralized water. These mutagenized cells were added to a 50 mL shake flask containing 9 mL SMMt and incubated for 21 days at 20 °C and 200 RPM. Maltotriose concentrations were analyzed at day 0, 19 and 21. After 21 days, two 100  $\mu$ L samples were transferred to fresh shake flasks containing SMMt and incubated until stationary phase. At the end of the second transfer, single cell isolates were obtained using the BD FACSAria™ II SORP Cell Sorter (BD Biosciences, Franklin Lakes, NJ) equipped with a 488 nm laser and a 70  $\mu$ m nozzle, and operated with filtered FACSFlow™ (BD Biosciences). Cytometer performance was evaluated by running a CST cycle with CS&T Beads (BD Biosciences). Drop delay for sorting was determined by running an Auto Drop Delay cycle with Accudrop Beads (BD Biosciences). Cell morphology was analysed by plotting forward scatter (FSC) against side scatter (SSC). Gated single cells were sorted into a 96 well microtiter plates containing SMMt using a “single cell” sorting mask, corresponding to a yield mask of 0, a purity mask of 32 and a phase mask of 16. The 96 well plates were incubated for 96 h at room temperature in a GENios Pro micro plate spectrophotometer (Tecan, Männedorf, Switzerland), during which period growth was monitored as OD<sub>660</sub>. After 96 h, biomass in each well was resuspended using a sterile pin replicator and the final OD<sub>660</sub> was measured. The 7 isolates with the highest final OD<sub>660</sub> were picked, restreaked and stocked as isolates IMS0637-643. PCR amplification of the *S. eubayanus*-specific *SeFSY1* gene and ITS sequencing confirmed that all 7 isolates were *S. eubayanus*.

#### Laboratory evolution in chemostats

Chemostat cultivation was performed in Multifors 2 Mini Fermenters (INFORS HT, Velp, The Netherlands) equipped with a level sensor to maintain a constant working volume of 100 mL. The culture temperature was controlled at 20 °C and the dilution rate was set at 0.03 h<sup>-1</sup> by controlling the medium inflow rate. Cultures were grown on 6-fold diluted wort supplemented with 10 g L<sup>-1</sup> additional maltotriose (Glentham Life Sciences), 0.2 mL L<sup>-1</sup> anti-foam emulsion C (Sigma-Aldrich, Zwijndrecht, the Netherlands), 10 mg L<sup>-1</sup> ergosterol, 420 mg L<sup>-1</sup> Tween 80 and 5 g L<sup>-1</sup> ammonium sulfate. Tween 80 and ergosterol were added as a solution as described previously (56). IMS0637-IMS0643 were grown overnight at 20 °C and 200 RPM in separate shake flasks on 3-fold diluted wort. The OD<sub>660</sub> of each strain was measured and the equivalent of 7 mL at an OD<sub>660</sub> of 20 from each strain was pooled in a total volume of 50 mL. The reactor was inoculated by adding 20 mL of the pooled culture. After overnight growth, the medium inflow pumps were turned on and the fermenter was sparged with 20 mL min<sup>-1</sup> of nitrogen gas and stirred at 500 RPM. The pH was not adjusted. Samples were taken weekly. Due to a technical failure on the 63<sup>rd</sup> day, the chemostat was autoclaved, cleaned and restarted using a sample taken on the same day. After a total of 122 days, the chemostat was stopped and 10 single colony isolates were sorted onto SMMt agar using FACS, as for IMS0637-IMS0643. PCR amplification of the *S. eubayanus* specific *SeFSY1* gene and ITS sequencing confirmed that all ten single-cell isolates were *S. eubayanus*. Three colonies were randomly picked, restreaked and stocked as IMS0750-752.

### Genomic isolation and whole genome sequencing

Yeast cultures were incubated in 50 mL Bio-One Cellstar Cellreactor tubes (Sigma-Aldrich) containing liquid YPD medium at 20°C on an orbital shaker set at 200 RPM until the strains reached stationary phase with an OD<sub>660</sub> between 12 and 20. Genomic DNA for whole genome sequencing was isolated using the Qiagen 100/G kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and quantified using a Qubit® Fluorometer 2.0 (Thermo Scientific).

Genomic DNA of the strains CBS 12357<sup>T</sup> and IMS0637-IMS0643 was sequenced by Novogene Bioinformatics Technology Co., Ltd (Yuen Long, Hong Kong) on a HiSeq2500 sequencer (Illumina, San Diego, CA) with 150 bp paired-end reads using PCR-free library preparation. Genomic DNA of the strains IMS0750 and IMS0752 was sequenced in house on a MiSeq sequencer (Illumina) with 300 bp paired-end reads using PCR-free library preparation. All reads are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the bioproject accession number PRJNA492251.

Genomic DNA of strains IMS0637 and IMS0750 was sequenced on a Nanopore MinION (Oxford Nanopore Technologies, Oxford, United Kingdom). Libraries were prepared using 1D-ligation (SQK-LSK108) as described previously (91) and analysed on FLO-MIN106 (R9.4) flow cell connected to a MinION Mk1B unit (Oxford Nanopore Technology). MinKNOW software (version 1.5.12; Oxford Nanopore Technology) was used for quality control of active pores and for sequencing. Raw files generated by MinKNOW were base called using Albacore (version 1.1.0; Oxford Nanopore Technology). Reads with a minimum length of 1000 bp were extracted in fastq format. All reads are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the bioproject accession number PRJNA492251.

### Genome analysis

For the strains CBS 12357<sup>T</sup>, IMS0637-IMS0643, IMS0750 and IMS0752, the raw Illumina reads were aligned against a chromosome-level reference genome of CBS 12357<sup>T</sup> (NCBI accession number PRJNA450912, <https://www.ncbi.nlm.nih.gov/>) (9) using the Burrows–Wheeler Alignment tool (BWA), and further processed using SAMtools and Pilon for variant calling (106-108). Heterozygous SNPs and INDELS which were heterozygous in CBS 12357<sup>T</sup> were disregarded. Chromosomal translocations were detected using Breakdancer (109). Only translocations which were supported by at least 10 % of the reads aligned at that locus were considered. Chromosomal copy number variation was estimated using Magnolya (110) with the gamma setting set to “none” and using the assembler ABySS (v 1.3.7) with a k-mer size of 29 (111). All SNPs, INDELS, recombinations and copy number changes were manually confirmed by visualising the generated .bam files in the Integrative Genomics Viewer (IGV) software (112). The complete list of identified mutations can be found in Supplementary Data File 1.

For strains IMS0637 and IMS0750, the nanopore sequencing reads were assembled *de novo* using Canu (version 1.3) (113) with –genomesize set to 12 Mbp. Assembly correctness was assessed using Pilon (108), and sequencing/assembly errors were polished by aligning Illumina reads with BWA (106) using correction of only SNPs and short indels (–fix bases parameter). Long sequencing reads of IMS0637 and IMS0750 were aligned to the obtained reference genomes and to the reference genome of CBS 12357<sup>T</sup> using minimap2 (114). The genome assemblies for IMS0637 and IMS0750 are available at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the bioproject accession number PRJNA492251.



### Molecular biology methods

For colony PCR and Sanger sequencing, a suspension containing genomic DNA was prepared by boiling biomass from a colony in 10  $\mu$ L 0.02 M NaOH for 5 min, and spinning cell debris down at 13,000 g. To verify isolates belonged to the *S. eubayanus* species, the presence of *S. eubayanus*-specific gene *SeFSY1* and the absence of *S. cerevisiae*-specific gene *ScMEX67* was tested by DreamTaq PCR (Thermo Scientific) amplification using primer pair 8572/8573 (115), and primer pair 8570/8571 (116), respectively. Samples were loaded on a 1 % agarose gel containing SYBR Green DNA stain (Thermo Scientific). GeneRuler DNA Ladder Mix (Thermo Scientific) was used as ladder and gel was run at a constant 100V for 20 min. DNA bands were visualized using UV light. For additional confirmation of the *S. eubayanus* identity, ITS regions were amplified using Phusion High-Fidelity DNA polymerase (Thermo Scientific) and primer pair 10199/10202. The purified (GenElute PCR Cleanup Kit, Sigma-Aldrich) amplified fragments were Sanger sequenced (BaseClear, Leiden, Netherlands) (117). Resulting sequences were compared using BLAST to available ITS sequences of *Saccharomyces* species and classified as the species to which the amplified region had the highest sequence identity. The presence of the *SeMALT* genes was verified by using Phusion High-Fidelity DNA polymerase and gene specific primers: 10491/10492 for *SeMALT1*, 10632/10633 for *SeMALT2* and *SeMALT4/2*, 10671/10672 for *SeMALT3*, 10491/10671 for *SeMALT13*, and 10633/10671 for *SeMALT413*. The amplified fragments were purified using the GenElute PCR Cleanup Kit (Sigma-Aldrich) and Sanger sequenced (BaseClear) using the same primers used for amplification.

### Plasmid construction

All plasmids and primers used in this study are listed in Table 3 and Supplementary Table S1, respectively. DNA amplification for plasmid and strain construction was performed using Phusion High-Fidelity DNA polymerase (Thermo Scientific) according to the supplier's instructions. The coding region of *SeMALT413* was amplified from genomic DNA of IMS0750 with primer pair 10633/10671. Each primer carried a 40 bp extension complementary to the plasmid backbone of p426-TEF-amds(16), which was PCR amplified using primer pair 7812/5921. The transporter fragment and the p426-TEF-amds backbone fragment were assembled (118) using NEBuilder HiFi DNA Assembly (New England Biolabs, Ipswich, MA), resulting in plasmid pUD814. The resulting pUD814 plasmid was verified by Sanger sequencing, which confirmed that its *SeMALT413* ORF was identical to the recombinant ORF found in the nanopore assembly of IMS0750 (Fig 3C).

**Table 3: Plasmids used during this study**

Name	Relevant genotype	Source
pUDP052	<i>ori</i> (ColE1) <i>bla</i> panARSopt <i>amdSYM</i> <i>ScTDH3</i> <sub>pr</sub> -gRNA <sub>SeSGA1</sub> - <i>ScCYC1</i> <sub>ter</sub> <i>AaTEF1</i> <sub>pr</sub> - <i>Spcas9</i> <sup>D147Y</sup> <i>P411T</i> - <i>ScPHO5</i> <sub>ter</sub>	(9)
pUDE044	<i>ori</i> (ColE1) <i>bla</i> 2 $\mu$ <i>ScTDH3</i> <sub>pr</sub> - <i>ScMAL12</i> - <i>ScADH1</i> <sub>ter</sub> <i>URA3</i>	(123)
p426-TEF-amds	<i>ori</i> (ColE1) <i>bla</i> 2 $\mu$ <i>amdSYM</i> <i>ScTEF1</i> <sub>pr</sub> - <i>ScCYC1</i> <sub>ter</sub>	(16)
pUD479	<i>ori</i> (ColE1) <i>bla</i> 2 $\mu$ <i>amdSYM</i> <i>ScTEF1</i> <sub>pr</sub> - <i>SeMALT1</i> - <i>ScCYC1</i> <sub>ter</sub>	(9)
pUD480	<i>ori</i> (ColE1) <i>bla</i> 2 $\mu$ <i>amdSYM</i> <i>ScTEF1</i> <sub>pr</sub> - <i>SeMALT2</i> - <i>ScCYC1</i> <sub>ter</sub>	(9)
pUD814	<i>ori</i> (ColE1) <i>bla</i> 2 $\mu$ <i>amdSYM</i> <i>ScTEF1</i> <sub>pr</sub> - <i>SeMALT413</i> - <i>ScCYC1</i> <sub>ter</sub>	This study

### Strain construction

To integrate and overexpress *SeMALT2* and *SeMALT413* ORFs in *S. eubayanus* CBS 12357<sup>1</sup>, *SeMALT2* and *SeMALT413* were amplified from pUD480 and pUD814 respectively with primers 13559/13560 that carried a 40 bp region homologous to each flank of the *SeSGA1* gene located on *S. eubayanus* chromosome IX. To facilitate integration, the PCR fragments were co-transformed with the plasmid

pUDP052 that expressed *SpCas9*<sup>D147Y P411T</sup> (119, 120) and a gRNA targeting *SeSGA1* (9). The strain IMX1941 was constructed by transforming CBS 12357<sup>T</sup> with 1 µg of the amplified *SeMALT2* expression cassette and 500 ng of plasmid pUDP052 by electroporation as described previously (120). Transformants were selected on SM<sub>Ac</sub>G plates. Similarly, IMX1942 was constructed by transforming CBS 12357<sup>T</sup> with 1 µg of the amplified *SeMALT413* expression cassette for *SeMALT413* instead of *SeMALT2*. Correct integration was verified by diagnostic PCR with primer pair 12635/12636 (Supplementary Fig S9). All PCR-amplified gene sequences were Sanger sequenced (BaseClear).

### Protein structure prediction

Homology modeling of the *SeMalT413* transporter was performed using the SWISS-MODEL server (<https://swissmodel.expasy.org/>) (121). The translated amino acid sequence of *SeMALT413* was used as input (Supplementary Fig S4). The model of the xylose proton symporter Xyle (PDB: 4GBY) was chosen as template (57). Models were built based on the target-template alignment using ProMod3. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodeled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. In case loop modelling with ProMod3 fails, an alternative model is built with PROMOD-II (122). 3D model was assessed and colored using Pymol (The PyMOL Molecular Graphics System, Version 2.1.1 Schrödinger, LLC.).

### Sequence analysis of *SpMTY1*

The sequence of *SpMTY1* was analyzed by aligning *ScMAL31*, *ScAGT1*, *ScMPH2* and *ScMPH3* from *S. cerevisiae* strain S288C (63) and *SeMALT1*, *SeMALT2*, *SeMALT3*, *SeMALT4* from *S. eubayanus* strain CBS 12357<sup>T</sup> (9) to the sequence of *SpMTY1* from *S. pastorianus* strain Weihenstephan 34/70 (20) using the Clone manager software (version 9.5.1, Sci-Ed Software, Denver, Colorado). The origin of nucleotides 969 to 1,639 of *SpMTY1* was further investigated using the blastn function of NCBI (<https://www.ncbi.nlm.nih.gov/>). The sequence was aligned against *S. cerevisiae* S288C (taxid:559292) to identify closely related homologues. In addition, *SpMTY1* was aligned against the complete nucleotide collection. To avoid matches with genomes harboring an *MTY1* gene, sequences from *S. pastorianus* (taxid:27292), *S. cerevisiae* (taxid:4932), *S. eubayanus* (taxid:1080349), *S. cerevisiae* x *S. eubayanus* (taxid:1684324) and *S. bayanus* (taxid:4931) were excluded. The most significant alignment was with nucleotides 1,043,930 to 1,044,600 of chromosome VII of *S. paradoxus* strain YPS138 (GenBank: CP020282.1). As the most significant alignment of these nucleotides to *S. cerevisiae* S288C (taxid:559292) was *ScMAL31*, the gene was further referred to as *SparMAL31*.

### Analytics

The concentrations of ethanol and of the sugars glucose, maltose and maltotriose were measured using a high pressure liquid chromatography (HPLC) Agilent Infinity 1260 series (Agilent Technologies, Santa Clara, CA) using a Bio-Rad Aminex HPX-87H column at 65 °C and a mobile phase of 5 mM sulfuric acid with a flow rate of 0.8 mL per minute. Compounds were measured using a RID at 35 °C. Samples were spun down (13,000 g for 5 min) to collect supernatant or 0.2 µm filter-sterilized before analysis. The concentrations of ethylacetate and isoamylacetate, methanol, propanol, isobutanol, isoamyl alcohol and diacetyl were determined as described previously (55).

## Acknowledgments

We thank Jan-Maarten Geertman (Heineken Supply Chain B.V.) for his support during the study. We are grateful to dr. EmilyClare Baker and dr. Chris Todd Hittinger for delaying the public release of the non-copy-edited version of their work during the revision process of our manuscript. This work was performed within the BE-Basic R&D Program (<http://www.be-basic.org/>), which was granted an FES subsidy from the Dutch Ministry of Economic Affairs, Agriculture and Innovation (EL&I).

## References

1. Libkind D, *et al.* (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA*:201105430.
2. Sampaio JP (2018) Microbe Profile: *Saccharomyces eubayanus*, the missing link to lager beer yeasts. *Microbiology* 164(9):1069-1071.
3. Dequin S (2001) The potential of genetic engineering for improving brewing, wine-making and baking yeasts. *Appl Environ Microbiol* 56(5-6):577-588.
4. Peris D, *et al.* (2014) Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Mol Ecol* 23(8):2031-2045.
5. Bing J, Han P-J, Liu W-Q, Wang Q-M, & Bai F-Y (2014) Evidence for a Far East Asian origin of lager beer yeast. *Curr Biol* 24(10):R380-R381.
6. Gayevskiy V & Goddard MR (2016) *Saccharomyces eubayanus* and *Saccharomyces arboricola* reside in North Island native New Zealand forests. *Environ Microbiol* 18(4):1137-1147.
7. Zastrow C, Hollatz C, De Araujo P, & Stambuk B (2001) Maltotriose fermentation by *Saccharomyces cerevisiae*. *J Ind Microbiol Biotechnol* 27(1):34-38.
8. Hebly M, *et al.* (2015) *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res* 15(3).
9. Brickwedde A, *et al.* (2018) Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357T. *Front Microbiol* 9:1786.
10. Gallone B, *et al.* (2016) Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. *Cell* 166(6):1397-1410 e1316.
11. Alves SL, *et al.* (2008) Molecular analysis of maltotriose active transport and fermentation by *Saccharomyces cerevisiae* reveals a determinant role for the AGT1 permease. *Appl Environ Microbiol* 74(5):1494-1501.
12. Chang Y, Dubin R, Perkins E, Michels C, & Needleman R (1989) Identification and characterization of the maltose permease in genetically defined *Saccharomyces* strain. *J Bacteriol* 171(11):6148-6154.
13. Naumov GI, Naumova ES, & Michels C (1994) Genetic variation of the repeated MAL loci in natural populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genetics* 136(3):803-812.
14. Charron MJ, Read E, Haut SR, & Michels CA (1989) Molecular evolution of the telomere-associated MAL loci of *Saccharomyces*. *Genetics* 122(2):307-316.
15. Voordeckers K, *et al.* (2012) Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS biology* 10(12):e1001446.
16. Marques WL, *et al.* (2017) Elimination of sucrose transport and hydrolysis in *Saccharomyces cerevisiae*: a platform strain for engineering sucrose metabolism. *FEMS Yeast Res* 17(1):fox006.
17. Han EK, Cotty F, Sottas C, Jiang H, & Michels CA (1995) Characterization of AGT1 encoding a general  $\alpha$ -glucoside transporter from *Saccharomyces*. *Mol Microbiol* 17(6):1093-1107.
18. Stambuk BU, da Silva MA, Panek AD, & de Araujo PS (1999) Active  $\alpha$ -glucoside transport in *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* 170(1):105-110.
19. Baker E, *et al.* (2015) The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol Biol Evol* 32(11):2818-2831.
20. Salema-Oom M, Pinto VV, Gonçalves P, & Spencer-Martins I (2005) Maltotriose utilization by industrial *Saccharomyces* strains: characterization of a new member of the  $\alpha$ -glucoside transporter family. *Appl Environ Microbiol* 71(9):5044-5049.
21. Dietvorst J, Londesborough J, & Steensma H (2005) Maltotriose utilization in lager yeast strains: *MTT1* encodes a maltotriose transporter. *Yeast* 22(10):775-788.
22. Kupiec M (2014) Biology of telomeres: lessons from budding yeast. *FEMS Microbiol Rev* 38(2):144-171.
23. Louis EJ & Haber JE (1992) The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* 131(3):559-574.
24. Louis EJ (1995) The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* 11(16):1553-1573.
25. Song W, Dominska M, Greenwell PW, & Petes TD (2014) Genome-wide high-resolution mapping of chromosome fragile sites in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*:201406847.
26. Pryde FE, Huckle TC, & Louis EJ (1995) Sequence analysis of the right end of chromosome XV in *Saccharomyces cerevisiae*: an insight into the structural and functional significance of sub-telomeric repeat sequences. *Yeast* 11(4):371-382.

27. Brown CA, Murray AW, & Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol* 20(10):895-903.
28. Bergström A, et al. (2014) A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* 31(4):872-888.
29. Dunn B, Richter C, Kvitek DJ, Pugh T, & Sherlock G (2012) Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res*:gr. 130310.130111.
30. Peter J, et al. (2018) Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556(7701):339-344.
31. Anderson MZ, Wigen LJ, Burrack LS, & Berman J (2015) Real-time evolution of a subtelomeric gene family in *Candida albicans*. *Genetics*:genetics. 115.177451.
32. Taylor JS & Raes J (2004) Duplication and divergence: the evolution of new genes and old ideas. *Annual review of genetics* 38:615-643.
33. Voordeckers K & Verstrepen KJ (2015) Experimental evolution of the model eukaryote *Saccharomyces cerevisiae* yields insight into the molecular mechanisms underlying adaptation. *Curr Opin Biotechnol* 28:1-9.
34. Ohno S, Wolf U, & Atkin NB (1968) Evolution from fish to mammals by gene duplication. *Hereditas* 59(1):169-187.
35. Ohno S (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. *Seminars in cell & developmental biology*, (Elsevier), pp 517-522.
36. Force A, et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151(4):1531-1545.
37. Lynch M & Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154(1):459-473.
38. Tang Y-C & Amon A (2013) Gene copy-number alterations: a cost-benefit analysis. *Cell* 152(3):394-405.
39. Christiaens JF, et al. (2012) Functional divergence of gene duplicates through ectopic recombination. *EMBO Rep* 13(12):1145-1151.
40. He X & Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169(2):1157-1164.
41. Rastogi S & Liberles DA (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol Biol* 5(1):28.
42. Deng C, Cheng C-HC, Ye H, He X, & Chen L (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc Natl Acad Sci USA*:201007883.
43. Des Marais DL & Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454(7205):762.
44. Hittinger CT & Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449(7163):677.
45. Mans R, Daran J-MG, & Pronk JT (2018) Under pressure: Evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. *Curr Opin Biotechnol* 50:47-56.
46. Bachmann H, Pronk JT, Kleerebezem M, & Teusink B (2015) Evolutionary engineering to enhance starter culture performance in food fermentations. *Curr Opin Biotechnol* 32:1-7.
47. Darwin C (2004) *On the origin of species, 1859* (Routledge, London).
48. Yona AH, et al. (2012) Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad Sci USA* 109(51):21010-21015.
49. Gresham D, et al. (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* 4(12):e1000303.
50. González-Ramos D, et al. (2016) A new laboratory evolution approach to select for constitutive acetic acid tolerance in *Saccharomyces cerevisiae* and identification of causal mutations. *Biotechnol Biofuels* 9(1):173.
51. Caspeta L, et al. (2014) Altered sterol composition renders yeast thermotolerant. *Science* 346(6205):75-78.
52. Papapetridis I, et al. (2018) Laboratory evolution for forced glucose-xylose co-consumption enables identification of mutations that improve mixed-sugar fermentation by xylose-fermenting *Saccharomyces cerevisiae*. *FEMS Yeast Res* 18(6):foy056.
53. Verhoeven MD, et al. (2018) Laboratory evolution of a glucose-phosphorylation-deficient, arabinose-fermenting *S. cerevisiae* strain reveals mutations in *GAL2* that enable glucose-insensitive l-arabinose uptake. *FEMS Yeast Res*.
54. Hong K-K, Vongsangnak W, Vemuri GN, & Nielsen J (2011) Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. *Proc Natl Acad Sci USA* 108(29):12179-12184.
55. Brickwedde A, et al. (2017) Evolutionary engineering in chemostat cultures for improved maltotriose fermentation kinetics in *Saccharomyces pastorianus* lager brewing yeast. *Front Microbiol* 8:1690.
56. Verduyn C, Postma E, Scheffers WA, & Van Dijken JP (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8(7):501-517.
57. Lam V, Daruwalla K, Henderson P, & Jones-Mortimer M (1980) Proton-linked D-xylose transport in *Escherichia coli*. *J Bacteriol* 143(1):396-402.
58. Henderson R & Poolman B (2017) Proton-solute coupling mechanism of the maltose transporter from *Saccharomyces cerevisiae*. *Sci Rep* 7(1):14375.

59. Yamashita I & Fukui S (1985) Transcriptional control of the sporulation-specific glucoamylase gene in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* 5(11):3069-3073.
60. Knijnenburg TA, et al. (2009) Combinatorial effects of environmental parameters on transcriptional regulation in *Saccharomyces cerevisiae*: a quantitative analysis of a compendium of chemostat-based transcriptome data. *BMC genomics* 10(1):53.
61. Vidgren V, Huuskonen A, Virtanen H, Ruohonen L, & Londesborough J (2009) Improved fermentation performance of a lager yeast after repair of its *AGT1* maltose and maltotriose transporter genes. *Appl Environ Microbiol* 75(8):2333-2345.
62. Gorter de Vries AR, et al. (2019) Laboratory evolution of a *Saccharomyces cerevisiae* x *S. eubayanus* hybrid under simulated lager-brewing conditions. *Frontiers in Genetics* 10:242.
63. Nakao Y, et al. (2009) Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Res* 16(2):115-129.
64. Vidgren V & Londesborough J (2012) Characterization of the *Saccharomyces bayanus*-type *AGT1* transporter of lager yeast. *J Inst Brew* 118(2):148-151.
65. Baker EP & Hittinger CT (2018) Evolution of a novel chimeric maltotriose transporter in *Saccharomyces eubayanus* from parent proteins unable to perform this function. *PLoS genetics* 15(4):e1007786.
66. Cousseau F, Alves Jr S, Trichez D, & Stambuk B (2013) Characterization of maltotriose transporters from the *Saccharomyces eubayanus* subgenome of the hybrid *Saccharomyces pastorianus* lager brewing yeast strain Weihenstephan 34/70. *Lett Appl Microbiol* 56(1):21-29.
67. Nguyen H-V, Legras J-L, Neuvéglise C, & Gaillardin C (2011) Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380T. *PLoS One* 6(10):e25821.
68. Cherry JM, et al. (2011) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40(D1):D700-D705.
69. Yan N (2015) Structural biology of the major facilitator superfamily transporters. *Annu Rev Biophys* 44:257-283.
70. Smit A, Moses S, Pretorius I, & Cordero Otero R (2008) The Thr<sup>505</sup> and Ser<sup>557</sup> residues of the *AGT1*-encoded  $\alpha$ -glucoside transporter are critical for maltotriose transport in *Saccharomyces cerevisiae*. *J Appl Microbiol* 104(4):1103-1111.
71. Blount ZD, Barrick JE, Davidson CJ, & Lenski RE (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489(7417):513.
72. Mugnier MR, Cross GA, & Papavasiliou FN (2015) The *in vivo* dynamics of antigenic variation in *Trypanosoma brucei*. *Science* 347(6229):1470-1473.
73. Hall JP, Wang H, & Barry JD (2013) Mosaic VSGs and the scale of *Trypanosoma brucei* antigenic variation. *PLoS Pathog* 9(7):e1003502.
74. Boothroyd CE, et al. (2009) A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*. *Nature* 459(7244):278.
75. Jackson AP, et al. (2012) Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proc Natl Acad Sci USA* 109(9):3416-3421.
76. Hovel-Miner G, Mugnier MR, Goldwater B, Cross GA, & Papavasiliou FN (2016) A conserved DNA repeat promotes selection of a diverse repertoire of *Trypanosoma brucei* surface antigens from the genomic archive. *PLoS Genet* 12(5):e1005994.
77. Gibbs MD, Nevalainen KH, & Bergquist PL (2001) Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. *Gene* 271(1):13-20.
78. Stemmer WP (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370(6488):389.
79. Cherry JR & Fidantsef AL (2003) Directed evolution of industrial enzymes: an update. *Curr Opin Biotechnol* 14(4):438-443.
80. Rogers RL, Bedford T, Lyons AM, & Hartl DL (2010) Adaptive impact of the chimeric gene *Quetzalcoat1* in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 107(24):10943-10948.
81. Dunn B, et al. (2013) Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet* 9(3):e1003366.
82. Hernandez-Rivas R, Hinterberg K, & Scherf A (1996) Compartmentalization of genes coding for immunodominant antigens to fragile chromosome ends leads to dispersed subtelomeric gene families and rapid gene evolution in *Plasmodium falciparum*. *Mol Biochem Parasitol* 78(1-2):137-148.
83. Carlson M, Celenza JL, & Eng FJ (1985) Evolution of the dispersed *SUC* gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biol* 5(11):2894-2902.
84. Jordan P, Choe J-Y, Boles E, & Oreb M (2016) Hxt13, Hxt15, Hxt16 and Hxt17 from *Saccharomyces cerevisiae* represent a novel type of polyol transporters. *Sci Rep* 6:23502.
85. Naumov GI, Naumova ES, & Louis EJ (1995) Genetic mapping of the  $\alpha$ -galactosidase *MEL* gene family on right and left telomeres of *Saccharomyces cerevisiae*. *Yeast* 11(5):481-483.
86. Teste M-A, François JM, & Parrou J-L (2010) Characterization of a new multigene family encoding isomaltases in the yeast *Saccharomyces cerevisiae*: the *IMA* family. *J Biol Chem*:jbc. M110. 145946.
87. Denayrolles M, de Villechenon EP, Lonvaud-Funel A, & Aigle M (1997) Incidence of *SUC-RTM* telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Curr Genet* 31(6):457-461.

88. Teunissen A & Steensma H (1995) The dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* 11(11):1001-1013.
89. Verstrepen KJ & Klis FM (2006) Flocculation, adhesion and biofilm formation in yeasts. *Mol Microbiol* 60(1):5-15.
90. Istace B, et al. (2017) *de novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6(2):1-13.
91. Salazar AN, et al. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN. PK113-7D. *FEMS Yeast Res* 17(7).
92. Guimaraes PM, François J, Parrou JL, Teixeira JA, & Domingues L (2008) Adaptive evolution of a lactose-consuming *Saccharomyces cerevisiae* recombinant. *Appl Environ Microbiol* 74(6):1748-1756.
93. Boles E & Oreb M (2018) A Growth-Based Screening System for Hexose Transporters in Yeast. *Methods Mol Biol* 1713:123.
94. Varzakas TH, Arvanitoyannis IS, & Baltas H (2007) The politics and science behind GMO acceptance. *Crit Rev Food Sci Nutr* 47(4):335-361.
95. Zheng X, D'Amore T, Russell I, & Stewart G (1994) Factors influencing maltotriose utilization during brewery wort fermentations. *J Am Soc Brew Chem* 52(2):41-47.
96. Verstrepen KJ, et al. (2003) Flavor-active esters: adding fruitiness to beer. *J Biosci Bioeng* 96(2):110-118.
97. García AI, García LA, & Díaz M (1994) Modelling of diacetyl production during beer fermentation. *J Inst Brew* 100(3):179-183.
98. Diderich JA, Weening SM, van den Broek M, Pronk JT, & Daran J-MG (2018) Selection of Pof-*Saccharomyces eubayanus* Variants for the Construction of *S. cerevisiae* × *S. eubayanus* Hybrids With Reduced 4-Vinyl Guaiacol Formation. *Front Microbiol* 9.
99. Mertens S, et al. (2015) A large set of newly created interspecific yeast hybrids increases aromatic diversity in lager beers. *Appl Environ Microbiol*:AEM. 02464-02415.
100. Hittinger CT, Steele JL, & Ryder DS (2018) Diverse yeasts for diverse fermented beverages and foods. *Curr Opin Biotechnol* 49:199-206.
101. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2017) Novel brewing yeast hybrids: creation and application. *Appl Microbiol Biotechnol* 101(1):65-78.
102. Bellon JR, et al. (2011) Newly generated interspecific wine yeast hybrids introduce flavour and aroma diversity to wines. *Appl Microbiol Biotechnol* 91(3):603-612.
103. Peris D, et al. (2017) Hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production. *Biotechnol Biofuels* 10(1):78.
104. Van den Broek M, et al. (2015) Chromosomal copy number variation in *Saccharomyces pastorianus* evidence for extensive genome dynamics in industrial lager brewing strains. *Appl Environ Microbiol*:AEM. 01263-01215.
105. Solis-Escalante D, et al. (2013) amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res* 13(1):126-139.
106. Li H & Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26(5):589-595.
107. Li H, et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
108. Walker BJ, et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
109. Chen K, et al. (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 6(9):677.
110. Nijkamp JF, et al. (2012) *De novo* detection of copy number variation by co-assembly. *Bioinformatics* 28(24):3195-3202.
111. Simpson JT, et al. (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res*:gr. 089532.089108.
112. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24.
113. Koren S, et al. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*:gr. 215087.215116.
114. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 1:7.
115. Pengelly RJ & Wheals AE (2013) Rapid identification of *Saccharomyces eubayanus* and its hybrids. *FEMS Yeast Res* 13(2):156-161.
116. Muir A, Harrison E, & Wheals A (2011) A multiplex set of species-specific primers for rapid identification of members of the genus *Saccharomyces*. *FEMS Yeast Res* 11(7):552-563.
117. Schoch CL, et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci USA* 109(16):6241-6246.
118. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343.
119. Bao Z, et al. (2014) Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol* 4(5):585-594.
120. Gorter de Vries AR, de Groot PA, Broek M, & Daran J-MG (2017) CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb Cell Fact* 16(1):222.

121. Biasini M, *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42(W1):W252-W258.
122. Guex N, Peitsch MC, & Schwede T (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 30(S1):S162-S173.
123. de Kok S, *et al.* (2011) Increasing free-energy (ATP) conservation in maltose-grown *Saccharomyces cerevisiae* by expression of a heterologous maltose phosphorylase. *Metab Eng* 13(5):518-526.

## Chapter 10: Outlook

Recent progress in genome sequencing and genome editing technologies have improved our understanding of the complex hybrid genomes of *S. pastorianus*. In addition to yielding chromosome-level genome assemblies, increased accuracy of long-read sequencing technologies may soon enable chromosome copy haplotyping (1), particularly in combination with emerging assembly algorithms for haplotype phasing (2, 3). Analogous to recent developments in *S. cerevisiae* and *S. eubayanus*, chromosome-level reference genomes will likely contribute to improved understanding of the genome complexity and plasticity of *S. pastorianus*, and to simplifying and accelerating non-GM strain improvement strategies such as mutagenesis and laboratory evolution (4-7). The emergence of Cas9 genome editing tools compatible with *S. pastorianus* enables the use for these high-quality genome assemblies for functional characterisation of genes (8), determination of targets for non-GM techniques (9) and reverse engineering after non-GM strain improvement methods (4). In addition, current developments in GM regulation outside the EU may lead to the direct applicability of genetically engineered strains, particularly when no heterologous DNA is introduced (10, 11). Consolidations in the brewing industry during the 20<sup>th</sup> century have transformed brewing companies into international conglomerates with broad portfolios of beer brands (12, 13). Such conglomerates are unlikely to adopt GM yeasts for brewing, as customer acceptance backlash may not be restricted to a specific beer brand or customer market, but could result in decreasing sales of their entire brand portfolio over all markets. However, the 21<sup>st</sup> century saw a revitalization of the declining beer market, resulting in the emergence of many new breweries commonly referred to as craft- and microbreweries (14, 15). Due to their small volumes and the presence of numerous competing beer brands, microbreweries generally strive towards clearly defined product identity to target highly specific customer segments (16, 17). GM-technology could be used to obtain characteristics which are popular in the microbrewery customer market, such as environmental sustainability and product uniqueness (18, 19). For example, the use of GM yeast without diacetyl production could reduce the energy requirements of lager brewing by alleviating the need for lagering, which typically requires cooling during time periods of about two weeks (20). Similarly, the introduction of genes for the production of hop flavours, can result in lesser water, land and energy usage due to decreased hop requirements (21). The introduction of genes for the production of novel flavour compounds can generate novel products clearly distinct from other brands (21), and fits into the recent commercial success of beers with fruity flavour additives, such as *Radler* (22). While GM microbreweries could theoretically target progressive market segments with high GM acceptance specifically, technological and financial hurdles to generate and implement genetically modified yeast have been prohibitive. However, the development of efficient gene-editing tools has considerably lowered such hurdles and popularized genetic editing, as illustrated in the extreme by the biohacking movement (23, 24). Despite these developments, the genetic complexity of *S. pastorianus* may continue to limit its genetic amenability, as illustrated by the emergence of unwanted loss of heterozygosity upon editing of heterozygous loci (25).

Sequencing of *S. pastorianus* genomes has revealed limited genetic diversity relative to the genetic diversity of *S. cerevisiae* and *S. eubayanus* strains (26-29). Due to their sexual isolation, the genetic diversity of allopolyploids is limited to the genomic material of their parental strains (30). The short evolutionary history of *S. pastorianus* strains resulted in limited accumulation of mutations (31).



Moreover, the genetic distinctness of Group 1 and Group 2 strains, despite their common ancestry, suggests that their occurrence has involved two population bottlenecks (32), possibly coinciding with the isolation of pure Group 1 strains from a Carlsberg brewery by Hansen and of Group 2 strains from a Heineken brewery by Elion (33-35). Despite their limited genetic divergence, Group 1 and 2 isolates display important phenotypic differences (36). Generation of novel hybrids has been successfully used to expand the genetic and phenotypic diversity of lager brewing strains, both with *S. cerevisiae* x *S. eubayanus* hybrids and with non-conventional lager brewing hybrids (37-39). In combination with mass mating and high throughput strategies, the generation of laboratory-made hybrids is a potent generator of novel lager brewing strains (40, 41). Such hybrids should be suitable for industrial application as they can be generated using non-GMO methods and are of sufficient genetic stability for several consecutive batches (4, 42, 43). As a result, laboratory-made lager brewing strains may complement or even replace traditional *S. pastorianus* strains for lager brewing in the near future. The long-running interest in *Saccharomyces* strain improvement has resulted in both GMO and non-GMO methods which can be applied directly to improve and diversify lager brewing performance, both for traditional *S. pastorianus* strains and for laboratory-made lager hybrids.

## References

1. Wenger AM, *et al.* (2019) Highly-accurate long-read sequencing improves variant detection and assembly of a human genome. *bioRxiv*:519025.
2. He D, Saha S, Finkers R, & Parida L (2018) Efficient algorithms for polyploid haplotype phasing. *BMC Genomics* 19(2):110.
3. Chin C-S, *et al.* (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13(12):1050.
4. Gorter de Vries AR, *et al.* (2019) Laboratory evolution of a *Saccharomyces cerevisiae* x *S. eubayanus* hybrid under simulated lager-brewing conditions. *Frontiers in Genetics* 10:242.
5. Brickwedde A, *et al.* (2018) Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357T. *Front Microbiol* 9:1786.
6. Mans R, Daran J-MG, & Pronk JT (2018) Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. *Curr Opin Biotechnol* 50:47-56.
7. Brouwers N, *et al.* (2019) In vivo recombination of *Saccharomyces eubayanus* maltose-transporter genes yields a chimeric transporter that enables maltotriose fermentation. *PLoS Genetics* 15(4):e1007853.
8. Gorter de Vries AR, de Groot PA, Broek M, & Daran J-MG (2017) CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. *Microb Cell Fact* 16(1):222.
9. Diderich JA, Weening SM, van den Broek M, Pronk JT, & Daran J-MG (2018) Selection of Pof *Saccharomyces eubayanus* variants for the construction of *S. cerevisiae* x *S. eubayanus* hybrids with reduced 4-vinyl guaiacol formation. *Front Microbiol* 9:1640.
10. Ishii T & Araki M (2017) A future scenario of the global regulatory landscape regarding genome-edited crops. *GM Crops Food* 8(1):44-56.
11. Waltz E (2016) Gene-edited CRISPR mushroom escapes US regulation. *Nature* 532(7599):293.
12. Poelmans E & Swinnen JF (2011) *A brief economic history of beer* (Oxford University Press, Oxford, UK).
13. Howard PH (2014) Too big to ale? Globalization and consolidation in the beer industry. *The Geography of Beer*, (Springer, New York), pp 155-165.
14. Carroll GR & Swaminathan A (2000) Why the microbrewery movement? Organizational dynamics of resource partitioning in the US brewing industry. *Am J Sociol* 106(3):715-762.
15. Ellis V & Bosworth G (2015) Supporting rural entrepreneurship in the UK microbrewery sector. *Br Food J* 117(11):2724-2738.
16. Maier T (2016) Sources of Microbrewery Competitiveness in the Czech Republic. *Agris* 8(4):14.
17. Thurnell-Read T (2014) Craft, tangibility and affect at work in the microbrewery. *Emot Space Soc* 13:46-54.
18. Carr AM (2017) Microbrewery consumer behavior. (University of Alabama Libraries, Tuscaloosa).
19. Williams AG & Mekonen S (2014) Environmental performance of traditional beer production in a micro-brewery. *Proceedings of the 9th International Conference on Life Cycle Assessment in the Agri-Food Sector*, (American Center for Life Cycle Assessment), pp 1535-1540.
20. Duong C, *et al.* (2011) Identification of Sc-type *ILV6* as a target to reduce diacetyl formation in lager brewers' yeast. *Metab Eng* 13(6):638-647.
21. Denby CM, *et al.* (2018) Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. *Nat Commun* 9(1):965.

22. Paixão SMCN (2015) Sagres Radler case study: attracting non-beer consumers to the beer category. (Católica-Lisbon School of Business & Economics, Lisbon).
23. Bennett G, Gilman N, Stavrianakis A, & Rabinow P (2009) From synthetic biology to biohacking: are we prepared? *Nat Biotechnol* 27(12):1109.
24. Yetisen AK (2018) Biohacking. *Trends Biotechnol* 36(8):744-747.
25. Gorter de Vries AR, et al. (2019) Allele-specific genome editing using CRISPR-Cas9 is associated with loss of heterozygosity in diploid yeast. *Nucleic Acids Res* 47(3):1362-1372.
26. Okuno M, et al. (2016) Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res* 23(1):67-80.
27. Peris D, et al. (2016) Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast *Saccharomyces eubayanus*. *PLoS Genet* 12(7):e1006155.
28. Gallone B, et al. (2016) Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166(6):1397-1410. e1316.
29. Peter J, et al. (2018) Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556(7701):339.
30. Mallet J (2007) Hybrid speciation. *Nature* 446(7133):279.
31. Monerawela C & Bond U (2017) Brewing up a storm: The genomes of lager yeasts and how they evolved. *Biotechnol Adv* 35(4):512-519.
32. Salazar AN, et al. (2019) Nanopore sequencing and comparative genome analysis confirm lager-brewing yeasts originated from a single hybridization. *bioRxiv*:603480.
33. Hansen EC (1883) Recherches sur la physiologie et la morphologie des ferments alcooliques. V. Methodes pour obtenir des cultures pures de *Saccharomyces* et de microorganismes analogues. *Compt Rend Trav Lab Carlsberg* 2:92-105.
34. Walther A, Hesselbart A, & Wendland J (2014) Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. *G3 (Bethesda)*:g3. 113.010090.
35. Gélinas P (2010) Mapping early patents on baker's yeast manufacture. *Compr Rev Food Sci Food Saf* 9(5):483-497.
36. Gibson BR, Storgårds E, Krogerus K, & Vidgren V (2013) Comparative physiology and fermentation performance of Saaz and Froberg lager yeast strains and the parental species *Saccharomyces eubayanus*. *Yeast* 30(7):255-266.
37. Nikulin J, Krogerus K, & Gibson B (2018) Alternative *Saccharomyces* interspecies hybrid combinations and their potential for low-temperature wort fermentation. *Yeast* 35(1):113-127.
38. Gonçalves P, Valério E, Correia C, de Almeida JM, & Sampaio JP (2011) Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One* 6(6):e20739.
39. Mertens S, et al. (2015) A large set of newly created interspecific yeast hybrids increases aromatic diversity in lager beers. *Appl Environ Microbiol*:AEM. 02464-02415.
40. Steensels J, et al. (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol Rev* 38(5):947-995.
41. Steensels J, Meersman E, Snoek T, Saels V, & Verstrepen KJ (2014) Large-scale selection and breeding to generate industrial yeasts with superior aroma production. *Appl Environ Microbiol*:AEM. 02235-02214.
42. Krogerus K, Magalhães F, Vidgren V, & Gibson B (2015) New lager yeast strains generated by interspecific hybridization. *J Ind Microbiol Biotechnol* 42(5):769-778.
43. Gorter de Vries AR, et al. (2019) Phenotype-independent isolation of interspecies *Saccharomyces* hybrids by dual-dye fluorescent staining and fluorescence-activated cell sorting. *Frontiers in Microbiology* 10:871.



## Acknowledgments

The four years that I spent pursuing a PhD in the industrial microbiology group have been a wonderful and formative experience. I could not imagine a place or a job which could have made me happier, and for that I am deeply thankful to the amazing people that surrounded me during my journey.

**Jean-Marc**, I am forever grateful for the perfect balance of guidance and freedom that you gave me over the years. You were the perfect supervisor for me and I could not have learned more under anybody else's supervision. You have always found time to discuss the challenges that I faced and to help me find the answers to my questions. While I could always count on your help, you also trusted me with increasing autonomy and freedom, thereby allowing me to develop myself. I particularly appreciated the role you entrusted me with in the projects of Philip, Hannes, Nick and Alex. I also want to thank you for having arranged funding for my project so well: costs were never an issue to perform experiments and that is a true luxury in science. You have my deepest gratitude and I wish you all the best with your future PhDs!

Of course, I had the great privilege of having two amazing promotors. **Jack**, my gratitude to you already began in 2012: your letter of recommendation may well have been my greatest asset when applying for my master's at ETH Zürich. Moreover, my time in the IMB group and our discussions after the completion of my BEP laid the foundations of my decision to embark on a PhD. Regarding my time at IMB, I want to thank you first and foremost for the wonderful working atmosphere that you have created at IMB. Your genuine interest in the personal and professional development of students, technicians, PhD students and postdocs, have shaped IMB with a collaborative spirit which I enjoyed tremendously. I have never seen the same combination of scientific ambition and *gezelligheid* in any other group, and I thank you for your tireless efforts to keep it that way. In your role as my promotor, I want to thank you specifically for your contagious energy and enthusiasm, for your wonderful (and sometimes crazy) ideas, and for the incredible speed and thoroughness with which you have given feedback on the multiple pieces which we wrote together. I also want to thank you for supporting my involvement in the departmental and faculty PhD committees, and for your openness when discussing issues within or concerning the group. Finally, I want to thank you for your support during my job search wonderings and wanderings.

During my PhD, I was able to build on the teachings of several scientists who supervised me in the past. **Dani**, thank you for your guidance during my BEP and for your friendship since then. Your passion for science (and for arguing about it!) have certainly inspired me and your teachings have kept resonating until today. In particular, your writing tips were really useful and the care and dedication with which you supervised students have been a great example for me. I am also grateful for the interest you have taken in my development since those days, and for being available as a mentor to me. **Ton**, I really appreciated our interactions during my BEP and finishing the acetic acid manuscript with you. Your pragmatic approach towards writing and towards science in general were invaluable lessons for me. You were one of the reasons why I wanted to return to IMB and I was very sad to see you leave. I am happy that you found your spot in Sweden, and it was great to visit you there. I hope to grab a pint with you soon and wish you lots of happiness with your nascent family! **Steven**, thank you for your very involved supervision: while you were really busy, you always managed to find the time to help me when needed. I hope to have made my students feel as cared

about as I did during my project under your supervision. **Sven**, thank you for the interesting feedback and questions about my project. I really appreciated our discussions about science, department politics and the educational programme before and after my project in your group, thank you for letting me voice my opinions in this way. **Tarik**, thank you for introducing me to the world of immunology and sequencing! My project with you taught me to think big when setting up experiments and to be open to new technologies and methods. Your energy and your dedication to work have made a great impression on me and I am really thankful for your willingness to find time for me outside working hours, especially after you started combining my supervision with your new job at Roche. Thank you for your mentoring since those days and for helping me figure out what I want and how to get there. It is always great to see you! **Sai**, thank you for all the great discussions we had about science and for all the fun times outside of work! An oddly specific thing I am grateful for: you introduced me to the academic beer cascade: where supervisors pay drinks for their students, who repay the system by buying drinks in the future for their own students. A wonderful system that I have applied with great pleasure. More seriously, I am happy to have contributed to the great work of your group and proud that you could use some of my data in your projects. Thanks for your efforts to find time for me when I am in Basel!

While I learned a lot from my supervisors, I learned at least as much from the students whose projects I had the privilege to supervise. **Roderick, Aafke, Anissa, Liset, Flip, Line, Ewout, Dion, Charlotte, Stephanie, Luuk, Maaike** and **Pascal**: I want to thank you for entrusting me with your supervision and for the great dedication and passion with which you have worked during your projects! Supervising you has been the highlight and the most satisfying part of my PhD adventure, and I am very proud of every single one of you. Your hard work has contributed greatly to the results presented in this thesis. More importantly, your involvement inspired me to seek the best of myself, and to try to help you develop yourselves. Working with you was also a lot of fun (both in the lab and in bar het Lab). I follow your fates with great interest and I remain happy to help you where I can. I hope to continue seeing you all! Flip and Charlotte, special thanks to you for accepting to be my paranymphs. It was great fun to supervise your MEPs, it is great fun to work with you as colleagues and I am sure it will be great fun for others that you are in charge of my cabaret.

Researching the genomes of lager brewing yeast would not have been the same without the involvement of our industrial partners at Heineken. **Niels**, I want to thank you for your great role during my PhD, in particular for your scientific input, your feedback on my manuscripts and your general enthusiasm about my research. I appreciated your communication style and the way in which you handled the whole FS10 collaboration. Furthermore, I am grateful for your efforts to make our work relevant industrially and to align the interests of Heineken and IMB. I fondly (but sometimes a bit vaguely) remember our conference adventures and our beer expeditions in Haarlem and Amsterdam. It has been great fun working with you, and I hope that we will remain friends beyond this project. I want to finish with thanking you for how you combined becoming a father and fighting for the success of some of our projects in the last months of my PhD. I wish you and your young family all the best! **Jan-Maarten**, I want to thank you for your enthusiasm, for your valuable suggestions and perhaps most of all for your support during my job search. Our discussions have helped me tremendously to explore what I want in a career and I felt supported by you as by a mentor. Thank you for sharing your taste in whiskey with me in Ireland; it would be an understatement to call that experience a revelation. **Tom**, I want to thank you for your know how and critical questions regarding fermentation and evolutionary set-ups: it was great to have

somebody who really knows what details are important. **Viktor**, thank you for the enthusiasm with which you have started your new position and for your interest in my wildly-divergent research topics.

In contrast to the dusty image of an ivory tower that some people may have regarding a PhD, my PhD project was conducted as a close collaboration between bioinformaticians and wet-lab biotechnologists. **Marcel**, I want to thank you first for the support you have given me in literally every single research line I worked on. Without you, I would have been completely lost in the seas of sequencing data we generated during the FS10 projects. Not only did you always find time to process data for me, you also took the time to think with me about how best to approach the different problems I encountered and to explain to me how these complex operations worked. I really enjoyed tackling these challenges together with you! Not only did this allow me to run scripts on my own (matrix-style screens!), it also formed a bioinformatics basis without which I would never have been able to work on CEN.PK113-7D and CBS 1483 with Alex. Beyond your bioinformatics input, I am grateful for the aura of calmness and peace you irradiate: talking with you has appeased me more than once. Thank you for all the discussions and relaxing tea breaks! **Nick** and **Alex**, it has been wonderful to do my PhD with two companions of fate. My project became a lot more fun when the two of you also started your projects, and our adventures at Boostcamp and in Dublin were a bonding experience that I am unlikely to forget. Alex, thanks for keeping your word in the “no trousers” bet, that was the most hilarious quarterly meeting ever, although closely followed by your Christmas trees and Alpacas presentation! I appreciate that you introduced me to the marvels of nanopore sequencing and bioinformatics, and that you invited me into the CEN.PK113-7D and CBS 1483 research lines. I learned a lot from both projects! **Nick**, I thank you for your input on many of my research lines, and for the close collaboration we had. I particularly enjoyed figuring out the *MAL7413* mystery with you. Thanks for all the fun we had at conferences in Dublin, Cork and Bariloche, and for all the adventures in Delft, Den Haag and Haarlem (and on the highway?!). If you ever need to get home, don’t forget to order your taxi at the Julianalaan, regardless of where you actually are! **Susan**, thank you for your support, among others with the FACS hybrids! You were tirelessly persistent and despite not having reached everything we had envisioned, your help was very valuable. I also want to thank you for all the many talks we had, in between experiments, ski slopes or on the bike after a workday or some late night partying! **Anja**, thanks for introducing me to the scientific (and consumptive) world of lager beer when I started my PhD. I remember lots of discussions about *MAL* genes, evolution and strain improvement. I also fondly remember PYFF in Lisbon, ISSY in Cork, CHEF gel adventures, Mario Kart evenings at your place and beer ascensions at the Locus Publicus! **Thomas**, thank you for your involvement in my bioinformatics endeavours and for your pragmatism. Trying to understand what Alex was up to was not always easy, thank you for lightening the mood during those meetings.

When choosing where to do my PhD, some professors could not believe their ears when I said that one of the determining advantages for IMB was the presence of great staff technicians, which ensure the technical knowhow of the group. **Pilar**, **Marijke** and **Erik**, thank you for keeping IMB afloat: without you, we PhDs wouldn’t be much more than children playing in a really expensive sandbox, especially during the start-up of our projects. Pilar, thank you for all the fun we’ve had over the years during our many lunches, dinners and party nights! You are someone who I always felt I could be really open with and I greatly value our friendship. I am glad we got the chance to work together directly on the Cas9 project. I know being the MolBio queen of IMB is not an easy job, so I am

thankful for all your efforts to keep things functioning and *gezellig* for all of us. Finally, thank you for drawing the cover of my thesis: I am really happy with the result and it is so much more meaningful because *you* made it. Marijke, thank you for our FACS adventures and for the great friendship that came from them. I remember I was a bit intimidated to spend two hours in the car with you to go to BD in my first year, but the two hours flew by with fun and meaningful discussions. All our tea breaks, dinners and FACS sessions since then have been just as pleasant! I greatly value your advice and your power to make me self-reflect, and I thank you for making me feel coached and cared for. I also thank you for introducing me to Laura. Laura: thanks for all the great pastries and discussions. I am sure things will go well for you, whatever adventure you decide to embark on in the coming years! Erik, thanks for your expertise in fermentation and laboratory set-ups, for the tea breaks and for introducing me and so many others to bouldering! Your stories almost invariably bring a smile to my face and it is great that you join so often for drinks and other activities of IMB. I also want to thank you for your sense of duty and for your reliability: whether it is cancelling plans to save a fermentation or staying behind to keep someone with car trouble company, you are someone who can truly be counted on.

I had the great pleasure of being a paranymph for, or together with, six of my favourite colleagues at IMB. **Robert**, thank you for dancing in your bee suit for Xavier and me: it was great fun to organize your cabaret. When I came to IMB you were the experienced PhD from whom I wanted to learn everything: I thank you for the patience and genuine pleasure with which you shared your knowledge with me and with everyone else at IMB. Your ability to involve others in your work, your attitude towards the students you supervise and the enthusiasm with which you talk about your work have been a shining example for me. I admire your choice for an academic career and trust that you will be a wonderful supervisor to many PhDs in the years to come. **Xavier**, thank you for saving Robert's cabaret with your musicality and your sense of rhythm. We have been companions of fate for the full four years of my PhD, and I have always felt lucky to be able to discuss a PhD's challenges and torments with you. Your sharp scientific mind has yielded ideas for my research in more than one of our discussions, and I enjoyed our talks on the workings of the academic world. As my go-to climbing buddy, I had a lot of fun pushing our limits with you and trying not to get out-competed by you: a great way to get our heads out of the work pressure! **Ioannis**, I think few people can say that they publicly dared to make fun of you during your PhD. For that and for so much more, I thank you for trusting me to arrange your cabaret. I have so many good memories from Boostcamp, PYFF, IMB and our regular dinners since you started at DSM. In addition to being a valued friend, you have been a great colleague to look up to: I admire your dedication and cut-the-crap pragmatism and I wish I had your persistence and resilience. I am happy to be housemates with you and Nicolò! **Hannes**, thanks for all the pizza evenings and for the paranymphing fun! I also want to thank you for our collaboration on the "Cas9 for weird yeast" work: I really enjoyed working with somebody so well organized. Smoothest manuscript writing process ever, if you ask me. **Paschmini**, we have always had so much fun together! Whether it was in the office, on the climbing walls or partying, I don't think I ever laughed as often as I did with you! Thank you for listening to all my dating (mis)adventures and for relativizing them where needed. I think I should also thank you, although reluctantly, for slowly converting me to your vegetarianism. I was really sad when you returned to Switzerland because I lost my buddy, but it was a lot of fun to organize your cabaret. I hope we stay in touch and keep having as much fun together as we always have! **Nicolò**, you and I were the perfect partners in crime for Jasmine's defence. Thank you for being a pleasant housemate, a caring colleague and a great

friend. I have to admit I still don't always get your sense of humour, but I have certainly grown to enjoy it. Thank you for your willingness to talk about the things that really matter in life, for sharing the things that trouble you and for genuinely caring about things that trouble me.

The cultivation of microorganisms is not possible by IMB standards without reproducible media, cultivation vessels and vast amounts of glassware. **Apilena, Astrid** and **Jannie**, thank you for how well you run the MSD kitchen. You always try to think along with our long-term needs, and I am thankful for the many occasions in which you saved my planning, for example with an autoclave session at an odd hour or a back-up stock of some much-needed chemical. Moreover, it is always nice to chat with you in the process! Apilena, thank you for welcoming me and others to a copious dinner at your home and for the delicious food you regularly share with me! Jannie, thank you for thinking along with our needs when we were shuffling around the first year practical programme.

**Pascale**, while you were not directly involved in my projects, I want to thank you for your scientific contributions. I also greatly enjoyed being a practical assistant on your course and getting the opportunity to contribute to some of the remodelling of its curriculum. I also thank my fellow practical assistants and instructors for all the great experiences!

I thank the Biotechnology department and the Applied Sciences faculty management for their investment to involve PhDs in discussions about policy issues, and for listening to our concerns. My involvement in the departmental PhD committee and in the faculty PhD council felt like meaningful and fruitful attempts to change things for the better. Particularly **Maita** and **Lucas** have been welcoming of our input and very empowering. I also thank my fellow PhD committee/council members for their dedication to improving the working environment of current and future PhDs.

While I cannot possibly thank everyone individually here, I am greatly thankful to everyone who contributed to making the IMB group a pleasant working environment, including the BSc and MSc students. The atmosphere of camaraderie and everyone's willingness to go the extra mile to keep things working smoothly are unparalleled in my experience. The culture of joining for after work activities, such as sports, dinners, movie nights, game nights and ski holidays, has brought me close to many of you. Moreover, I have felt truly cared for, and I have come to care for the personal well-being of many of you. Thank you all!

**Anna**, thank you so much for all of our adventures. I knew from the moment you joined IMB that we would get along really well, and your energy filled the second half of my PhD with life. What started with an irresistible affinity to party together soon became a strong bond that I hope never to lose. I care about you deeply.

**Heren van Cosa Nostra**, hartelijk dank voor alle lol die wij over de jaren beleefd hebben. Onze stedentrips, party-nachten, ski vakanties, nieuwjaarsvieringen, zelf gebrouwen cider/wijn/bier sessies en andere ondernemingen waren een zeer welkome afwisseling van mijn PhD, al moest ik er af en toe wel echt van bijkomen voor ik weer door kon. Jaren geleden vormden wij een club en ik vind het onvoorstelbaar waardevol dat wij nog steeds zo hecht zijn. Ik vind het leuk om te zien hoe wij langzamerhand uit het wilde van de eerste jaren groeien, en daarvoor een steeds diepere band in de plek komt. We worden langzamerhand toch wel echt volwassen, maar de lol gaat gewoon door. **Elgar, Rufus** en **Pieter**, dank voor alle ontspannende avonden die wij als huisgenoten gedeeld hebben tijdens mijn PhD. Samen konden wij altijd elkaars tegenslagen doen vergeten en onze meevallers



uitgebreid vieren, wat gelukkig tot een flinke champagne consumptie geleidt heeft. **Iris en Noortje**, dank dat jullie hier ook aan bijgedragen hebben!

**Ioana**, thank you for your support and for believing in me when I decided to return to the Netherlands for my PhD. You made me learn so much about myself and I think back of our time together with melancholy. I admire your compassion and your willingness to forgive others for their flaws, and I am deeply proud of how you succeeded at your own PhD. I am grateful that we are still in touch and I still highly value your advice and opinions.

Lieve **papa** en **mama**, woorden zijn onvoldoende om mijn dank aan jullie uit te drukken. Jullie hebben mij gemaakt tot wie ik ben en hebben mij geweldig ondersteund tijdens mijn PhD. Pap, zoals ik in het voorwoord van dit geschrift heb proberen te verwoorden, heb ik enorme bewondering voor je en denk ik dat je de oorsprong van veel van mijn drijfveren bent. Mam, ik denk dat ik aan jou mijn voorliefde voor argumenteren en mijn vurigheid daarin te danken heb! Ook heb je ons pro-activiteit en zelfstandigheid met de paplepel ingegoten en heb ik dankzij jou de gewoonte om werk te maken van wat ik wil en om mensen gewoon aan te spreken als ik iets met ze bereiken wil. Dank jullie wel dat jullie ons altijd met zoveel liefde en respect opgevoed hebben. Men zegt wel eens dat je het moet hebben van je familie; mijn wiegje had nergens beter kunnen staan dan bij jullie. **Philip** en **Sophie**, ik vind het heel erg fijn dat wij met zijn allen zo goed overweg kunnen en dat het altijd gezellig is als wij elkaar zien! Het is ontzettend leuk om mijn passie voor wetenschap en voor biologie met jullie te delen. Onze bezoeken aan elkaars studentenleven waren altijd heel erg leuk en ik kijk er naar uit jullie tijdens jullie PhDs te komen bezoeken! Ik ben super trots op jullie, en wens jullie veel succes!

Finally, I would like to thank lady luck for the mostly favourable dice throws that I have been handed over the years. I am particularly fond of the great gifts evolution has given me in the various laboratory evolution experiments. I wish the same luck upon others, and would like to remark that some misfortunes can result in valuable insights, as illustrated by chapter 5 of this thesis. While I do not believe in any higher entity guiding my dice, I wish to acknowledge the lucky star under which my thesis seems to have taken place.

## Curriculum Vitae

Arthur Roelof Gorter de Vries was born on the 7<sup>th</sup> of February 1993 in Voorburg, the Netherlands. As a result of his father's employment as a chemical process engineer, Arthur grew up in Germany between 1994 and 2000, and in France between 2000 and 2010, where he received his secondary education at the *Cité Scolaire Internationale de Lyon*, specialized in Biology, and obtained a French *Baccalauréat Scientifique* and a German *Abitur*. In 2010, Arthur enrolled in the Life Science and Technology bachelor programme at the University of Leiden and the Delft University of Technology. He minored in advanced Life Sciences and Technologies and performed his bachelor thesis in Prof. Jack Pronk's Industrial Microbiology group by studying acetic acid tolerance of yeast under the supervision of dr. Daniel González-Ramos. In 2013, Arthur started a master's in Biotechnology at the Swiss Federal Institute of Technology (ETH Zürich), during which he was also president of the university's Biotechnology Student Association. Arthur performed a six month research project on screening for novel antibiotics in Prof. Sven Panke's Bioprocess Laboratory under the supervision of dr. Steven Schmitt, and spent twelve months researching murine B-cell immune responses in Prof. Sai Reddy's Laboratory for Systems and Synthetic Immunology under the supervision of dr. Tarik Khan. In 2015, Arthur returned to the Industrial Microbiology group in Delft to embark on a PhD studying the genetics of lager brewing yeast under the supervision of dr. Jean-Marc Daran, in collaboration with Heineken. During his PhD, Arthur supervised thirteen student research projects, co-authored various scientific publications, was inventor on two patent applications and represented his fellow PhDs in the departmental and faculty PhD committees. The results generated during Arthur's PhD project form the basis of this thesis. In June 2019, Arthur joined the Amsterdam office of the global strategy consultant Roland Berger.



## List of Publications

Khan TA, Friedensohn S, Gorter de Vries AR, Straszewski J, Ruscheweyh H-J, Reddy ST (2016) **Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting**, *Science advances* 2(3):e1501371.

González-Ramos D, Gorter de Vries AR, Grijsseels SS, van Berkum MC, Swinnen S, van den Broek M, Nevoigt E, Daran J-MG, Pronk JT, van Maris AJA (2016) **A new laboratory evolution approach to select for constitutive acetic acid tolerance in *Saccharomyces cerevisiae* and identification of causal mutations**, *Biotechnology for Biofuels* 9:173.

Gorter de Vries AR, Pronk JT, Daran J-MG (2017) **Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts**, *Applied and Environmental Microbiology* 83:e03206-16.

Salazar AN\*, Gorter de Vries AR\*, van den Broek M, Wijsman M, de la Torre Cortés P, Brickwedde A, Brouwers N, Daran J-MG, Abeel T (2017) **Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN. PK113-7D**, *FEMS Yeast Research* 17(7):fox074.

Gorter de Vries AR\*, de Groot PA\*, van den Broek M, Daran J-MG (2017) **CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus***, *Microbial Cell Factories* 16:222.

Juergens H, Varela JA, Gorter de Vries AR, Perli T, Gast VJM, Gyurchev NY, Rajkumar AS, Mans R, Pronk JT, Morrissey JP, Daran J-MG (2018) **Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid**, *FEMS yeast research* 18 (3), foy012.

Gorter de Vries AR, Couwenberg LGF, van den Broek M, de la Torre Cortés P, ter Horst J, Pronk JT, Daran J-MG (2018) **Allele-specific genome editing using CRISPR-Cas9 is associated with loss of heterozygosity in diploid yeast**, *Nucleic acids research* 47(3):1362-1372.

Brouwers N\*, Gorter de Vries AR\*, van den Broek M, Weening SM, Elink Schuurman TD, Kuijpers NGA, Pronk JT, Daran JM-G (2019) ***In vivo* recombination of *Saccharomyces eubayanus* maltose-transporter genes yields a chimeric transporter that enables maltotriose fermentation**, *PLoS Genetics* 15(4):e1007853.

Gorter de Vries AR, Voskamp MA, van Aalst ACA, Kristensen LH, Jansen L, van den Broek M, Salazar AN, Brouwers N, Abeel T, Pronk JT, Daran J-MG (2019) **Laboratory evolution of a *Saccharomyces cerevisiae* x *S. eubayanus* hybrid under simulated lager-brewing conditions**, *Frontiers in Genetics* 10:24.

Gorter de Vries AR\*, Koster CC\*, Weening SM, Luttik MAH, Kuijpers NGA, Geertman J-MA, Pronk JT, Daran J-MG (2019) **Phenotype-independent isolation of interspecies *Saccharomyces* hybrids by dual-dye fluorescent staining and fluorescence-activated cell sorting**, *Frontiers in Microbiology* 10:871.

\* These authors contributed equally.

Salazar AN\*, Gorter de Vries AR\*, van den Broek M, Brouwers N, de la Torre P, Kuijpers NGA, Daran JMG, Abeel T (2019) **Nanopore sequencing and comparative genome analysis confirm lager-brewing yeasts originated from a single hybridization**, bioRxiv 603480.

Gorter de Vries AR, Pronk JT, Daran J-MG **Lager brewing yeasts in the era of modern genetics**, manuscript in preparation.

Brouwers N, Brickwedde A, Gorter de Vries AR, van den Broek M, Weening SM, van den Eijnden L, Diderich JA, Bai F-Y, Pronk JT, Daran J-MG **Maltotriose consumption in hybrid *Saccharomyces pastorianus* is heterotic and results from regulatory cross-talk between parental sub-genomes**, manuscript in preparation.

Dutch Patent Application N2020912

European Patent Application EP18213697.8

\* These authors contributed equally.



