THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Exploring lactose metabolism in *Sungouiella intermedia* for sustainable bioproduction

A pathway to new cell-factory design

KAMESHWARA V.R. PERI

Department of Life Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2024

Exploring lactose metabolism in *Sungouiella intermedia* for sustainable bioproduction
A pathway to new cell factory design
KAMESHWARA V.R. PERI
ISBN 978-91-8103-103-4

© KAMESHWARA VENKATA RAMANA PERI, 2024.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5561
ISSN 0346-718X

Division of Industrial Biotechnology
Department of Life Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover: “*The yeast we might not want, but the one we need.*”
Artwork by: Rawdha Al-Amri

Printed by Chalmers Digitaltryck,
Gothenburg, Sweden 2024

“The world only makes sense if you force it to.”

-Batman: The Dark Knight. Volume 1 #4

Preface

This dissertation partially fulfils the requirements for obtaining the degree of Doctor of Philosophy at the Department of Life Sciences at Chalmers University of Technology. The work was supported by FORMAS grant no. 2017-01417 awarded to Associate Professor Cecilia Geijer. The PhD studies were carried out between May 2020 and November 2024 under the supervision of Associate Professor Cecilia Geijer and the co-supervision of Professor Lisbeth Olsson and examination by Professor Carl Johan Franzén.

Most of the work in this thesis was carried out at the Division of Industrial Biotechnology at the Department of Life Sciences at Chalmers University of Technology. CRISPR-Cas9 gene editing method development was initiated in collaboration with the lab of Professor Uffe Mortensen from the Department of Biotechnology and Biomedicine at Denmark Technical University. Genome-scale metabolic model reconstruction was performed in collaboration with Iván Domenzain, PhD from the Division of Systems and Synthetic Biology at the Department of Life Sciences at Chalmers University of Technology.

Kameshwara V.R. Peri,
November 2024

Exploring lactose metabolism in *Sungouiella intermedia* for sustainable bioproduction
A pathway to new cell-factory design

Kameshwara V. R. Peri
Division of Industrial Biotechnology – Department of Life Sciences
Chalmers University of Technology

Abstract

The world around us is rich in diversity, offering resources that have enabled humans to build thriving civilizations. However, the ever-increasing demand for finite resources has led to their overuse, causing ecosystem imbalances, accelerating climate change and destabilizing societies. To tackle the challenges of our rapidly changing world, technological innovations that integrate sustainable production practices are necessary. Microbial fermentation technology has greatly benefited from innovations such as integration of new microorganisms as cell-factories and industrial side-streams as alternate raw materials into existing manufacturing processes, thereby reducing waste and enhancing the efficiency of bioproduction.

In this thesis potential microbial cell-factories on lactose-rich substrates have been explored, starting from bioprospecting and profiling of different yeast species found in the tropical niche of Nigeria, to characterizing a superior lactose grower, *Sungouiella intermedia* (formerly *Candida intermedia*). First, development of the first ever genome-editing toolbox for *S. intermedia* using split-marker and marker-less CRISPR-Cas9 based techniques enabled in-depth characterization of its lactose and galactose metabolism. Second, genomics, transcriptomics and deletion mutants' phenotyping were leveraged to elucidate novel transcriptional and regulatory machinery orchestrated by a unique gene cluster in this yeast. Third, simulations from a reconstructed genome-scale metabolic model for *S. intermedia* revealed an alternate galactose metabolic pathway called the oxidoreductive pathway enabling overflow metabolism of galactitol to attain redox homeostasis. Finally, cell-factory applications for *S. intermedia* were demonstrated by the production of galactitol and its derivative, the natural sweetener, tagatose, on lactose-rich industrial side-stream cheese whey permeate.

In summary, this thesis presents the non-conventional yeast, *S. intermedia*, as a promising cell-factory, highlighting the tools and strategies used to explore its intrinsic metabolic capabilities. These insights pave the way for biotechnological innovations and the advancement of sustainable bioprocesses.

Keywords: Bioprospecting, *Sungouiella intermedia*, gene-editing toolbox, lactose metabolism, genome-scale metabolic model, cell-factory, industrial side-stream

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Karl Persson, Vanessa Oneyma, Princess Nwafor, Kameshwara V R Peri, Chioma Amadi, Jonas Warringer, Cecilia Geijer. (2024). "Lactose-assimilating yeasts with high fatty acid accumulation uncovered by untargeted bioprospecting". *(Submitted)*

II. Kameshwara V R Peri, Fábio Faria-Oliveira, Adam Larsson, Alexander Plovie, Nicolas Papon, Cecilia Geijer. (2023). "Split-marker-mediated genome editing improves homologous recombination frequency in the CTG clade yeast *Candida intermedia*." FEMS Yeast Res. 2023 Jan 4;23:foad016. DOI: 10.1093/femsyr/foad016.

III. Hanna D Alalam, Kameshwara V R Peri, Marta Parmigiani, Tomas Strucko, Uffe Mortensen, Cecilia Geijer. (2024). "A Highly Efficient CRISPR/Cas9 Toolbox for *Sungouiella intermedia*." *(Manuscript)*

IV. Kameshwara V R Peri, Le Yuan, Fábio Faria Oliveira, Karl Persson, Hanna D Alalam, Lisbeth Olsson, Johan Larsbrink, Eduard J Kerkhoven, Cecilia Geijer. (2024). "A unique metabolic gene cluster regulates lactose and galactose metabolism in the yeast *Candida intermedia*". Appl Environ Microbiol 0:e01135-24. DOI: 10.1128/aem.01135-24

V. Kameshwara V R Peri*, Iván Domenzain*, Hanna D. Alalam, Abril Valverde Rascon, Jens Nielsen, Cecilia Geijer. (2024). "Model-driven elucidation of lactose and galactose metabolism via oxidoreductive pathway in *Sungouiella intermedia* for cell-factory application*." (Manuscript)*

*The authors contributed equally

Contribution Summary

I. I contributed to experimental work for growth and metabolite characterization of lactose growing yeasts on cheese whey using HPLC and the writing and editing of the manuscript.

II. I planned and performed the experimental work. I analyzed the data, interpreted the results, and wrote the manuscript. I edited the manuscript together with other authors.

III. I contributed to experimental work involving initial assessment of replicating plasmids and tested CRISPR for integration of gene constructs. I also contributed to the writing and editing of the manuscript.

IV. I planned and performed the experimental work. I analyzed the data, interpreted the results, and wrote the manuscript. I edited the manuscript together with other authors.

V. I planned and executed the experimental work and contributed to the model reconstruction work led by Iván Domenzain. I performed the data analysis, interpretation of the results and manuscript writing together with Iván Domenzain. I edited the manuscript together with other authors.

Table of Contents

Preface V

Abstract VII

List of Publications VIII

Contribution Summary IX

Table of Contents X

Abbreviations XII

Chapter 1. Introduction 1

1.1. Aim of the thesis 3

Chapter 2. Bioprospecting for yeast discovery 7

2.1 Yeasts and their ecological niche 7

2.2 Bioprospecting yeasts in Nigeria – an untargeted approach 9

2.2.1 High-throughput screening discovers lactose growing yeasts 9

2.2.2 Identification of lactose growing yeasts 10

2.2.3 Lactose growers – future cell factories? 13

Chapter 3. Tools to explore a non-conventional yeast 15

3.1 Exploring a non-conventional yeast 15

3.2. Genome editing tools for *S. intermedia* 16

3.2.1. Testing transformation protocol & identification of functional selection marker 18

3.2.2. Improvement of targeting efficiency 19

3.2.4. Marker recycling to enable multiple deletions 20

3.3 System level understanding using Genome Scale Metabolic Models 25

3.3.1 Reconstruction of *Sint*GEM for *S. intermedia* 25

Chapter 4. Lactose metabolism in *S. intermedia* 29

4.1. Abundance of lactose-growing yeasts 29

4.2 Genetic determinants of lactose metabolism and regulation in yeasts 31

4.2.1 A unique gene cluster partakes in lactose and galactose metabolism in *S. intermedia* ... 33

4.2.2 Elucidation of a novel lactose and galactose regulation 33

4.3 An alternate pathway for galactose metabolism 38

4.3.1 *Sint*GEM-mediated exploration of the oxidoreductive pathway 40

4.3.2 Complex regulatory network and pathway redundancy 42

4.3.3 Role of dual-cofactor specificity of the Aldose reductase Xyl1_2 43

4.3.4 Elucidation of galactitol overflow metabolism in *S. intermedia* 44

Chapter 5. Cell-factory application of *S. intermedia* 47

5.1. Cell-factory design for sustainable bioproduction 47

5.1.1 Microbial production of galactitol and tagatose 48

5.1.2 Galactitol-producing chassis strain 48

5.1.3 Tagatose production strain 49

5.2. Process improvement for efficient bioproduction 50

5.2.1 Titer 51

5.2.2 Rate 51

5.2.3 Yield 52

5.3. Production of galactitol and tagatose using industrial side-stream as raw material . 54

Chapter 6. Conclusions and outlook 57

6.1. Conclusions 57

6.2. Outlook 61

Acknowledgements 65

References 67

Abbreviations

NCY	Non-conventional yeast
HTP	High throughput
CBS	Centraalbureau voor Schimmelcultures
DSB	Double-strand break
NHEJ	Non-homologous end joining
HR	Homologous recombination
MMEJ	Micro-homology mediated end-joining
ARS	Autonomously replicating sequence
GEM	Genome-scale metabolic model
FBA	Flux balance analysis
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamie adenine dinucleotide phosphate
CW	Cheese whey
CWP	Cheese whey permeate
DLP	De-lactosed permeate

Chapter 1. Introduction

The industrial revolution marked a shift from agrarian societies to manufacturing-driven economies, enabling faster production, diverse industries, and improved living standards. However, this rapid advancement, fueled by heavy reliance on finite natural resources, has led to ecological imbalance, pollution, and climate change, posing threats to human existence [1]. Throughout history, humanity has relied on Earth's biodiversity, from forests providing timber to grasslands and aquatic ecosystems supplying food and other natural products, while also depending on crucial microscopic actors, invisible to the naked eye but essential to our existence [2]. The depletion of many of these critical resources necessitates a re-evaluation of how we use and manage natural systems to ensure their sustainability for future generations [3].

Microorganisms play a vital role in improving human well-being and advancing our societies, from supporting health (e.g. gut [4] and skin microbiota, combating pathogens [5]), to their use in fermentation processes to produce food [6] and medicines [7]. Industrial fermentation utilizes bacteria, filamentous fungi, and yeasts as “cell-factories” to produce valuable chemicals. Bacteria like *Escherichia*, *Clostridium*, and *Pseudomonas* are used to produce metabolites like lactic acid, butanol, and ethanol but often face challenges like phage infections and intolerance to high metabolite concentration or pH change [8, 9]. Filamentous fungi, while capable of growing on complex substrates and secreting valuable hydrolytic enzymes, are limited by the availability of molecular tools and difficulties in metabolomic analysis [10, 11]. In contrast, yeasts offer several advantages, including high metabolic rates, genetic tractability, and tolerance to low pH, high temperatures, and organic solvents [12-14]. These traits combined with available tools for gene editing and metabolomic analysis, make yeasts ideal for applications in food, pharmaceuticals, fine chemicals, and biofuels production [6, 7].

Among yeasts used in bioproduction, *Saccharomyces cerevisiae* (baker's yeast) is the most widely applied, and serves as the primary workhorse of industrial-scale fermentations [6, 15]. With over 1,830,000 articles (according to Google Scholar as of October 2024), the vast research on *S. cerevisiae* highlights its importance both as a model organism and as a cell factory. *S. cerevisiae* has contributed to studies on cell-cycle regulation, autophagy and eukaryotic transcription [6]. In addition to the traditional food and beverages like beer, wine and bread [6], products derived from *S. cerevisiae* include, but are not limited to: insulin [16]; human serum albumin [17], succinic acid

[18]; ethanol [13], n-butanol [19], anti-cancer compound vinblastine [20] and the anti-malarial drug precursor artemisinic acid [21]. However, *S. cerevisiae* has limitations in certain applications, such as its inability to metabolize sugars like xylose and lactose naturally, and its preference for ethanol production, which often hampers the synthesis of non-ethanol products at high titers [9, 15, 22]. As a result, there is growing interest in exploring alternative yeasts that are better suited for specific applications. This has led to increased focus on "non-conventional yeasts" (NCYs), which exhibit unique metabolic traits and can offer solutions to some of the challenges associated with *S. cerevisiae* [15].

Compared to *S. cerevisiae*, NCYs are generally understudied and have limited or no genetic engineering tools. Therefore, they are not widely applied in industrial settings. Nevertheless, many NCYs have gained attention as emerging model organisms due to their unique metabolic traits. For example, *Yarrowia lipolytica* and *Schizosaccharomyces pombe* are valuable for research on diseases like obesity and Alzheimer's [23-25] while others, such as *Rhodotorula toruloides* and *Lipomyces starkeyi*, are notable for lipid production, and *Debaryomyces hansenii* for its halotolerance and osmotolerance [26, 27]. Some NCYs can also grow on various industrial side-streams, such as xylose-rich lignocellulosic biomass (*Blastobotrys mokoenaii* and *Scheffersomyces stipitis*), and lactose-rich cheese whey (CW) (*Kluyveromyces lactis* and *K. marxianus*), making them promising candidates for industrial fermentation applications [28, 29]. Many of these favorable traits are attributed to their ecological niches; for instance, *S. stipitis* live in wood-consuming beetles' guts, utilizing xylose and xylan, while *D. hansenii* thrives in high-salt environments [29]. Therefore, extensive exploration of their natural diversity and the development of tools to elucidate the metabolic versatility of NCYs are essential to enable their broader use both as model organisms and cell factories [3].

Since the industrial-scale aerobic fermentation for the production of citric acid in 1919 [7], the range of products produced through microbial fermentation has greatly expanded [3, 30-36]. Innovations in this field have made the technology resistant to geographic, climate and supply-chain instabilities and diverse in product range. Nevertheless, fermentation-based production is still often outcompeted by fossil carbon-based chemical production due to high production costs and long lead times for product development and return of investments [3, 7]. A key aspect of enhancing the sustainability of bioproduction processes is the use of low-cost substrates. Thus, the field has focused on transitioning from expensive pure sugars to cheaper but more complex industrial or domestic side/waste-streams [37]. For instance, lignocellulosic biomass from the paper and pulp industry, which is rich in xylose and glucose, and lactose-rich cheese whey (CW) from the dairy industry, are promising raw materials for microbial bioproduction [38, 39]. A by-product of cheese and yogurt production, CW is produced in vast quantities (0.9 kg of whey is generated from every

1 kg of milk; about 55 million tonnes CW produced in 2022 in the EU), and presents an environmental problem [40, 41]. Being rich in nutrients and containing 4-5% lactose, CW has a high biological and chemical oxygen demand, making its treatment expensive [27, 42-44]. While some dairy companies utilize CW in food products, many smaller companies lack the resources to valorize it [45]. Due to its abundance and low cost, whey is a suitable raw material for microbial bioproduction of value-added products using lactose-metabolizing yeasts [37, 44].

Whereas *S. cerevisiae* does not naturally metabolize lactose, the dairy yeasts *K. lactis* and *K. marxianus* are well-characterized in terms of lactose metabolism and are used in various cell-factory applications with lactose-rich substrates for producing biomass, ethanol, and both endogenous and heterologous enzymes [15, 41]. Recent studies involving *D. hansenii* provide another example of a NCY with potential as a cell factory that can utilize whey as a substrate for recombinant protein production [27, 46]. Further diversification of products using dairy side-streams could be achieved by exploring the potential of other lactose-metabolizing yeasts. However, lactose metabolism is relatively uncommon among yeasts, and most of these lactose-utilizing species remain understudied, with limited genetic and physiological insights available on their lactose metabolism [47]. Thus, genomic and transcriptomic exploration of these yeasts will be essential to discover and develop future cell-factories with diverse product portfolios and well suited for industrial applications.

1.1. Aim of the thesis

The overarching aim of this thesis was to investigate the potential of NCYs, for sustainable bioproduction using lactose-rich side streams. Specifically, the study aimed to explore the genetic and metabolic factors involved in lactose and galactose metabolism in the NCY *Sungouiella intermedia* (formerly *Candida intermedia*) and evaluate its suitability as a cell factory.

The following research questions were formulated and addressed in this thesis:

- *Which lactose-metabolizing yeasts, with traits suitable for cell-factory application can be discovered from untargeted bioprospecting?*

To discover new yeast strains and species, untargeted bioprospecting of yeasts was performed in tropical West Africa as described in **Chapter 2**. A high-throughput (HTP) screening platform was used to assess 1996 yeast isolates in 70 different environments and a subsequent focus was on characterizing the identified lactose-assimilating yeasts. These strains were phylogenetically identified by sequencing of the Internal transcribed spacer (ITS) genomic regions, and their further characterization was performed on lab media and industrial side-stream CW. This revealed several strains with interesting growth phenotypes and cell-factory potential (**Paper I**).

- *Having discovered a new strain with interesting properties, what tools must be developed to enable genetic and metabolic characterization of a NCY like the lactose-metabolizing* S. intermedia*?*

To enhance our ability to study the NCY *S. intermedia*, genome editing tools including split-marker (**Paper II**) and CRISPR-Cas9 (**Paper III**) techniques, specifically tailored for this yeast, were developed (explored in **Chapter 3**). Computational biology tools, reinforced with experimental characterization, were used to develop the first-ever genome-scale metabolic model for *S. intermedia,* named *Sint*GEM*,* enabling further insights into the lactose and galactose metabolism of *S. intermedia* (**Paper V**).

- *What are the genetic and metabolic determinants of lactose metabolism in* S. intermedia?

The ability to construct gene deletion mutants of *S. intermedia* and phenotyping them, along with transcriptomic and bioinformatic analysis, led to an in-depth understanding of the genetic determinants of lactose metabolism in this yeast (**Chapter 4)**. In particular, the discovery and characterization of a unique and novel gene cluster encoding important regulators of lactose and galactose metabolism provided new insights into yeast regulatory networks and metabolic

pathways (**Paper IV**). Additionally, a knowledge-matching approach, combining experimental and computational analysis, was used to understand overflow metabolism in *S. intermedia* (**Paper V**).

- *How can the knowledge gained for* S. intermedia *enable valorisation of lactose-rich industrial side-streams to produce added-value compounds?*

In **Chapter 5**, the knowledge gained from the extensive characterization of *S. intermedia* was used to explore possible cell-factory applications for this yeast on lactose-rich industrial side-streams. Moreover, utilizing genetic tools developed for this yeast and guided by computational predictions, metabolic engineering and bioprocess optimization were carried out to improve the production of the polyol galactitol and its derivative tagatose using cheese whey permeate (CWP) as the feedstock (**Paper V**).

Chapter 6 concludes on the key findings and their implications, while also exploring potential future directions for research and applications based on the results.

Chapter 2. Bioprospecting for yeast discovery

This chapter describes the discovery of novel yeast species through untargeted bioprospecting in the tropical West African region. The initial part of the chapter focuses on bioprospecting to leverage the diverse and vast microbial biodiversity. The use of HTP screening to characterize new yeast species with a particular focus on yeasts capable of metabolizing lactose and subsequently, their identification, is described. Final part of the chapter covers the assessment of the potential of strains isolated from West Africa to produce industrially relevant products such as bioethanol, organic acid or lipids on industrial side-stream, cheese whey.

2.1 Yeasts and their ecological niche

Yeasts display a diverse array of heterotrophic metabolic capabilities, enabling them to effectively utilize various nutrients and flourish in ecosystems spanning all continents, as well as in major aquatic and terrestrial biomes [47]. As shown by Groenewald and colleagues the wide yeast diversity is reflected in various culture collections, which show that yeasts have been isolated from an array of distinct environments, highlighting their adaptability to different ecological niches such as fruit surfaces, plant exudates, soil, insects, rotting wood, and tree bark (Fig. 1A) [48]. The same study gives an overview of the continents of origin (data collected from 200 countries) of these strains preserved in the participating collections (Fig. 1B). The least number of strains were in the culture collections from the continents Africa and Australia (including New Zealand). Additionally, while many collections from a country/region preserve mostly strains from that region, large European collections, such as the CBS yeast strain collection from The Westerdijk Fungal Biodiversity Institute in The Netherlands, have an evenly spread geographic distribution of strains [48].

Over the last 20 years, nearly half of newly described species have been sampled in Asia (480 species). Europe and North America contributed 227 and 208 species, while species discovered in South America amounted to 135. Along with Australasia, Africa and Antarctica were among the most under-sampled places for new yeasts discovered since 2001 (Fig. 1C) [49]. These findings

suggest that bioprospecting in Africa could significantly expand the yeast repository and uncover valuable traits for industrial use.

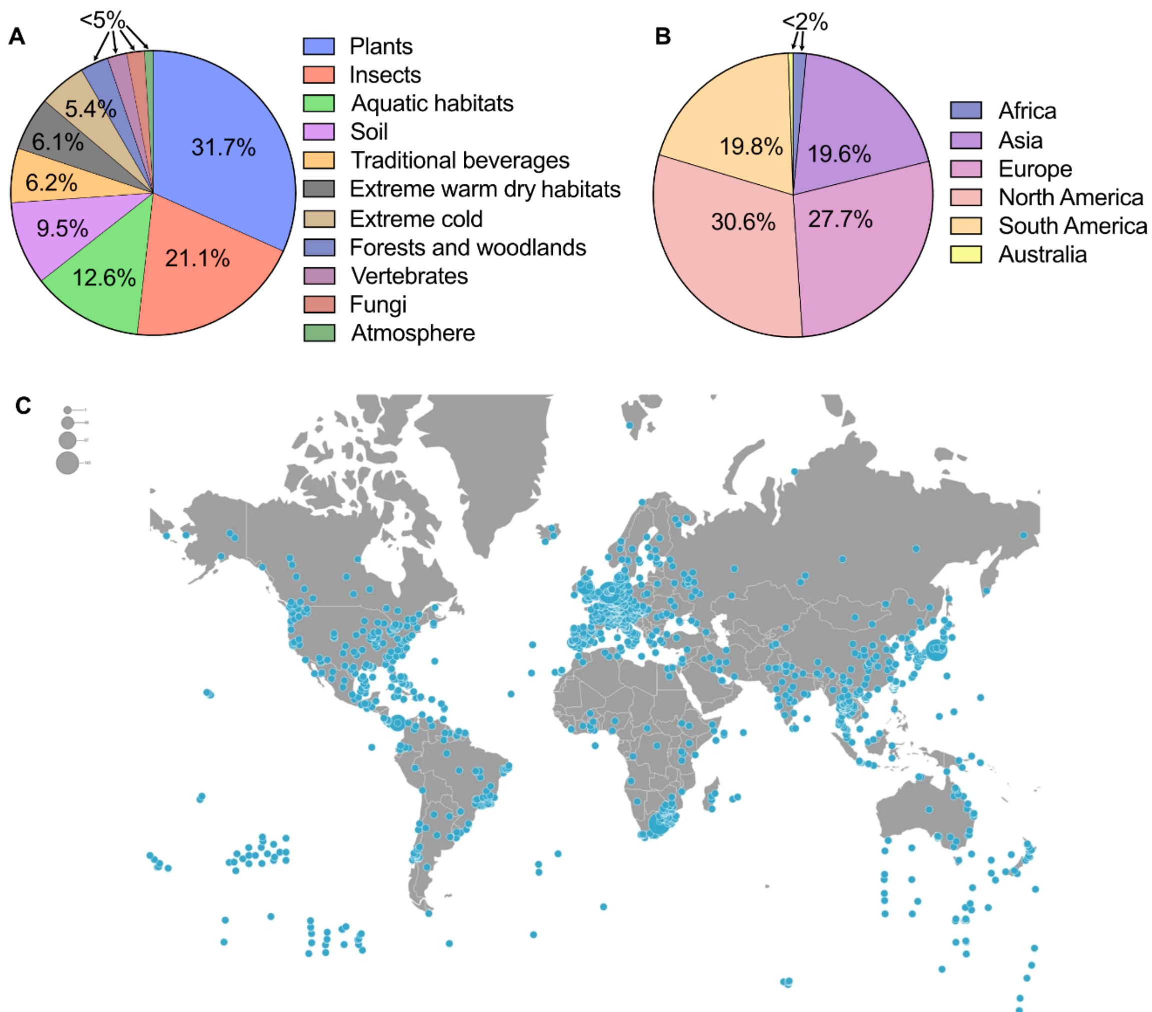

Figure 1. **Yeasts and their ecological niches.** Curators of 41 collections from 27 countries supplied data representing a total of 58,095 yeast strains. A) Pie chart depicting different ecological niches or sources from where yeasts have been isolated. The number of isolates from each niche/source is depicted in percentage and comprises of the total number of niches recorded in this study. B) Pie chart depicting abundance of yeasts isolated from different continents. The number of isolates from each continent is depicted in percentage as recorded in this study. Data adapted from [48] under CC BY 4.0 license. C) Global map showing the origin of yeast strains from the public CBS collection (Westerdijk Fungal Biodiversity Institute, The Netherlands). Figure taken from [49] under CC BY 4.0 license.

2.2 Bioprospecting yeasts in Nigeria – an untargeted approach

The objective of bioprospecting is to isolate yeasts from various biomes, and evaluate their potential for biotechnological applications [50]. Bioprospecting can be untargeted where yeasts may be isolated from diverse environments without a predefined target or specific application in mind. Untargeted bioprospecting can offer a broad assessment of microbial diversity within an ecosystem and increase the likelihood of discovering species with unique traits [51]. For instance, studies assessing the yeast biodiversity of mangrove ecosystems have identified diverse species, including *Candida*, *Geotrichum*, and *Kluyveromyces*, highlighting their potential in the production of enzymes, xylitol, biofuels, and single-cell oils, among other applications [52]. Conversely, targeted bioprospecting focuses on isolating microbes adapted to specific environmental conditions, often of industrial relevance, such as high temperatures, salinity, or low water availability [53, 54]. For instance, studies aimed at identifying new yeasts for beer production screened for strains from high-sugar niches, such as flower nectar and cacao fermentations, to select those capable of growing in high sugar concentrations, with potential applications in high-gravity beer fermentations [55].

In the study described in **Paper I**, an untargeted approach was employed to isolate yeasts capable of growing on non-selective media, thereby enabling the cultivation of a wide range of yeast species. As mentioned in the previous section, many regions in Africa remain understudied in terms of yeast biodiversity, leaving much of the continent largely unexplored [56]. To address this, the first country-wide yeast bioprospecting campaign in Nigeria was conducted by researchers and students from the University of Nigeria, and the collection was then characterized at the University of Gothenburg and Chalmers. Nigeria was chosen due to its underexplored tropical regions, characterized by diverse ecosystems and a tropical climate.

2.2.1 High-throughput screening discovers lactose growing yeasts

HTP screening technologies have been extensively applied in diverse research areas, including the screening for improved production strains in the food industry and the rapid testing of compounds in the pharmaceutical industry to identify biological activity for drug discovery [57, 58]. Bioprospecting can significantly benefit from HTP screening methods, such as HTP robotics [54] and ultra-HTP droplet-based microfluidic systems [59] which accelerate the phenotyping of novel strains by enabling the simultaneous handling of a large number of isolates at once. Additionally,

some of these systems allow for precise colony imaging and accurate estimation of colony size and growth rate, thereby enhancing the quality of screening protocols.

In this study, the “Scan-o-matic” HTP screening system was used to screen yeasts (1,536 colony arrays per agar plate) based on growth parameters such as colony size, growth rate and yield under varying environmental or experimental conditions [60]. A total of 1996 isolates were characterized on 70 different environments, including one with lactose as carbon source. On lactose, 27.4% of the isolates exhibited rapid growth (doubling time of <3h). This finding was particularly noteworthy given that lactose-assimilation is rare among yeasts, and the bioprospecting approach was untargeted and did not include lactose-rich niches such as dairy farms or dairy products. (**Paper I: Fig. 2D**) [47]. Thus, even in niches which are not rich in lactose, yeasts may possess the genetic machinery necessary for the assimilation of this carbon source, underlining the importance of untargeted bioprospecting for finding strains of significant industrial interest.

2.2.2 Identification of lactose growing yeasts

203 lactose-growing isolates were identified in the Nigerian strain collection. The isolates were assigned to respective genera and species using PCR for amplification of the ribosomal Internal Transcribed Spacer (ITS) region, which is the most used genetic marker for the molecular identification of fungi, especially in environmental sequencing and molecular ecology studies [61, 62]. Comparison of the resulting ITS sequences with sequenced type yeasts in the NCBI BLAST database identified 30 distinct lactose-utilizing yeast species from the *Ascomycota* and *Basidiomycota* phyla (**Paper I: Table 1**). Although the majority of isolates passed the cut-off of 98.41% ITS sequence similarity [62], 48 isolates were below the threshold, which included strains assigned to *Sungouiella intermedia*, *Candida pseudocylindracea*, *Candida tropicalis*, *Kurtzmaniella quercitrusa*, *Moesziomyces antarcticus*, *Papiliotrema laurentii*, *Pichia kudriavzevii*, *Rhodotorula mucilaginosa*, *Ustilago sparsa* and *Yarrowia galli*, indicating that they may represent potential novel species.

The results from ITS sequencing of the lactose growing yeasts identified *S. pseudointermedia* (57 isolates) and its sister species *S. intermedia* (19) as two of the three most predominant lactose-growing species in the collection (Fig. 2A). Given the diverse geographic and source origins of the isolates (**Paper I: Table 1**), it was inferred that these isolates are not just genetic clones, and that lactose growth is a common trait in these species. Previous works on strain isolation, where isolates

were identified belonging to the *Candida* genus, showed that *S.* (formerly *Candida*) *intermedia* was in fact one of the most abundant species found in diverse niches [63]. In fact, many *S. intermedia* strains have been isolated from diverse environments such as alcoholic and citric acid fermentation substrates, soil, water sources both sea and river, fruits, as well as food products such as orange juice and yoghurt [64]. Additionally, a strain belonging to this species was identified as a contaminant in an adaptive laboratory evolution experiment of *S. cerevisiae* growing on xylose-rich lignocellulosic hydrolysates in our lab over a decade ago. Since then, this strain, CBS 141442, has been extensively studied for its xylose metabolism and, its lactose metabolism has been studied in **Paper II**-**V** of this thesis [64-68].

Furthermore, phylogenetic analysis using the neighbor-joining method was performed for the isolates belonging to *S. intermedia* and *S. pseudointermedia* (as per ITS sequencing). Our lab strain CBS 141442, the type-strain *S. intermedia* CBS 572 and additional *S. intermedia* strains from the CBS culture collection were also included in the analysis, as well as the *S. pseudointermedia* type strain CBS 6918 (Fig. 3B). Interestingly, the *S. intermedia* isolates appeared to be distributed across this branch of the phylogenetic tree. In the case of the *S. pseudointermedia* (57 isolates in total), 39 isolates seemed to be closely related to the type-strain CBS 6918, while the other 18 were closer to different *S. intermedia* strains. These results highlight the limitations of using ITS sequences alone for naming species (and phylogenetic analyses such as this one). However, several of the strains from both the CBS and Nigerian collections may also be misidentified and could potentially represent new species. As new species are continuously being identified, reoccurring classification efforts are essential to ensure a clear and updated understanding of yeast diversity and taxonomy. For this thesis, a very appropriate example of such work is the recent reclassification of species within the polyphyletic genus *Candida*, which reassigned *S. intermedia* and *S. pseudointermedia* to the newly established genus *Sungouiella* [69].

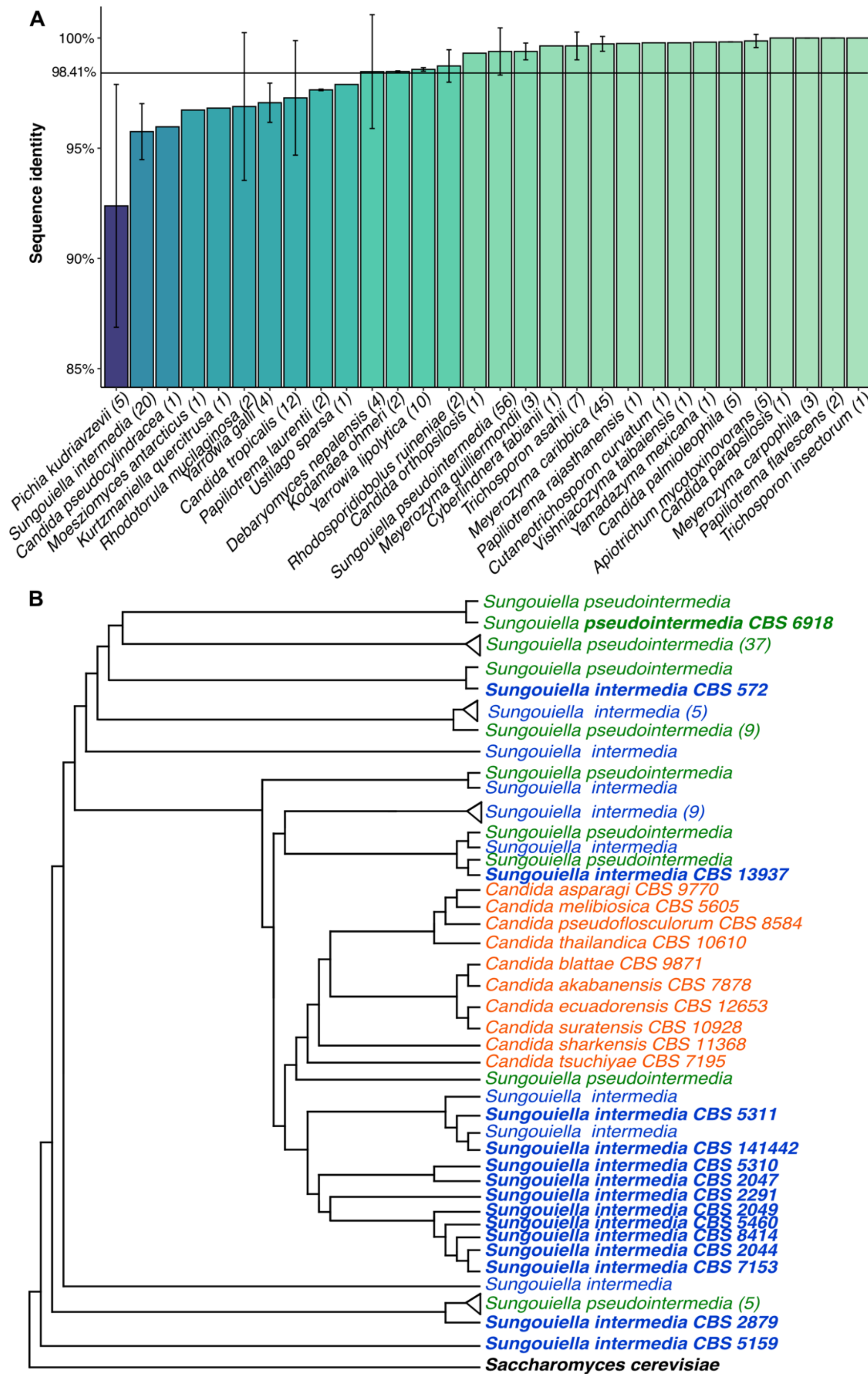

Figure 2. Identification of lactose-growing yeasts derived from the untargeted bioprospecting of Nigeria. A) Bar graph showing the results from ITS sequencing for lactose-growing yeasts. Yeast names as derived from Blast comparison are depicted on x-axis and their respective sequence identity on y-axis. The cut-off value of 98.41% (shown by solid line) was used to classify the isolates to respective yeast species. Data adapted from **Paper I**. B) Phylogenetic tree (neighbor-joining analysis) for isolates belonging to *S. intermedia* (strains in blue; type strain (bold) – *S. intermedia* CBS 572) and *S. pseudointermedia* (strains in green; type-strain (bold)- *S. pseudointermedia* CBS 6918) as per ITS sequencing. Closely related species are represented in red and the outgroup represented by *S. cerevisiae* in black.

2.2.3 Lactose growers – future cell factories?

Evaluating yeast strains on industrially relevant substrates is a crucial step in assessing their potential as future cell factories. In this study, lactose-assimilating yeast strains were cultivated in whey using microtiter plates, followed by metabolite analysis via HPLC and, lipid production assessment in selected strains. The isolates exhibited diverse growth patterns and varying final yields, with 94.6% of the strains achieving an average doubling time of 2.17 hours in whey (**Paper I: Fig. 3A**). While the basidiomycetous yeasts *Moesziomyces antarcticus*, *Ustilago sparsa*, *Papiliotrema laurentii*, and *Papiliotrema rajasthanensis* nearly completely consumed the available lactose, ascomycetes such as *S. intermedia* and *S. pseudointermedia* displayed moderate lactose consumption, with significant variation across different strains (**Paper I: Fig. 3B**). This bioprospecting approach successfully identified both known and novel strains capable of growing in lactose-rich industrial side-streams, highlighting their potential for lactose remediation.

Metabolite profiling revealed that certain isolates, including *S. intermedia*, *Debaryomyces nepalensis*, *Meyerozyma caribbica*, and *Y. lipolytica*, produced ethanol in concentrations ranging from 0.5 to 2.3 g/L (**Paper I: Fig. 3C**). Additionally, strains such as *Trichosporon asahii*, *Trichosporon insectorum*, and *Cutaneotrichosporon curvatum* produced approximately 0.5 g/L of fumarate, a compound of significant interest to the food, pharmaceutical, and chemical industries due to its potential for conversion into therapeutic drugs and its application in polymerization processes [70]. The study also identified lactose-assimilating yeasts capable of accumulating lipids constituting more than 20% of their dry cell weight, classifying them as oleaginous [71]. Of the 14 selected isolates, eight accumulated lipids exceeding 20%, with *Apiotrichum mycotoxinivorans* (40%), *Papiliotrema laurentii* (36%), *Moesziomyces antarcticus* (34%), and *Cutaneotrichosporon curvatum* (31%) demonstrating the highest lipid accumulation (**Paper I: Fig. 5B**). These strains exhibit strong potential as cell factories for lipid production using lactose as a carbon source.

As evidenced by HTP screening in the bioprospecting efforts, *S. intermedia* strains displayed rapid growth in lactose-containing media as well as cheese whey (**Paper I: Fig. 3A & 4A**). Given these promising results, we decided to investigate the genetic and metabolic basis of lactose metabolism in *S. intermedia*. However, as is the case with many NCYs, the lack of genetic and metabolic tools for *S. intermedia* posed a challenge. The development of such tools, which is covered in **Chapter 3** of this thesis, enabled deeper exploration of the strain's metabolic potential.

Key take-away point from Chapter 2:

- Bioprospecting in unexplored regions can help discover new yeast species with industrial potential.
- HTP technologies can enable screening of large number of yeast isolates enabling their discovery and characterization in a fast and efficient manner.
- ITS sequencing identified lactose-utilizing yeasts belonging to both known and potentially new species. Among the 30 lactose-metabolizing species identified, the most abundant were *S. intermedia* and its sister species *S. pseudointermedia*.
- Growth and metabolite analysis identified new yeast candidates for production of industrially relevant compounds like ethanol, fumarate, and lipids.

Chapter 3. Tools to explore a non-conventional yeast

The third chapter describes the development of tools that enabled both experimental and computational exploration of *S. intermedia*. The first part of the chapter describes the development of genome editing tools for this yeast starting with the split-marker based technique and followed by the marker-less CRISPR-Cas9 mediated approach. Thereafter, the development of the first curated genome-scale metabolic model (GEM) for *S. intermedia* is described.

3.1 Exploring a non-conventional yeast

Bioprospecting for yeasts offers insights into yeast biodiversity and the potential to discover NCYs that can convert sugars into industrially valuable metabolites [72]. But to fully unlock the potential of a NCY whether as a model organism or a cell factory—a detailed understanding of the molecular mechanisms underlying the trait of interest is necessary. This involves studying both genetic and metabolic factors that shape the desired phenotypes, thereby linking genotype to phenotype [9].

Recent advancements in genome sequencing have made it faster and more cost-effective to sequence yeast species and strains, which also expedites the development of genomic tools to explore their metabolic potential [9]. The combination of experimental and computational tools is particularly beneficial for studying understudied yeast species. Experimental tools for genome editing enables the perturbation of genetic networks, while omics technologies, such as transcriptomics, can reveal how these perturbations affect yeast response to genetic manipulations or environmental shifts. Computational tools such as genome-scale metabolic models (GEM) offer a systems-level view of cellular function, both as a knowledge database as well as a predictive tool to understand genotype-phenotype relations. This combined approach has significantly advanced the understanding of new and emerging yeast species, opening doors for innovative biotechnological applications [9, 14, 73].

The interesting properties of *S. intermedia* make it an excellent candidate for further exploration using genomic tools. Therefore, this chapter elaborates the work presented in **Papers**

II, **III**, and **V,** focusing on the development of genome editing (split-marker and CRISPR-Cas9) and computational tools (GEM).

3.2. Genome editing tools for *S. intermedia*

S. intermedia CBS141442 is a haploid yeast belonging to the *CTG* clade of ascomycetous yeasts, where the universal *CTG* codon is predominantly translated into a serine rather than a leucine [74]. The previously published *de novo* genome sequencing and annotation of the *S. intermedia* strain CBS 141442 [66], paved the way to develop genome editing tools for this yeast.

The development of genome editing tools for yeasts relies on innate DNA repair mechanisms that respond to DNA double-strand breaks (DSBs). DSBs, induced by radiation, toxic chemicals, or endogenous processes, can cause genome instability and cell death if unrepaired [75]. To repair DSBs, cells activate two major pathways: non-homologous end joining (NHEJ) and homologous recombination (HR) (Fig. 3). NHEJ directly reseals breaks without the need for extended homology and is active throughout the cell cycle. This error-prone pathway, facilitated by the Ku70/Ku80 heterodimer and proteins Nej1 (Non-homologous end-joining protein 1) and Dnl4-Lif1 (DNA ligase 4 and ligase-interacting factor 1), often results in chromosome translocations, gene amplifications, and transposable element movements [76]. NHEJ can also lead to random insertions or deletions, potentially disrupting open reading frames. Moreover, alternative NHEJ mechanisms exist that do not use ku70/80 proteins, including microhomology-mediated end joining (MMEJ) and single-stranded annealing (SSA), and they utilize short homologous sequences (1–4 bp) during repair [77, 78].

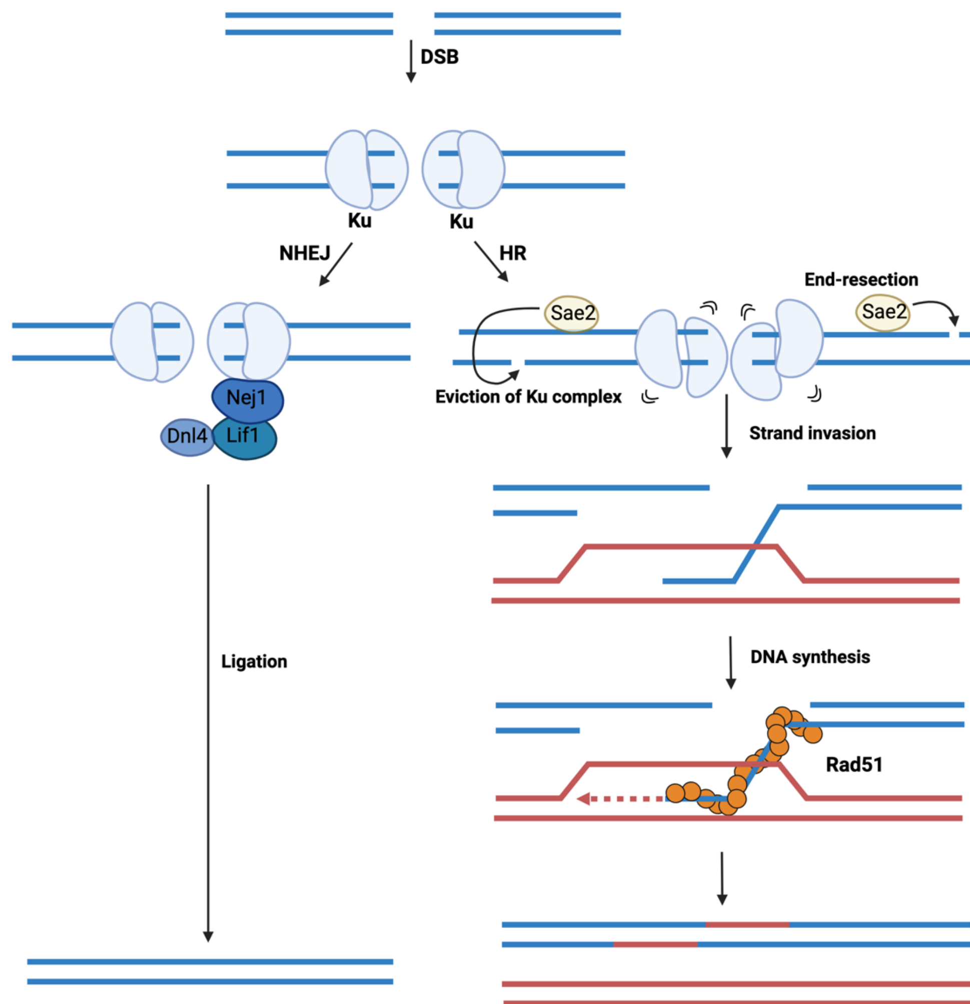

Figure 3: The conserved mechanisms of DSB repair by NHEJ and HR in budding yeast. On the left, Ku encircles the DSB ends and recruits the Lif1-Dnl4 and Nej1 ligase complex to facilitate NHEJ. On the right, HR is initiated by Sae2 and Ku mediated DNA resection, followed by action of Rad51 protein, which locates and invades a homologous donor sequence, initiating DNA repair synthesis. Figure adapted from [79] under CC BY 4.0 license.

In contrast, HR is a precise repair mechanism that ligates broken ends using homologous donor DNA from homologous chromosomes or sister chromatids [77]. In budding yeast, concerted action of Sae2 (Sporulation in the absence of Spo eleven 2) and Ku stimulate DNA resection for the Rad protein to bind. Then, Rad51 searches for a homologous template for repair resulting in a precise repair [79]. HR is mainly active during the S and G2 phases in the cell cycle, where sister chromatids are available. Factors such as the proximity of the donor and recipient DNA and, nuclear compartmentalization also contribute to the availability of donor DNA for HR [80]. Although both mechanisms co-exist, HR is the predominant repair mechanism in *S. cerevisiae* [81], while NHEJ is dominant in many NCYs [77, 82].

The development of a genome editing technique for *S. intermedia* was performed in a step-by-step manner.

1) testing transformation protocol & identification of functional selection marker

2) improvement of targeting efficiency

3) Marker recycling to enable multiple deletions

Each step has been separately explained in the following paragraphs.

3.2.1. Testing transformation protocol & identification of functional selection marker

Shen and colleagues demonstrated the successful expression of a *NAT1* gene from *Streptomyces noursei*, codon-optimized for *C. albicans*, in other *CTG* clade yeasts for selection on nourseothricin and encoding nourseothricin acetyltransferase [83]. Due to its close phylogenetic relationship with *Candida albicans* within the *CTG* clade, it was hypothesized that *S. intermedia* would display similar sensitivity to nourseothricin. As expected, *S. intermedia* showed sensitivity, indicating that the *Ca*NAT1 marker could function effectively as a selection marker in this yeast.

Several transformation protocols have been tested for *S. cerevisiae* and *C. albicans*, with electroporation and heat-shock/lithium-acetate (LiAc) being the most preferred [84]. Electroporation is believed to induce transient pores in the membrane, facilitating DNA entry, while the LiAc method enhances membrane permeability using PEG and a heat shock treatment to facilitate uptake of foreign DNA [85, 86]. In *C. albicans*, electroporation was found to be the most efficient transformation method as compared to heat-shock/LiAc and spheroplast. Likewise, when testing electroporation and heat-shock/LiAc methods in *S. intermedia*, electroporation yielded higher transformation efficiency, suggesting that *S. intermedia*'s membrane responds more effectively to electric pulses than to chemical treatments [87]. A high-efficient transformation result using electroporation suggested that *S. intermedia* could express the *CaNAT1* marker (**Paper II: Fig. 1**).

3.2.2. Improvement of targeting efficiency

Improving targeting efficiency can be achieved by increasing homologous recombination (HR)-mediated integration of donor DNA, rather than relying on the error-prone and ectopic non-homologous end joining (NHEJ) mechanism. One method involves adding long stretches (1–2 kb) of DNA homologous to the target site in the genome to the marker gene, enhancing HR at the integration site [88]. In addition, using chemical modifiers such as hydroxyurea to synchronize the cell cycle in the S/G2 phase during transformation can increase the likelihood of HR over NHEJ [89].

Results for transformation with a linearized DNA fragment containing *CaNAT1* cassette flanked on both sides by 1000 bp of homology to the *ADE2* locus in *S. intermedia*, and subsequently complementing with cell-cycle synchronization by hydroxyurea, confirmed that NHEJ pathway is the dominant pathway in *S. intermedia* for DSB repair, as most transformants (~96%) displayed random integration of the deletion cassette (**Paper II: Fig. 2D**).

3.2.2.1 Split-marker mediated genome editing

Another approach to increase targeting efficiency that has been successfully used in yeasts and filamentous fungi for targeted gene deletions is the split-marker technique [90]. It works by dividing the linear heterologous DNA in two parts, where the two fragments are designed to be homologous at the overlap of the marker and either upstream or downstream of the integration locus. Since the marker gene is split, it only becomes functional if homologous recombination (HR) between the overlapping fragments restores the complete marker (Fig. 4) [91]. Thus, only transformants in which the two halves recombine and integrate into the genome can grow on antibiotic selection plates. In *S. intermedia*, the split-marker technique resulted in significant increase in targeting efficiency to an average of 63.9% (**Paper II: Fig. 3C**), demonstrating that this method promotes homologous recombination, ensuring the proper integration of the full cassette at the target site.

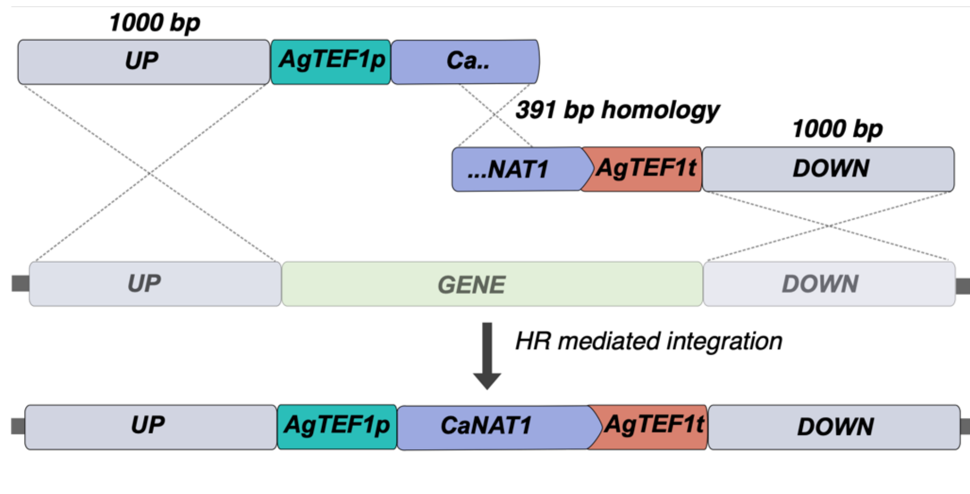

Figure 4. Illustration of the split-marker mediated genome editing in *S. intermedia*. Truncated fragments of a split-marker cassette are shown with homology of 391 bp at marker overlap. Each fragment contains half of the *CaNAT1* marker gene fused with 1000 bp of either up- or downstream homology to the target site. Recombination at the marker overlap site and at the homology arms to the up- or downstream sequence of the target gene, results in the integration of the complete fragment causing disruption or replacement of the target gene. Figure taken from **Paper II**.

Several factors can influence the targeting efficiency of genome editing, such as genomic loci, size of the target DNA and length of the homology arms [88]. The varying targeting efficiencies for three additional loci in *S. intermedia* indicated that the split-marker method shows locus (and its size) dependent variation, but the targeting efficiencies are significantly improved compared to those obtained with a linear cassette in both synchronized and non-synchronized cells (**Paper II: Fig. 3C**).

3.2.4. Marker recycling to enable multiple deletions

3.2.4.1 Split-marker with FLP recombinase

The split-marker method enabled construction of multiple single gene deletion mutants in *S. intermedia.* To enable multiple gene deletions within the same strain background, it was desired that the *CaNAT1* marker could be recycled and re-used. Thus, the plasmid pJK863 containing a *FLP* ('flip') gene, encoding a FLP-recombinase, was tested (Fig. 5) [92].

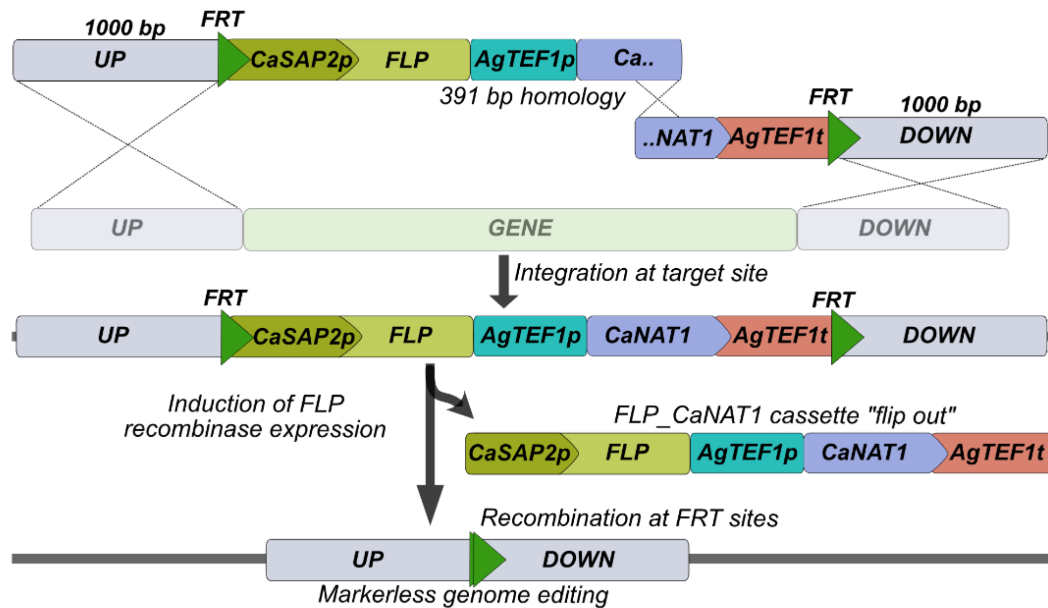

Figure 5. Split marker mediated genome editing in *S. intermedia* using the pJK863 plasmid. The cassette contains the dominant *CaNAT1* antibiotic selection marker which is split into two truncated fragments. Each fragment contains 1000 bp homology to the target site, either up- or downstream. As depicted, one of the fragments contains the *FLP* recombinase while both fragments contain the FRT site. HR results in the formation of a complete *caNAT1* marker and integration of the linear cassette at the target site. Subsequently, induction of the *FLP* recombinase results in the region between the FRT sites to “flip-out” due to recombination of the FRT sites, thus enabling marker-less genome editing. Figure taken from **Paper II**.

Split-marker transformation using the pJK863 plasmid resulted in low targeting efficiencies, approx. 10% and lower, for all the tested loci (**Paper II: Fig. 4B**). The reasons for the varying targeting efficiencies using the two integrative plasmids under the same experimental setups were not investigated in this thesis, but may be attributed to the size of the integration fragment as compared to the homology arms and the integration locus [88]. Additionally, the efficiency of flipping out the deletion cassettes was low, likely due to the fact that the *FLP* gene was expressed under a *C. albicans* promoter.

Characterization of model systems and development of chassis strains often require the deletion of more than one gene in the same strain background. However, using only one selection marker makes this process inefficient [9]. Therefore, the marker-free genome editing method Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) proteins (CRISPR-Cas9), was tested for advancing genome editing in this yeast.

3.2.4.2 Genome editing using CRISPR-Cas9 in S. intermedia

The CRISPR/Cas9 method originates from the discovery of an adaptive immune defense mechanism that exists in bacteria and archaea against invading viral particles. This system enables these organisms to target viral DNA, providing the foundation for its application in genome editing [93]. CRISPR-based genome editing has been applied in different organisms [94-97] and specifically, application of CRISPR-Cas9 in *S. cerevisiae* has accelerated the industrial workhorse's applications in the field of biotechnology [98]. Further development of CRISPR-Cas genome editing has enabled expansion to many other yeasts such as *S. pombe* [99], *P. pastoris* [100], *C. albicans* [101], *Y. lipolytica* [102], *S. bombicola* [103] to name a few.

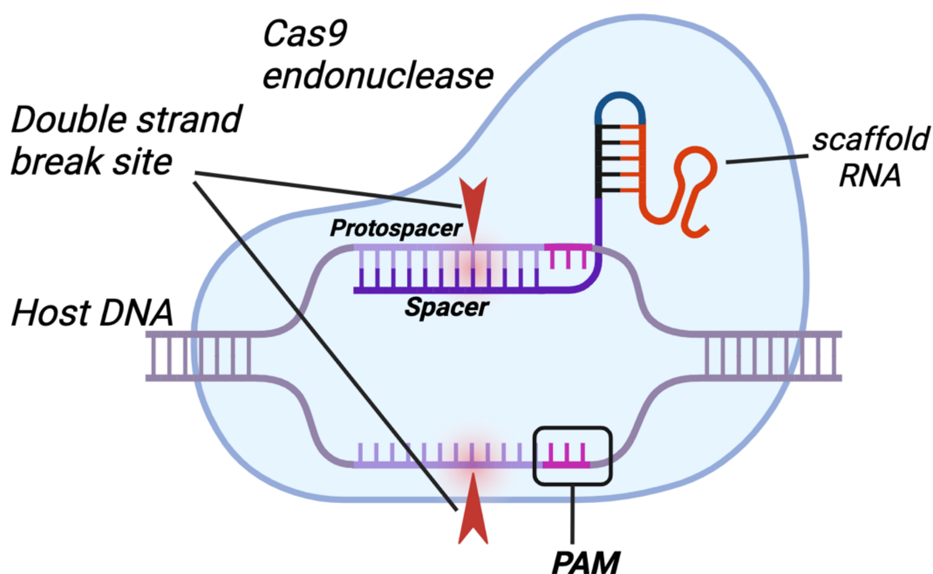

Figure 6. Schematic representation of the CRISPR/Cas9 mechanism in yeast. The single guide RNA (sgRNA) consists of two parts: the scaffold RNA (tracrRNA and the structural part of crRNA) and a 20-nucleotide spacer RNA. Once the Cas9 endonuclease associates with the sgRNA, it binds to the target site in the genomic DNA and undergoes a conformational change, cleaving both DNA strands (indicated by red triangles) 3-4 base pairs upstream of the PAM site. Following the introduction of a double-strand break (DSB), the preferred homology-directed repair (HDR) pathway repairs the break using donor DNA.

The CRISPR system relies on two key components: a CRISPR-associated (Cas) protein and a guide RNA (gRNA) (Fig. 6). The Cas endonuclease cuts DNA when directed by the gRNA to a specific locus. The most used Cas nuclease is Cas9 from *Streptococcus pyogenes*. The gRNA consists of two parts: scaffold RNA (a transactivating RNA (tracrRNA) and crispr RNA (crRNA)) which acts as a binding scaffold for Cas endonuclease and a 17-20 nucleotide spacer that is complementary to the target DNA. Cas9 recognizes its target with the help of a protospacer adjacent motif (PAM), a specific sequence (5'-NGG-3') on the non-target DNA strand. Cas nucleases from different species

recognize different PAM sequences. Cas9 cuts 3-4 nucleotides upstream of the PAM, generating a double-strand break (DSB), which is then repaired by the cell's DSB repair mechanisms [98].

The development of CRISPR-Cas9 based genome editing for *S. intermedia* (**Paper III**) was performed similar to many NCY such as *D. hansenii* [104], *K. phaffi* [105], *C. lusitaniae* [106], where Cas9 and gRNA are expressed from the same replicative plasmid. Expression from the same plasmid was selected with the intention of simplifying the cloning process. Additionally, the relative difficulty in 'flipping-out' the marker gene contributed to the decision of not integrating *CAS9* into the genome. While plasmid-based expression is an attractive option in *S. cerevisiae* owing to the vast availability of replicating plasmids, this is a limitation in the case of *S. intermedia* due to no known autonomously replicating sequence (ARS) available for plasmid maintenance in this yeast [105]. ARSs are short (<100 bp) DNA sequences which serve as the binding sites for the origin of replication machinery and enable plasmid replication in the strain [107]. Plasmids recently developed for CRISPR genome editing in *D. hansenii* (another CUG-Ser1 clade yeast) were tested in *S. intermedia* and results confirmned that the ARS functional in *D. hansenii* could be employed in *S. intermedia* [104]. Thus, multiple plasmids originally developed for *D. hansenii*, containing different combinations of sgRNA expression cassettes based on RNA polymerase (Pol) II and III designs and both heterologous and endogenous promoters, were also tested in *S. intermedia* [104]. However, none of the plasmids gave positive results for deletion of *ADE2* gene (red colonies), suggesting a need for a CRISPR system tailored specifically for *S. intermedia*.

Thus, a CRISPR plasmid was developed for *S. intermedia,* comprising of the *CaNAT1* selection marker, a codon optimized Cas9, gRNA under tRNA based promoter and with a self-cleaving hepatitis delta virus (HDV) ribozyme. Codon Adaptation Index (CAI) was used to assess the preferred codon usage for *S. intermedia* and the CUG-optimized *CAS9* from *S. pyogenes,* originally codon optimized for *S. stipitis* was chosen (**Paper III: Fig. 1A**) [108]. Moreover, endogenous and constitutive promoters *SiRNR2* and *SiTAF6,* of medium or weak strength respectively, were chosen to avoid strong expression of *CAS9*, which can be toxic to the cell [104, 109]. The design of gRNA was adapted from previous works in *S. cerevisiae* and *C. albicans*, which show the correlation of high cellular levels of sgRNAs with improved efficiency of Cas9-mediated genome targeting by using tRNA promoter based guide designs [110, 111]. Additionally, the abundance of the gRNA is also correlated to the presence of self-cleaving hepatitis delta virus (HDV) ribozyme, which protects the 5' end of the sgRNA from 5' exonucleases (**Paper III: Fig 1B**). The resulting plasmid is shown Fig. 7.

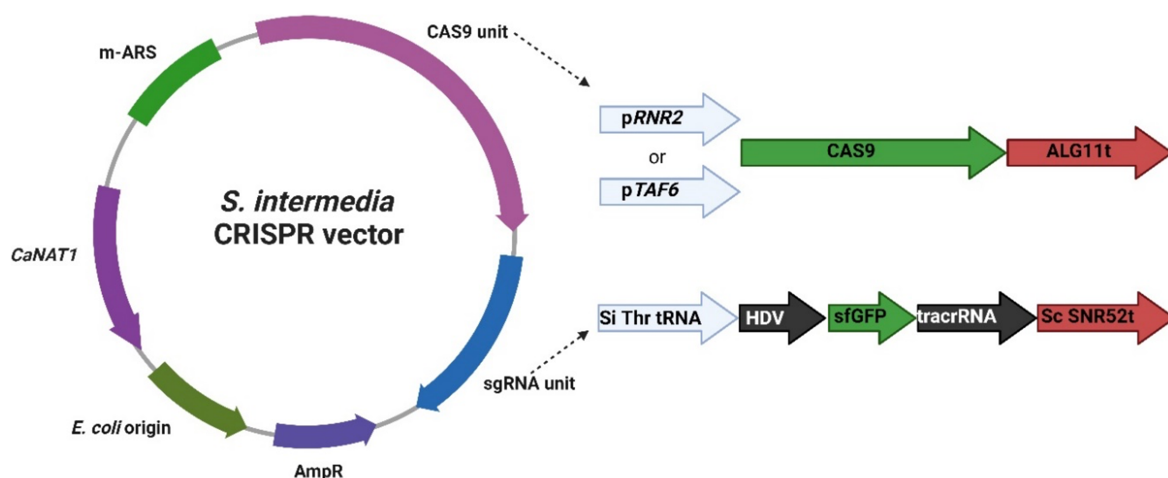

Figure 7. Plasmid map of CRISPR engineering plasmid for *S. intermedia*. The map details all the designed parts as well as a more detailed depiction of both the *CAS9* expression unit and sgRNA expression unit. HDV: hepatitis delta virus; sfGFP: super-folder green fluorescent protein. Figure taken from **Paper III**.

Validation of the CRISPR plasmid developed for *S. intermedia* was performed using X-GAL-based blue/white screening [112]. Using *S. intermedia b*-galactosidase encoding *LAC4* as a target gene, CRISPR transformants were screened by plating on X-GAL plates containing galactose as carbon source, as galactose is an inducer of lactase in *S. intermedia* [68]. Transformation using a medium strength *RNR2* promoter endogenous to *S. intermedia* resulted in 94.5% efficiency. The use of weaker strength *SiTAF6* promoter resulted in targeting efficiency of 90.4% (**Paper III: Fig. 3D**). Since no repair fragment was provided in this experiment, the DSB repair is likely mediated by NHEJ. Thus, the customized CRISPR-Cas9 system for *S. intermedia* resulted in high targeting efficiency (90-94%), where *LAC4* gene indels led to the loss of lactase activity, producing white colonies.

In metabolic engineering workflows, it is desirable to have the option to not only delete or knock-out genes but also be able to integrate heterologous gene cassettes [9]. To determine the applicability of the CRISPR-Cas9 system for targeted gene integration in *S. intermedia*, a *LAC4* deficient strain with an internal frameshift (*LAC4Δ610-618*), was constructed. On X-GAL plates, the colonies of this mutant do not turn blue. Integration of a short double stranded oligo (45bp homology on each side of the Cas9 cutting site) enables restoration of *LAC4* locus and appearance of blue colonies (just as in the wild-type strain). Transformation of the *LAC4Δ610-618* allele strain with a repair fragment under the same experimental setup resulted in a targeting efficiency of 14% indicating NHEJ dominance in *S. intermedia* (**Paper III: Fig. 4C**). In contrast to the significant increase in targeting efficiency observed for many other NCYs upon deletion of *KU70* gene, [104, 106] in *S. intermedia* deleting *KU70* resulted in only 4% increase in targeting efficiency.

3.3 System level understanding using Genome Scale Metabolic Models

Systems biology integrates mathematical analysis and computational modelling to uncover interactions within biological systems, linking molecular components to physiological functions and phenotypes through quantitative reasoning [113]. Two key approaches of systems biology are: top-down and bottom-up. The top-down approach looks at the behavior of the system as a whole and uses genomic, transcriptomic, proteomic, and metabolomics to comprehend the functions of parts of the metabolic network [114]. In contrast, the bottom-up approach uses molecular properties to construct models to predict system properties [14, 114]. GEMs are knowledge databases containing information about the genome and metabolism pertaining to an organism/cell, as well as predictive tools that use functions of linear programing to enable systematic explorations of genotype-phenotype relationships as well as predictions of metabolic engineering strategies [113].

GEMs have been widely used to understand metabolism in NCYs such as *K. phaffii*, *Y. lipolytica*, *S. stipitis*, *Issatchenkia orientalis*, and *R. toruloides*, providing insights into their diverse metabolic capabilities [115, 116]. For example, GEMs have been employed in *K. phaffii* to explore subcellular protein localization [117], and in *S. stipitis* to investigate its oxidoreductive metabolism [118]. The extensive application of GEMs in these yeasts highlights the growing role of systems biology as a key strategy for understanding and enhancing metabolic functions, designing novel cell factories for valuable chemicals, and deciphering molecular mechanisms underlying complex metabolic processes. [73]. Thus, it was desired to reconstruct a GEM also for *S. intermedia*.

3.3.1 Reconstruction of *Sint*GEM for *S. intermedia*

GEM reconstruction is facilitated by tools like COBRA (Constraint-Based Reconstruction and Analysis) and RAVEN (Reconstruction, Analysis, and Visualization of Metabolic Networks) [119, 120]. RAVEN, which operates within the MATLAB environment, supports both manual and automated GEM reconstruction. It enables both, *de novo* model construction by importing species-specific data from the KEGG database to generate an initial draft which can be further refined, or using a model of a phylogenetically related species as the draft for further curation. In this study, *SintGEM* was reconstructed using the RAVEN toolbox in several steps as outlined below.

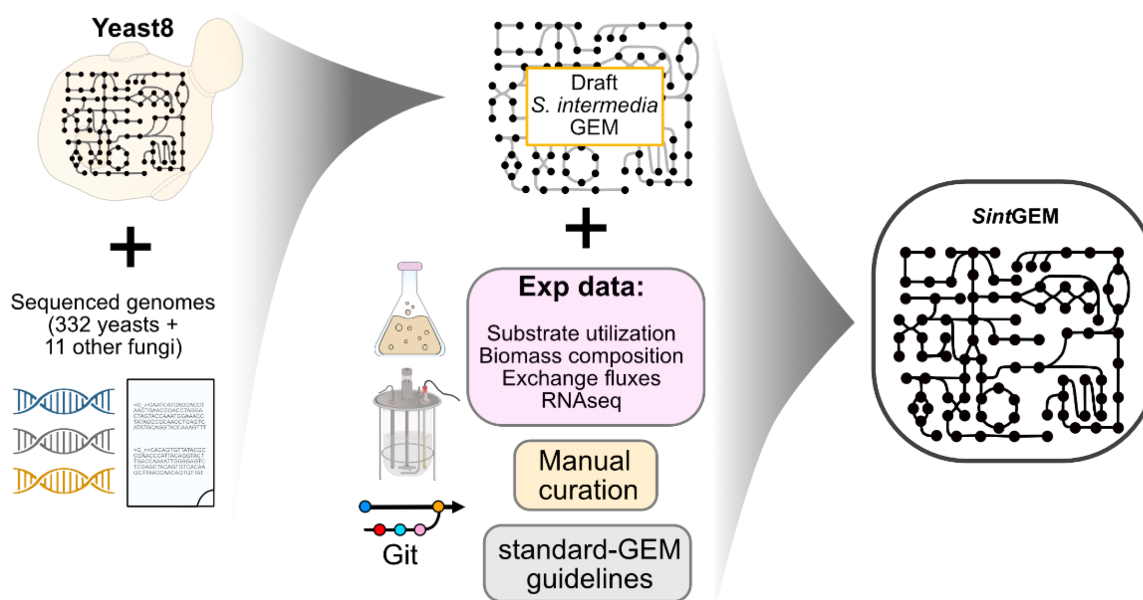

Figure 8. Illustration of the steps involved in the construction of *Sint*GEM from the draft Yeast8 model of *S. intermedia*. Reconstruction of the *Sint*GEM using the draft model for *S. intermedia* based on yeast8 and gene orthology predictions. Data-driven manual curation yielded a high-quality GEM consisting of 1070 genes, 3991 reactions, 2759 metabolites and 14 cellular compartments. As a part of this thesis, *Sint*GEM was further manually curated and experimental data was incorporated, resulting in the final version of *Sint*GEM used in **Paper V** to understand lactose and galactose metabolism in *S. intermedia*. Figure taken from **Paper V**.

Step 1: Creating a Draft Reconstruction

*Sint*GEM was generated using an automated workflow based on the consensus GEM for *S. cerevisiae* (Yeast8) as a scaffold, along with orthology searches across 332 sequenced yeasts and 11 other fungi by Lu and colleagues (Fig. 8) [121]. One of the ssGEMs generated in this process was for *S. intermedia* and was further curated as a part of the work in **Paper V**.

Step 2: Manual curation

Although automated methods generate draft reconstruction in an efficient manner, manual curation and inspection remain necessary to eliminate potential errors and inconsistencies [122]. The manual curation of *SintGEM* focused on refining the assimilation pathways for galactose and lactose in *S. intermedia* (**Paper V**). The draft model lacked an exchange reaction for lactose uptake, likely due to its absence in Yeast8, as *S. cerevisiae* does not metabolize lactose. This reaction was introduced into the model. Some steps of the galactose oxidoreductive pathway, seen in species like *Aspergillus niger*, *Aspergillus nidulans*, and *Trichoderma reesei*, were already present [123-125]. Further refinement involved integrating *S. intermedia*-specific experimental data, including

biomass composition, total protein, and total lipid measurements from glucose-limited chemostat cultures [126].

Omics data mapping enriches a GEM with context-specific information about the organism [113]. In this study, RNA-seq data for *S. intermedia* grown on different carbon sources was mapped onto the reconstructed network and allowed further understanding over the expression pattern of genes involved in the hypothesized oxidoreductive pathway in *S. intermedia* as discussed in **Chapter 4** [64].

Step 3: Model simulations and predictions

Reporter metabolite analysis integrates gene expression data with the structure of metabolic networks. This algorithm identifies "reporter metabolites," which are metabolites surrounded by significant changes in the expression of nearby enzymes. In **Paper V**, reporter metabolite analysis was applied to investigate the presence of an alternate galactose metabolic pathway in *S. intermedia*, detecting subtle yet coordinated gene expression changes in response to controlled perturbations and uncovering regulatory "hotspots" in the metabolic network [127].

Linear programming techniques, such as flux balance analysis (FBA), are used to solve linear equations representing an objective function, which is a quantifiable phenotype that is maximized or minimized in *in silico* simulations [128]. The function that is maximized or minimized is called an objective function, such as biomass production or ATP generation. Constraints are applied by setting upper and lower limits for reactions, such as the maximum carbon uptake rate or cellular growth rate. In this thesis (**Paper V**), FBA was employed to validate the functionality of the reconstructed *SintGEM* for growth on glucose, galactose, and lactose and better understand *S. intermedia* metabolism.

An extension of FBA, called flux sampling, can be used to fully explore a metabolic network. Unlike FBA, which optimizes a single objective function, flux sampling generates multiple flux distributions without a predefined objective. This approach allows for a range of possible flux distributions based on the constraints applied to the network, with each distribution representing a different attainable phenotype for the cell under those conditions [129]. In addition to FBA, flux sampling was employed in **Paper V** to explore various flux scenarios predicted by *SintGEM* for growth on galactose and lactose, providing deeper insights into the cell's metabolic processes.

Overall, the tools developed for *S. intermedia* enabled its exploration both *in vivo* and *in sillico* focusing on elucidating its lactose metabolism in the following **Chapter 4**.

Key take-away point from Chapter 3:

- Experimental and computational tools were developed to study the genetic and metabolic makeup of *S. intermedia*.
- Genome editing tools were facilitated by an electroporation-based transformation protocol and the identification of *CaNAT1* as a dominant selectable marker in *S. intermedia*.
- The split-marker technique using pJK795 achieved high targeting efficiency of 63.4% in *S. intermedia*. However, when coupled with FLP-recombinase (plasmid pJK863), targeting efficiency was significantly reduced (~10%).
- Designing and testing a CRISPR plasmid containing different CRISPR components optimized for efficient expression in *S. intermedia* resulted in high targeting efficiency (>90%) for NHEJ-mediated gene mutations, while substantially lower (~18%) for HR-mediated gene integration.
- The first curated genome-scale metabolic model (GEM) for *S. intermedia*, *Sint*GEM, was reconstructed using *Yeast8* as a draft model. Manual curation of the network topology, combined with experimental data and RNA-seq gene expression data on various carbon sources, refined the model. This allowed for predictions using reporter metabolite analysis, flux balance analysis (FBA), and flux sampling.

Chapter 4. Lactose metabolism in *S. intermedia*

This chapter reviews the work presented in **Paper IV** and **Paper V**, which focuses on advancing our understanding lactose of metabolism in *S. intermedia*. The primary emphasis is on lactose uptake and hydrolysis and the catabolism of the galactose moiety of the disaccharide. First, the genetic determinants underlying lactose metabolism in *S. intermedia* are identified, followed by exploration of their roles through the characterization of gene deletion mutants. Thereafter, the chapter describes the discovery and elucidation of an alternate galactose catabolic route in *S. intermedia*.

4.1. Abundance of lactose-growing yeasts

Lactose, the disaccharide composed of glucose and galactose linked by a β-1,4-glycosidic bond, is primarily produced by lactating mammals. While most fermented dairy products rely on the fermenting capabilities of lactic acid bacteria, some lactose-metabolizing yeasts such as *Debaryomyces hansenii* and *Geotrichum candidum* are highly prevalent on surface-ripened cheeses. Various *Kluyveromyces* species have also been isolated from the core of some cheese [130]. *S. cerevisiae*, *Candida kefyr* and *K. marxianus* are the predominant yeast species in the fermented drink kefir [131]. Although lactose-utilizing yeasts are present in various fermented foods and drinks, only 6% of 852 sequenced yeast species have been shown to metabolize lactose, making it a relatively rare trait (Fig. 9A) [132]. Among those capable of lactose metabolism, *Kluyveromyces lactis* and *K. marxianus* have been extensively studied as alternative model organisms in genetics and physiology [133-142] and are also considered as valuable cell factories for processes utilizing lactose-rich substrates [143-145]. However, most other lactose-metabolizing yeasts remain understudied, with little known about the genetics underlying their lactose metabolism. Additionally, as depicted in Fig. 9B, lactose-assimilating yeasts are dispersed throughout the phylogenetic tree, suggesting large genetic differences between the lactose-growers. Exploring these yeasts at the genomic and transcriptomic levels may uncover distinct regulatory and metabolic systems, shedding light on their metabolic diversity. Furthermore, a better understanding of these understudied yeasts can enable the development of future cell-factories with diverse product portfolios and tailored for industrial applications.

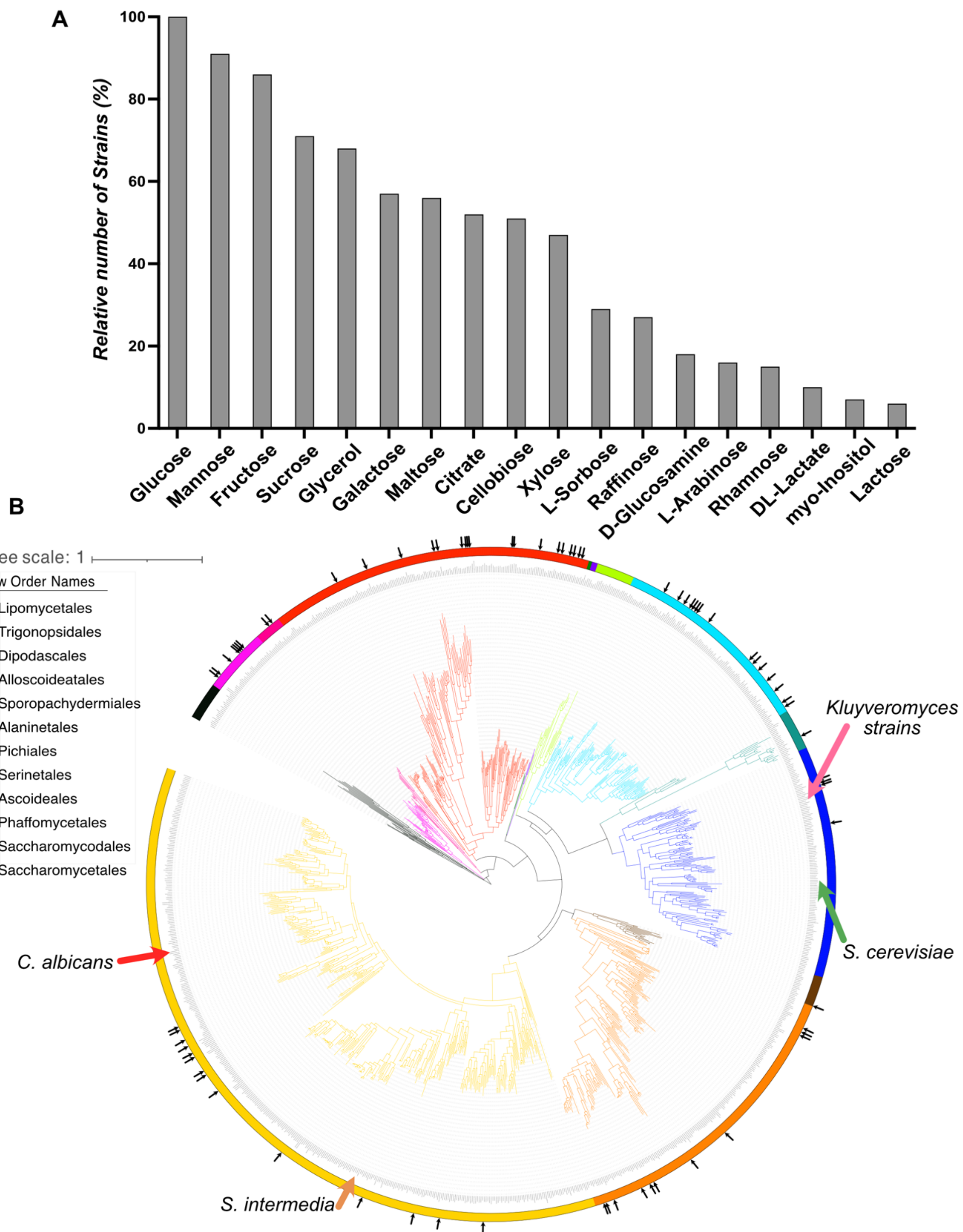

Figure 9. Abundance of lactose growing yeasts. A) Bar graph showing the diversity in substrate metabolism in different yeasts. On the y-axis, the relative percentage of yeasts (%) of the total 852 species characterized is visualized in a bar plot per substrate. Figure adapted from [132] under CC BY 4.0 license. B) Phylogenetic tree of 1154 yeast species color coded by subphylum. Black arrows show the phylogenetic spread of different lactose growers as shown by Opulente et al., 2024 [132]. Color arrows highlight the placement of the model yeasts *S. cerevisiae* (green), *K. lactis* and *K. marxianus* (blue), *C. albicans* (red) in the phylogenetic tree, and the orange arrow indicates the location of *S. intermedia*. Figure adapted from [132] under CC BY 4.0 license.

Previous works have shown that *S. intermedia* can grow on lactose as the sole carbon source, [47, 124, 132] and this yeast has been studied for the valorization of lactose from whey [146]. In **paper IV (Fig. 1B)**, three *S. intermedia* strains (CBS 141442, PYCC 4715 and type strain CBS 572) were grown along with 24 lactose-growing yeasts in minimal media containing lactose. Comparison of doubling time showed that two of the *S. intermedia* strains (CBS 4715 and CBS 141442) were among the top five fastest growing, next to the dairy yeasts *K. marxianus* and *K. lactis*, under the tested conditions (**Paper IV: Fig. 1B**).

4.2 Genetic determinants of lactose metabolism and regulation in yeasts

Metabolism involves interconnected pathways that perform chemical transformations within the cell. In eukaryotes, genes encoding proteins that participate in the same pathway are sometimes co-localized within the genome, forming so called metabolic gene clusters. Although these genes are transcribed from individual promoters, they are often co-regulated and work together to support a shared cellular function, such as sugar metabolism [147]. Several gene clusters have been extensively studied to understand gene regulation as well as evolutionary forces contributing to ecological adaptations [124]. In the well-studied dairy yeast *K. lactis*, lactose metabolism involves the concerted action of two gene clusters, the *LAC* and *GAL* clusters (Fig. 10A). The *LAC* cluster, responsible for uptake and hydrolysis of lactose, comprises of two genes, *LAC12* and *LAC4*. The *LAC12* gene encodes a membrane permease that facilitates lactose uptake that is divergently expressed under the same promoter as the *LAC4* gene encoding a β-galactosidase, responsible for intracellular lactose hydrolysis into glucose and galactose. The glucose moiety of lactose is metabolized through glycolysis, while galactose is processed via the Leloir pathway. The Leloir enzymes are encoded by the conserved *GAL* cluster genes *GAL1*, *GAL10* and *GAL7*, present in many yeasts including *S. cerevisiae, K. lactis* and *C. albicans* (Fig. 10B) [148].

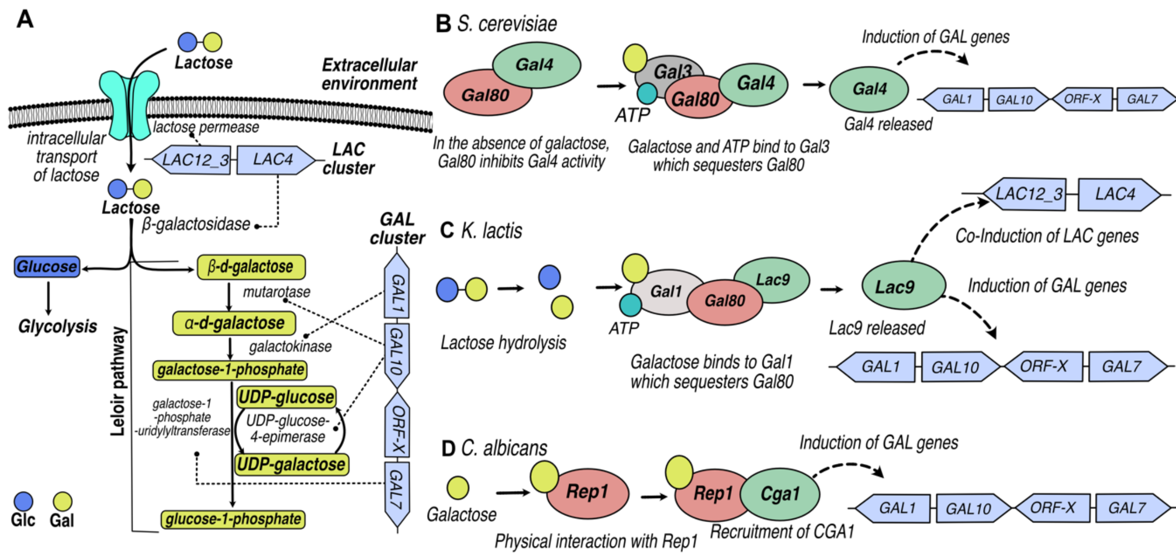

Figure 10. Lactose metabolism and regulation in yeasts. (A**)** Lactose uptake from the extracellular environment may be enabled by the membrane permease and subsequently hydrolyzed by the β-galactosidase enzyme. Genes coding for the permease (*LAC12*) and β-galactosidase (*LAC4*) are co-localized forming a cluster. Glucose is metabolized via glycolysis while galactose via the Leloir pathway. Metabolites in the Leloir pathway is depicted in yellow boxes while the genes encoding the enzymes partaking in this pathway are in a conserved cluster, called the *GAL* cluster. (B) Regulation of galactose metabolism in *S. cerevisiae* by the Gal3-Gal80-Gal4 system where galactose and ATP induce Gal3 to bind Gal80 resulting in the activation of Gal4. Thus, Gal4 induces expression of the structural *GAL* genes. (C) Regulation of galactose and lactose genes in *K. lactis* mediated by the bi-functional Gal1. The *Sc*Gal3 homolog in *K. lactis* (Gal1) is induced by galactose (or galactose derived from lactose) resulting in sequestering Gal80 and relieving Lac9, which in turn activates the interconnected galactose and lactose metabolic genes in this yeast. (D) Graphic representation of the *C. albicans* galactose regulatory system through Rep1 and Cga1. Galactose physically binds to Rep1 resulting in recruitment of Cga1, and the complex ultimately induces the structural genes responsible for galactose metabolism in this yeast. Figure adapted from **Paper IV**.

While the metabolic role of the *GAL* cluster is similar in *K. lactis*, *S. cerevisiae* and *C. albicans*, there are differences observed at the level of regulation. As depicted in fig. 10, galactose regulation in *S. cerevisiae* is mainly attributed to the interactions of Gal3, Gal4 and Gal80 [136], where the regulator Gal3 has duplicated and diverged from the galactokinase Gal1 and lost its enzymatic activity [136]. In *K. lactis*, the transcriptional activator Lac9, ortholog of Gal4 in *S. cerevisiae*, activates the expression of both the *LAC* and *GAL* cluster genes. This ensures a coordinated regulation of lactose uptake, hydrolysis, and subsequent galactose catabolism. *K. lactis* harbors only one Gal1, which has retained the dual functionality of a regulator and an enzyme [141]. In contrast, *C. albicans* galactose regulation does not involve the Gal1-Gal80-Gal4 proteins but instead relies on Rep1 and Cga1 [149].

4.2.1 A unique gene cluster partakes in lactose and galactose metabolism in *S. intermedia*

The genome of *S. intermedia* contains both the conserved *LAC* and *GAL* clusters, and the genes in two clusters are upregulated on galactose and lactose as compared to glucose. Additionally, *S. intermedia* harbors a novel cluster comprising several genes seemingly involved in galactose and lactose metabolism: a putative transcriptional regulator gene *LAC9*, a galactokinase gene *GAL1*, an epimerase-encoding gene *GAL10*, and *XYL1* encoding one of the three previously studied xylose/aldose reductases [64]. Enzyme assay-based characterization suggests that Xyl1_2 has specificity for galactose (**Paper V: Fig. S1**) [64]. Thus, this novel cluster was named the *GALLAC* cluster as it contains homologs of genes associated to both galactose and lactose metabolism. All *GALLAC* cluster genes except *LAC9* were highly upregulated in both galactose and lactose, indicating a possible metabolic or regulatory role of this cluster.

Next, comparative genomics of 332 genome-sequenced ascomycetous yeasts was performed to determine the uniqueness of the *GALLAC* cluster to *S. intermedia* and the evolutionary origin of this cluster. Interestingly, the analysis suggested that the *GALLAC* cluster is unique to *S. intermedia*, and that the formation of the cluster is attributed to gene duplication events within the species. Thus, the genes in the *GALLAC* cluster were annotated *LAC9_2, GAL1_2, GAL10_2* and *XYL1_2* (**Paper IV: Fig. 3D & E**).

Since this study was published, the number of ascomycetous yeast species with sequenced genomes has expanded to 1,154 [132]. Although no close relatives to *S. intermedia* except the sister species *S. pseudointermedia* can grow on lactose, a reanalysis may be required to confirm if the *GALLAC* cluster remains unique to *S. intermedia*.

4.2.2 Elucidation of a novel lactose and galactose regulation

The study of mutant strains with alterations or deletions in their nucleotide sequences has long been a fundamental approach in biology [150], and this approach was used in **Papers IV** and **V** to characterize the genes within the three clusters. By employing the split marker method (described in **Chapter 2**), gene deletion mutants for each of the three clusters were constructed (Fig. 11A) and then characterized for growth and metabolite production (Fig, 11B, C) providing important insights into sugar metabolism in the different strains (**Paper IV** and **V)**. The mutant phenotyping confirmed

the significant roles of the *GAL* cluster for growth on galactose and the *LAC* cluster for growth on lactose as was expected based on studies on the model yeasts [151, 152] (Fig. 11A). Additionally, the significant growth defects observed for the *GALLAC* cluster deletion mutant (*gallacΔ*) on both galactose and lactose suggested that the conserved *GAL* and *LAC* clusters alone are not sufficient for metabolism of these carbon sources (Fig. 11A). Instead, there seemed to be an interdependence between the three clusters.

Next, individual gene deletion mutants of the *GAL* and *GALLAC* clusters were constructed and characterized (Fig. 11 B, C). Deleting *GAL1* from the *GAL* cluster resulted in a severe growth defect on galactose, similar to the *GAL* cluster mutant phenotype, indicating that Gal1 plays a key role in galactose metabolism via Leloir pathway, either as galactokinase or/and a regulator [153]. However, the presence of the paralog *GAL1_2* in the *GALLAC* cluster did not complement the *GAL1* deletion, indicating a different role or a different regulatory system in place for this protein.

Similar characterization of individual gene deletions in the *GALLAC* cluster revealed significant roles of *LAC9_2* and *GAL1_2* genes. Detailed analysis showed that the *lac9_2Δ* mutant exhibited a longer lag phase on both galactose and lactose compared to wild-type (Fig. 11 B, C), suggesting a regulatory role, similar to what has been observed for *Kl*Lac9 [133]. The *lac9_2Δ* could still grow on both carbon sources, which could be due to complementation by the non-clustered paralogous genes *LAC9* (45% identity to Lac9_2) and *GAL4* (18% identity to Lac9 and 19% to Lac9_2) also present in the genome of *S. intermedia*. However, deletion of *LAC9* or *GAL4* alone did not lead to any noticeable growth defects on either galactose or lactose, suggesting that *LAC9_2* is the main transcriptional regulator (**Paper IV: Fig. S6**).

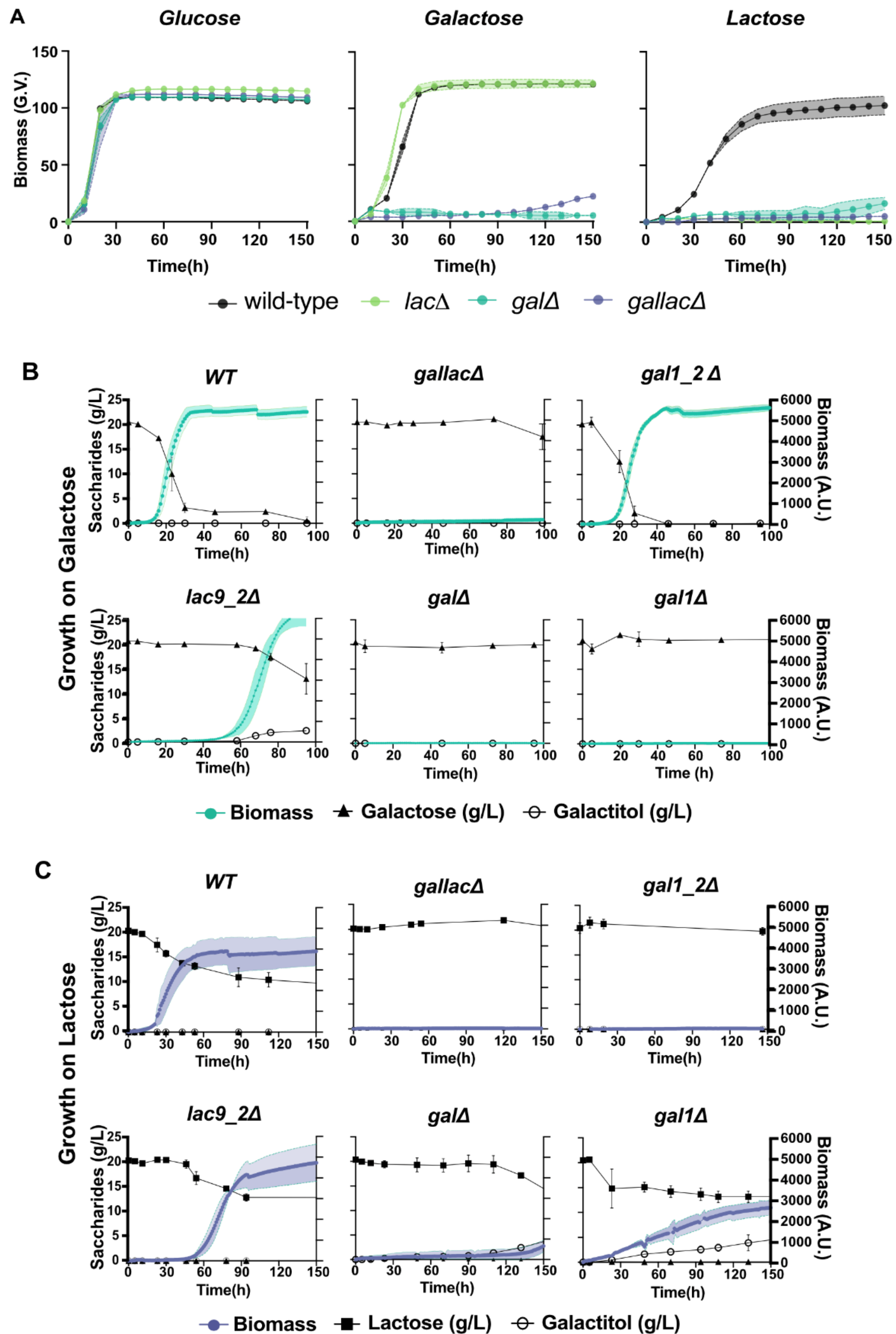

Figure 11. Elucidation of a novel lactose and galactose metabolism in *S. intermedia*. A) Cluster deletion mutants of *S. intermedia* were characterized by growth on glucose, galactose, and lactose. Legend shows the wild-type strain (black), *LAC* cluster mutant (light green), *GAL* cluster deletion mutant (dark green), and *GALLAC* cluster deletion mutant (purple), depicted in the graph with biomass as (G.V.) corresponding to growth based on pixel counts as determined by GrowthProfiler90, on the y-axis, against time (h) on the x-axis. Data are represented as mean ± standard deviation (shaded region) for biological triplicates. Figure taken from **Paper IV**. B &C) Growth and metabolite profiles of cluster mutants *gallacΔ*, *galΔ* and individual gene deletions *gal1_2Δ*, *lac9_2Δ* and *gal1Δ* for growth on B) galactose or C) lactose containing media in a cell

growth quantifier. Graphs represent biomass (filled circle; gal—dark green; lac—purple) on the right y-axis, consumption of respective sugars (filled triangle for galactose in g/L or filled square for lactose in g/L), and metabolite production (open circle for galactitol in g/L) on the left y-axis (depicted by saccharides (g/L)), plotted against time (h) on x-axis. Data are represented as mean ± standard deviation (shaded region for biomass and bars for sugars and metabolites) for biological triplicates. Figure taken from **Paper IV**.

To understand how Lac9_2 regulates transcription of genes involved in lactose and galactose metabolism, transcription factor binding motif analysis was performed using Multiple Expectation maximizations for Motif Elicitation (MEME) (**Paper IV: Fig. 5A**). MEME has been used to determine binding motifs of transcription factors in multiple studies on model yeasts [148, 154, 155]. The MEME software looks for motifs, which are short, conserved patterns, and then compares these motifs with databases assigning them score of significance (p-value). Such analysis for binding motifs in the three clusters identified motifs for Lac9 (Gal4) in the *GALLAC* cluster, specifically in the promoter regions of *GAL10_2* and *XYL1_2*/*GAL1_2*. To verify the bioinformatic prediction, gene expression analysis using qPCR was performed for *XYL1_2*, *GAL1_2* and *GAL10_2* in the wild-type and *lac9_2Δ* strains grown on glucose, galactose and lactose (**Paper IV: Fig. 5B**). A significant drop in fold change was observed for all three *GALLAC* genes in *lac9_2Δ* as compared to wild-type strain, confirming Lac9_2 as a transcriptional activator of the *GALLAC* cluster.

Interestingly, Lac9/Gal4 motifs were not identified in the promoters of the *LAC* or *GAL* cluster genes, indicating that other transcription factors regulate these genes. This aligns with previous studies showing that *S. intermedia* and many other CUG-Ser1 clade yeasts lack Lac9/Gal4 binding sites in their *GAL* cluster promoters [156]. Furthermore, the absence of *GAL80* in the *S. intermedia* genome strongly suggests that *S. intermedia* has evolved a distinct mechanism for regulating galactose and lactose metabolism, clearly different from the Gal1/3-Gal80-Gal4 regulon described in *S. cerevisiae* and *K. lactis* [134, 157].

The previously mentioned *gal1_2Δ* mutant could grow on galactose, but not on lactose (Fig. 11B, C). This result indicated that perhaps Gal1_2 has a regulatory role in lactose metabolism by acting on the *LAC* cluster in *S. intermedia*. Like the investigation of the function of *LAC9_2* in *S. intermedia*, the *GAL1_2* regulatory role was investigated using different approaches. To differentiate whether the *S. intermedia GAL1* and *GAL1_2* functioned as enzymes, they were heterologously expressed in the *S. cerevisiae* BY4741 *gal1* knockout strain. *S. intermedia* Gal1 and Gal1_2 proteins could complement *ScGAL1* deletion (**Paper IV: Fig. 6A**), demonstrating that they both have galactokinase activity.

To determine if Gal1_2 and/or Gal1 regulate lactose metabolism via β-galactosidase enzyme activity, the wild-type, *lacΔ*, *gal1Δ* and *gal1_2Δ* strains were grown on galactose or lactose, and enzyme activity assay was performed for samples taken at different timepoints during growth. A clear distinction in enzyme activities for *gal1Δ* as compared to the *gal1_2Δ* mutant confirmed that the Gal1_2 is an important regulator of lactase activity, as its deletion led to a complete growth arrest and diminished lactose uptake and hydrolysis. Lastly, qPCR results provided evidence for Gal1_2-mediated regulation of *LAC* cluster occurring at the transcriptional level. These results suggest that the original Gal1 has retained its primary function as the main galactokinase, while Gal1_2 has evolved to serve as a key regulator.

Overall, the findings in **Paper IV** reveal the presence of a novel and unique *GALLAC* gene cluster in *S. intermedia*, in addition to the *LAC* and *GAL* clusters, that together govern galactose and lactose metabolism. As illustrated in Fig. 12, this work highlights the interdependence among these three clusters and the central role of the *GALLAC* cluster in orchestrating a seemingly unique regulatory system within *S. intermedia*. Due to limitations in time and advanced characterization tools for *S. intermedia*, a complete picture of this intriguing system could not be achieved. However, the functional differentiation of the two Gal1 proteins in this yeast, the Lac9-mediated *GALLAC* cluster regulation and the interdependence among the cluster genes presents a fascinating example of the regulatory complexity of yeast metabolism.

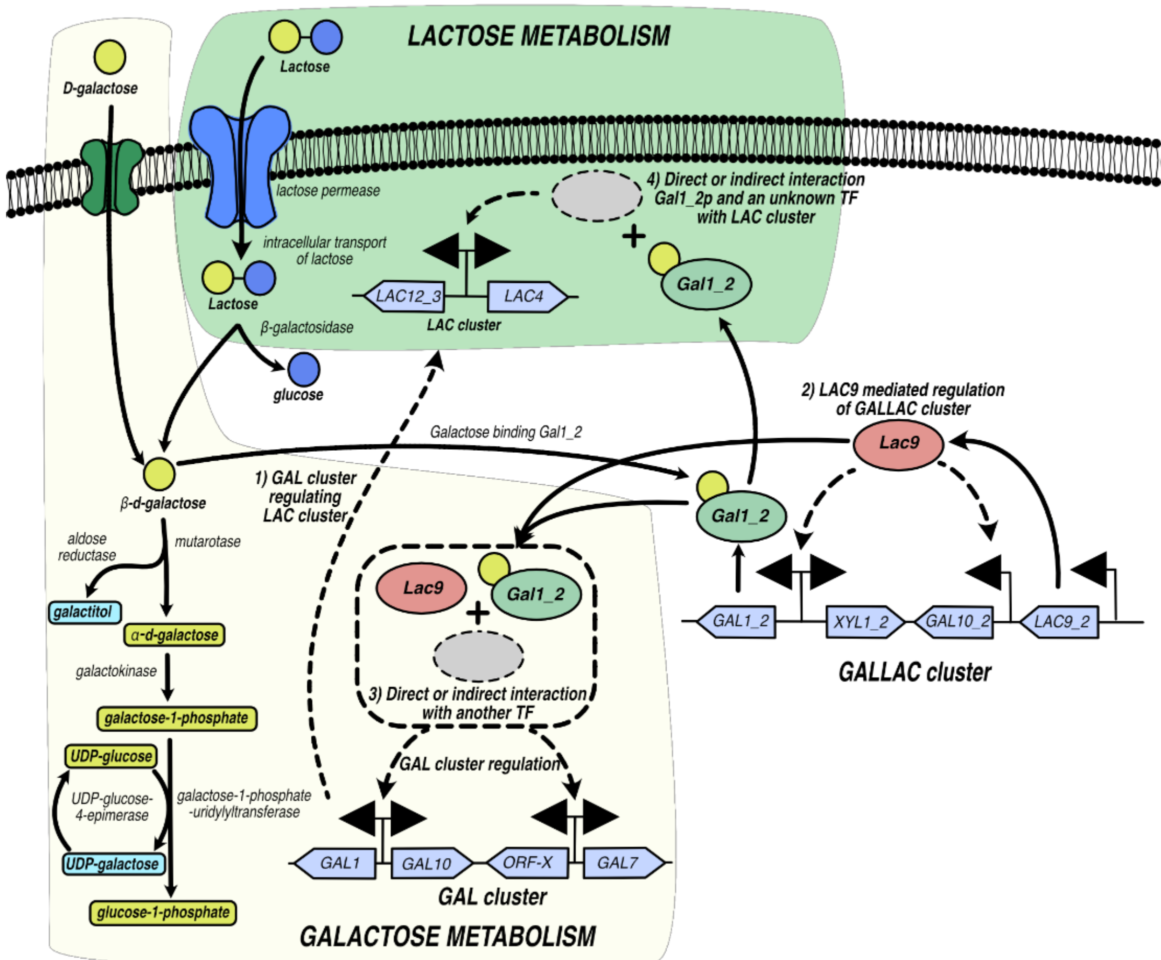

Figure 12. Depiction of lactose (green box) and galactose (light yellow box) metabolism in *S. intermedia*. Instances of cluster interdependence are depicted with 1) the *GAL* cluster having a regulatory effect on the *LAC* cluster as *galΔ* cannot grow on lactose; 2) regulation of *GALLAC* cluster by the transcription factor *Ci*Lac9; 3) on galactose, Lac9 and Gal1_2 interact directly or indirectly resulting in the regulation of *GAL* cluster gene(s), thus affecting *S. intermedia*'s growth; and 4) on lactose, Gal1_2 from the *GALLAC* cluster regulates the *LAC* cluster at a transcriptional level. This effect of Gal1_2 can be speculated to be indirect due to the inability of Gal1_2 to bind DNA or protein based on predicted structure. Graphical representation also illustrates the overflow metabolism in *S. intermedia* through aldose reductase-mediated conversion of galactose to galactitol. Figure taken from **Paper IV**.

4.3 An alternate pathway for galactose metabolism

As mentioned earlier in section 4.2, the Leloir pathway is the main galactose metabolic pathway and has been extensively studied in model yeasts [158]. However, filamentous fungi such as *A. niger* and *A. nidulans* possess an additional galactose metabolic pathway called the oxidoreductive pathway (Fig. 13A) [159]. The oxidoreductive route of galactose metabolism contains a sequence of steps of oxidation and reduction, converting galactose to fructose-6-phosphate which is then

metabolized via glycolysis [160]. Additionally, a recent study used machine learning-based predictions (validated by growth and biochemical characterization) to show that yeasts belonging to the genus *Ogataea* and the clade containing *Candida auris*, possess the oxidoreductive pathway [124]. The prediction was based on two features: a) growth on galactose despite the absence of *GAL* genes and, b) growth on galactitol, the reduction product of galactose and the intermediate of the oxidoreductive pathway [123].

Three observations from the work in **Paper IV** were crucial in shaping the hypothesis for **Paper V** that *S. intermedia* also metabolizes galactose via an oxidoreductive pathway. First, the synteny of the *GALLAC* cluster includes the *XYL1_2* gene encoding an aldose reductase that can convert galactose to galactitol. This gene is situated among galactose and lactose metabolic and regulatory genes and shares a promoter with *GAL1_2* that regulates the *LAC* cluster (Fig. 13B). Second, growth of the *galΔ* strain on lactose was accompanied by the production of galactitol, detected extracellularly in the media at high yield (90% of the theoretical maximum yield from the galactose moiety), indicating the action of an aldose reductase (Fig. 13C). Third, the production of galactitol was sometimes followed by its consumption, suggesting the presence of a galactitol catabolic pathway downstream of the galactose-to-galactitol conversion step [123, 160]. The hypothesis was further strengthened by the observation that three *S. intermedia* wild-type strains could grow on galactitol, again suggesting the presence of genes products downstream of Xyl1 that facilitate the catabolism of galactitol.

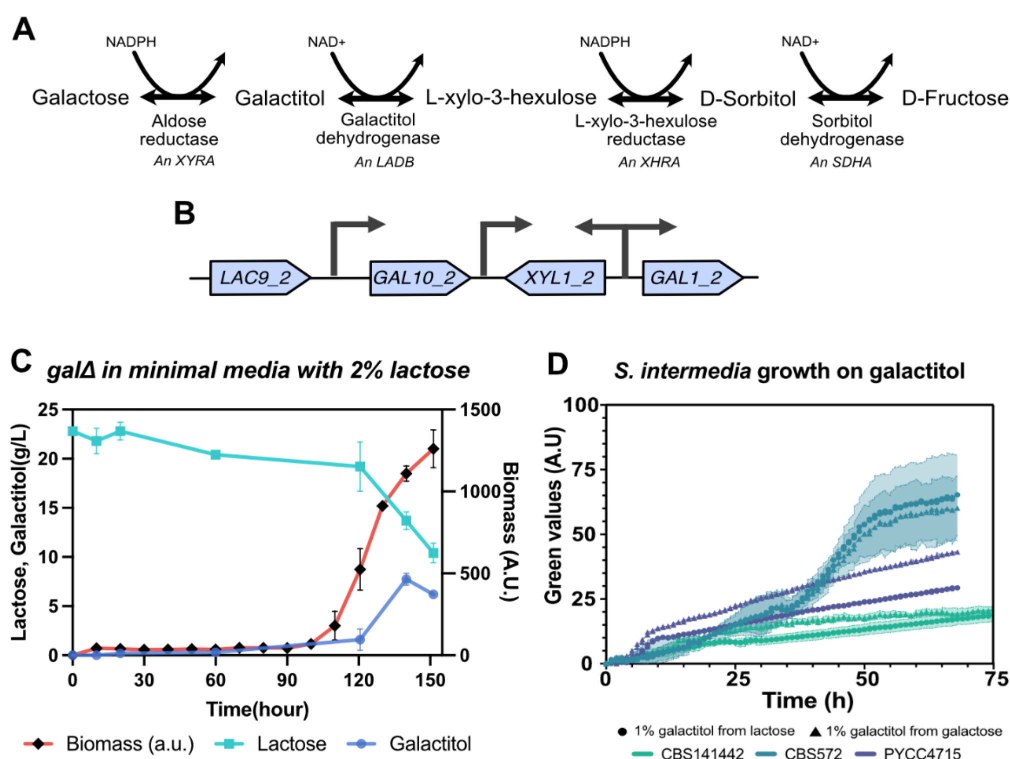

Figure 13. Alternate galactose metabolic pathway in *S. intermedia*. A) Oxidoreductive pathway in *A. nidulans* with the respective enzyme encoding genes from the work of Mojzita et al., 2012 (Figure adapted from [123]). B) *GALLAC* gene cluster synteny with the *GAL1_2* (lactose regulator) and *XYL1_2* (aldose reductase) under the same promoter and regulated by the transcriptional regulator Lac9_2. C) Growth, lactose consumption and galactitol production of *S. intermedia GAL* cluster mutant *galΔ* in minimal media containing 2% lactose. D) Growth of *S. intermedia* strains CBS 141442 (isolated from our lab); CBS 572 (type strain for *Candida intermedia*), PYCC 4715 (isolated from sewage in Portugal). Figure taken from **Paper V**.

4.3.1 *Sint*GEM-mediated exploration of the oxidoreductive pathway

In this study, *Sint*GEM was used for multiple reasons. First, the work of Lu and colleagues provided easy access to the automatically generated model for *S. intermedia* [161]. Second, analysis of the pathway was possible from a genome-wide perspective (considering pathway dependencies and compartmentalization), where *Sint*GEM enabled to interactively analyse and visualize a stoichiometrically feasible metabolic network. Third, the availability of RNAseq data for *S. intermedia* presented the opportunity to add a regulatory layer to the model. However, as a stoichiometry-based model, *Sint*GEM is limited by its steady-state assumption overlooking time-dependent changes and simplified regulatory effects due to lack of enzyme and protein constraints [162].

An initial assessment of the network topology of lactose metabolism in *S. intermedia* revealed a missing lactose uptake step, likely due to using Yeast8 as the draft model, which was

therefore manually added. The reconstructed network in *Sint*GEM included most of the steps of the oxidoreductive pathway (taking filamentous fungi as reference), highlighting the effectiveness of the homology-based reconstruction pipeline and the role of metabolic profiling [163]. The metabolites D-tagatose and L-sorbose were found to be closely linked to the oxidoreductive pathway in *SintGEM*. Fekete and colleagues reported that while they observed the conversion of galactitol to L-sorbose by alditol dehydrogenase in *A. nidulans*, D-tagatose could also be expected as a potential product [164]. The model contained exchange reactions associated to these metabolites as well, while the exchange reaction for galactitol was introduced according to experimental observation (Fig. 13C).

After curating the biomass composition and energy requirements in the model, gene expression data was incorporated. Combined analysis of the gene expression profile and pathway fluxes in *Sint*GEM on galactose and lactose suggested that, although Leloir pathway is the dominant pathway (as evidenced by the *galΔ* phenotype on galactose, Fig. 11B), the oxidoreductive pathway can also carry flux to maximize carbon assimilation. Reporter metabolite analysis on galactose indicated regulatory hot spots around metabolites in the oxidoreductive pathway, further supporting the hypothesis (Fig. 14). Random flux sampling was performed to explore conditions in which the oxidoreductive pathway is active (either alone or alongside the Leloir pathway) during galactose and lactose metabolism. The highest frequency of pathway usage occurred under oxygen limited growth, as expected for redox metabolism [165], and certain cases of carbon and nitrogen limitation were also predicted to activate this pathway [166].

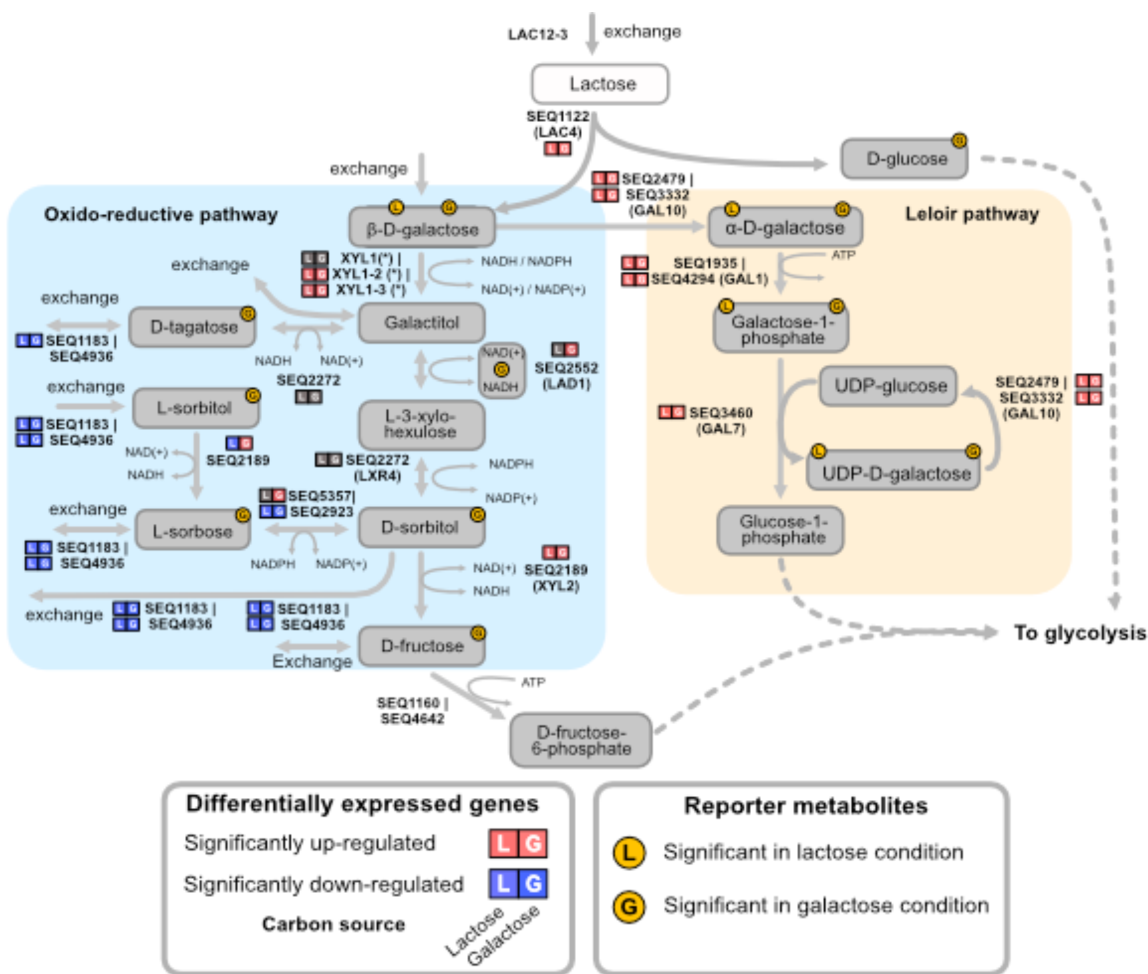

Figure 14. Transcriptional insights support activity of the *Sint*GEM-predicted oxidoreductive pathway during growth on lactose and galactose. Differential expression profiles and reporter metabolite analysis of genes and pathway intermediates of the Leloir pathway and the hypothesized oxidoreductive pathway in *S. intermedia* wild-type strain on galactose and lactose. Differential expression of genes is depicted by squares indicating directionality (red = upregulation; blue = downregulation) and the utilized carbon source (indicated by L = lactose and G = galactose). Similarly, reporter metabolites are represented by yellow circles indicating the carbon source in which the metabolite is characterized. Figure taken from **Paper V.**

4.3.2 Complex regulatory network and pathway redundancy

Elucidation of oxidoreductive pathway constitutive genes and their function could provide us with better insights about galactitol metabolism in *S. intermedia*. Therefore, genes involved in this hypothesized pathway, as identified through the *Sint*GEM, were specifically targeted for deletion and tested for growth on galactitol and sorbitol (**Paper V: Fig.7**). Interestingly, the deletion of *XYL1_2* gene completely arrested growth on galactitol but not on sorbitol, indicating a direct or indirect role in galactitol assimilation. However, individual gene deletions of downstream

oxidoreductive genes (*LAD1*, *LXR4* or *XYL2*) did not arrest growth neither on galactitol nor sorbitol while Leloir pathway mutants (*gal1Δ*, *gal7Δ* and *gal10Δ*) failed to grow on galactitol (**Paper V: Fig.S7**). While the phenotypes observed for downstream oxidoreductive pathway deletion mutants bring forward the possible redundancy in the oxidoreductive metabolism of *S. intermedia*, just like in filamentous fungi, the Leloir gene deletion mutants indicate a possible regulation of the oxidoreductive pathway by the *GAL* genes [125]. Orthology searches indicated that the tested downstream genes (*LAD1*, *LXR4* and *XYL2*) have no predicted paralogs in *S. intermedia*, suggesting that the observed phenotype may be due to the high substrate promiscuity of oxidoreductive enzymes [125, 160]. In other words, other oxidoreductases may compensate for the loss of these enzymes by acting on similar substrates, despite a lack of sequence similarity [160]. Alternatively, the phenotype may result from carbon flow through alternative pathways, not accounted for by *SintGEM*, involving galactitol, a limitation of reconstructed models for NCYs [9].

4.3.3 Role of dual-cofactor specificity of the Aldose reductase Xyl1_2

Microorganisms face challenges in the environments around them, which require metabolic flexibility that may be conferred by redox metabolism. Redox processes play crucial roles in several key situations such as substrate and oxygen gradients, stresses caused by variation in pH or the detoxification of reactive oxygen species [165, 167]. Among the many electron carriers in the cell, the oxidoreductive pathway involves NADH/NADPH. In **paper V**, the roles of NADH and NADPH have been emphasized to better understand the oxidoreductive pathway, especially in the context of lactose and galactose metabolism and galactitol as an overflow metabolite. The redox balance, mediated by these cofactors, is also critical for understanding how *S. intermedia* utilizes alternative metabolic routes when the Leloir pathway is disrupted, and for exploring how carbon flux is redirected under various growth conditions.

Aldose reductases use predominantly NADPH as the preferred cofactors [168]. However, the *XYL1_2* encoded aldose reductase in *S. intermedia* shows dual cofactor specificity (**Paper V: Fig. S1**) and this was curated in the model [64]. Flux sampling over all limiting conditions tested (carbon, oxygen and nitrogen) for growth revealed that NADH is the stoichiometrically optimal cofactor in *Sint*GEM. However, dual cofactor usage was observed in 42% of simulations with fluxes through the oxidoreductive pathway, which greatly influences the physiology of the cell. The simulations of the stoichiometric model suggest that the advantages of dual cofactor specificity are associated with trade-offs to ensure NADPH availability for anabolic reactions (e.g. membrane lipids) [116].

4.3.4 Elucidation of galactitol overflow metabolism in *S. intermedia*

Overflow metabolism has been extensively studied in *S. cerevisiae*, particularly in relation to ethanol formation. Ethanol is considered an electron sink, serving as a mechanism that enables the yeast to rapidly metabolize glucose while diverting excess electrons away from cellular metabolism into the production of ethanol. This allows *S. cerevisiae* to maximize its growth rate under glucose-rich conditions, even though the production of ethanol is less energetically efficient than respiration. This phenomenon is a classic example of how yeasts balance metabolic flux to accommodate environmental conditions and resource availability [165].

Having detected galactitol extracellularly for the *galΔ* mutant grown in lactose (Fig. 13C), *Sint*GEM was probed to better understand specific limiting conditions that result in galactitol production in *S. intermedia*. Random sampling and FBA simulations were performed to get unbiased secretion phenotypes under carbon, oxygen and nitrogen limiting conditions on galactose (**Paper V: Table 1**).

Under carbon limitation, only a minor fraction of the random sampling simulations displayed secretion of galactitol (less than 1% for all cases) (Table 1). Additionally, FBA simulations for growth under galactose-limiting conditions predicted complete respiration of the carbon source with no secretion of galactitol (**Paper V: Fig.6A**). Similarly, under oxygen limitation, random sampling of the *SintGEM* solution space showed that galactitol was secreted, but coupled with ethanol secretion, on both galactose and lactose, for most of the tested simulations (>94% of the cases) (Table 1). FBA for growth with oxygen as the limiting factor, showed that almost 60% of the simulations predicted coupled secretion of ethanol and galactitol using galactose as the carbon source. (**Paper V: Fig. 6D**).

Table 1: Percentage of simulations for each secreting ethanol and oxidoreductive pathway intermediates in FBA random sampling using *Sint*GEM at different limiting conditions. Percentage is calculated over 10,000 simulations run for each combination of galactose or lactose and nutrient limitation.

Metabolite	Galactose			Lactose		
	C-limited	O2-limited	N-limited	C-limited	O2-limited	N-limited
Ethanol	0.13%	94.84%	0%	0.24%	100%	0
Galactitol	0.8%	60.51%	0.33%	0.11%	100%	42.93%

Bioreactor cultivation of *S. intermedia* wild-type strain performed in aerobic (21% O_2 in inlet air) condition with galactose as sole carbon source indicated that galactitol production contradicted

model prediction of carbon limitation. Significant levels of galactitol production (between 36-60 hours) coincided with maximum cellular growth, the highest rate of galactose consumption and the highest (but non-depleting) oxygen uptake rates (Fig.15A). However, galactitol production was followed by its consumption, observed when galactose approached depletion in the media. Model predictions of oxygen limitation aligned well with experimental data as both ethanol and galactitol were observed in low-oxygen reactors (2% O_2 in inlet air) on galactose (Fig. 15B).

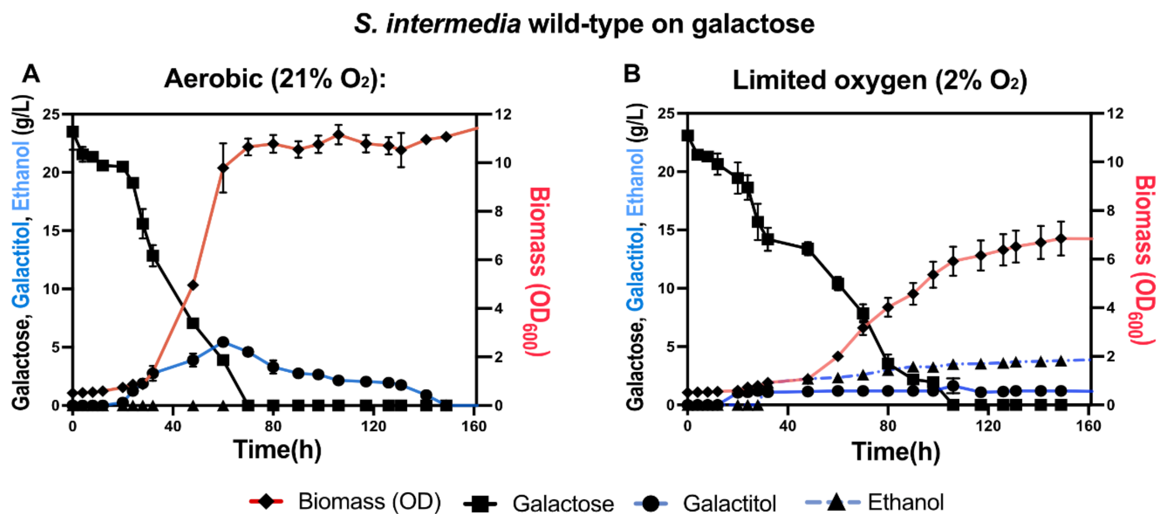

Figure 15. Experimental validation of *Sint*GEM predictions to understand galactitol overflow metabolism in *S. intermedia*. A & B) Growth curves and metabolite profiles for *S. intermedia* wild-type strain grown on galactose under different levels of oxygenation. Biomass in red (OD_{600}) is depicted on the right y-axis and the metabolites (Galactose, Galactitol and Ethanol in g/L) on the left y-axis, plotted against time. Figure taken from **Paper V.**

Secretion of galactitol by *S. intermedia* wild-type growing on galactose irrespective of the oxygen levels indicated that its production might not be strictly dictated by oxygen availability and is a potential overflow metabolite. Galactitol secretion in aerobic conditions may result from incomplete respiration of galactose entering the cell, likely due to an insufficient catabolic enzyme system. This limitation could stem from the high protein demand needed to support rapid growth rates, similar to other overflow metabolites [165].

Knowledge gained about lactose and galactose metabolism in *S. intermedia* and its trait of galactitol production paved way to a cell-factory application for this yeast in lactose-rich substrate, as explored in **Chapter 5**.

<u>Key take-away point from Chapter 4:</u>

1. *S. intermedia* CBS 141442 is an efficient lactose-metabolizing yeast.

2. The *GALLAC* cluster in *S. intermedia* forms a novel and unique regulatory system through interactions with the conserved *LAC* and *GAL* clusters, dissimilar to the regulatory systems seen in *S. intermedia*, *K. lactis*, and *C. albicans*.

3. The study of gene clusters suggests that *S. intermedia* may utilize an alternative galactose metabolic pathway, the oxidoreductive pathway. Detailed analysis using *Sint*GEM provided insights into cofactor balancing and overflow metabolism.

4. Our knowledge-matching approach using model predictions and experimental validation indicated that galactitol is an overflow metabolite in *S. intermedia*.

Chapter 5. Cell-factory application of *S. intermedia*

This chapter summarizes the work presented in **Paper V** on developing *S. intermedia* as a cell-factory using lactose-rich industrial side streams as raw material. It begins by detailing the strain engineering steps undertaken to enhance galactitol production in *S. intermedia*. The subsequent improvements in production metrics, including titer (the concentration of product formed), rate (the speed at which the product is produced), and yield (the efficiency of converting the feedstock into the desired product), are then discussed. Finally, the engineered strain and optimized process conditions are applied to test galactitol and tagatose production using cheese whey permeate as the feedstock.

5.1. Cell-factory design for sustainable bioproduction

Characterizing a strain's inherent metabolism, with the aid of gene editing and systems biology tools, is crucial for realizing its potential in cell-factory applications for bio-based production [9]. However, to fully achieve its production potential for industrial applications, further strain development and process optimization are often needed. In addition to metabolic engineering strategies, it is important to thoroughly evaluate strain performance under process conditions by measuring Titers, Rates, and Yields (TRY). Testing both the chassis strain, a stable platform for precursor production, and final strains in conditions mimicking actual production environments provides valuable insights for enhancing bioproduction efficiency and sustainability [37]. This is especially relevant when renewable raw materials, such as plant biomass or industrial side streams (e.g., from the dairy industry), are used in the production process.

Converting dairy byproducts into added-value products has gained attention in both science and industry. Even though whey protein is commercially recovered and used in beverages, sauces, infant foods and formula, and whey permeate is used in fertilizers, animal feed, and as bulking agents, large quantities of whey are still unused and end up in waste-water treatment plants [169, 170]. Studies using lactose-rich, de-proteinized whey rich as feedstock for fermentation have demonstrated its potential for producing various biofuels and bioproducts, including bioethanol, 2,3-butanediol, biogas, hydrogen, and polyhydroxyalkanoates [41].

5.1.1 Microbial production of galactitol and tagatose

Galactitol, a sugar alcohol or polyol, has applications in the pharmaceutical industry for producing di-anhydrogalactitol, an anti-tumor agent, [171] and in the food industry as a precursor of the natural sweetener tagatose [30]. Galactitol is commercially produced through the hydrogenation of galactose derived from plant biomass, but production is energy-intensive and costly. As an alternative, microbial-based production methods have been developed, including the use of NCYs such as *R. toruloides* using galactose and *Metschnikowia pulcherrima* on cheese whey [172, 173]. Tagatose, a derivative of galactitol, is a low-calory sweetener naturally found in the gum of *Sterculia setigera* trees, lichens, and in trace amounts in milk and other dairy products [174, 175]. Commercially, tagatose is produced via galactose isomerization, a process that suffers from low yields and the excessive use of metal hydroxides and acids, negatively affecting both cost and the environment [30]. Microbial production of tagatose has been demonstrated using isomerases [172], epimerase [176] and oxidoreductive enzymes [30] using substrates such as lactose or fructose which are derived from industrial side-streams.

5.1.2 Galactitol-producing chassis strain

Chassis strains are engineered through gene deletions or overexpressions to regulate the supply of precursors or cofactors and to facilitate substrate and product transport across cellular membranes. These strains provide a foundation for the expression of heterologous pathway genes, with the aim of overproducing the target metabolite. Further improvements can be achieved using the "push-pull-block" strategy, where precursor supply is enhanced ("push"), flux is directed towards the desired pathway ("pull"), and losses to competing pathways are prevented ("block") [177]. Applying this approach in a yeast strain grown on lactose allows carbon partitioning, with glucose used for cell maintenance and galactose for product formation, as demonstrated in *S. cerevisiae* for tagatose production. Although Lu and colleagues achieved a high tagatose yield (62.7% of the theoretical maximum) by overexpressing heterologous genes, the main limiting factor for further yield improvement was low availability of the precursor galactitol [30].

As outlined in **Chapter 4**, the *S. intermedia* strain *galΔ* lacking the Leloir pathway accumulated significant amounts of galactitol, achieving approximately 90% of the theoretical maximum yield (0.526 g galactitol/g consumed lactose) when grown on lactose in shake flasks. While this strain has potential as a chassis for tagatose production, galactitol was eventually consumed by the yeast,

reducing the availability of the precursor (**Chapter 4: Fig. 13C**). Deleting genes in the proposed oxidoreductive pathway was intended to block flux and ensure complete galactitol secretion. However, only the deletion of *XYL1_2* resulted in a noticeable growth defect on galactitol, indicating a regulatory role for this protein in galactitol metabolism (**Paper V: Fig. 7**). The double deletion mutant (galΔxyl1_2Δ) still produced galactitol on lactose, though at lower levels than the galΔ strain, indicating that while Xyl1_2 contributes to galactitol metabolism, it is not the only aldose reductase acting on galactose in *S. intermedia*. Further research is required to determine if galactitol assimilation occurs in this double mutant. Similar analyses were conducted for other downstream genes in the hypothesized oxidoreductive pathway. Double deletions of *LAD1*, *LXR4* and *XYL2* genes in the *galΔ* background did not fully inhibit growth on galactitol, warranting further investigation.

5.1.3 Tagatose production strain

In a previous study, galactitol-2-dehydrogenase-encoding gene *GDH* from *Rhizobium leguminosarum* (*RlGDH*), which exhibits high specific activity towards galactitol, was heterologously expressed in *S. cerevisiae* to produce tagatose [30, 178]. To enable the redirection (pull) of flux of galactitol towards tagatose production, Lu and colleagues expressed the *RlGDH* gene under a strong Sc*TDH3* promoter [30]. A similar approach was adopted for tagatose production in the *S. intermedia* galactitol chassis strain. Codon-adjusted *RlGDH* gene was synthesized and expressed under two strong endogenous promoters driving the expression of the genes *PET9* (ATP:ADP transport activity) and *PGA26* (filamentation and biofilm activity). The selection of these promoters was based on their high relative gene expression on five different carbon sources as derived from RNAseq gene expression data [64]. Using the CRISPR-Cas9 genome editing technique developed for *S. intermedia*, successful integration of *PET9_RlGDH_ACT1* cassette was achieved at the *ADE1* locus enabling the “pull” strategy for tagatose production in this strain (referred to as *galΔ ade1*::*GDH*) (Fig.16).

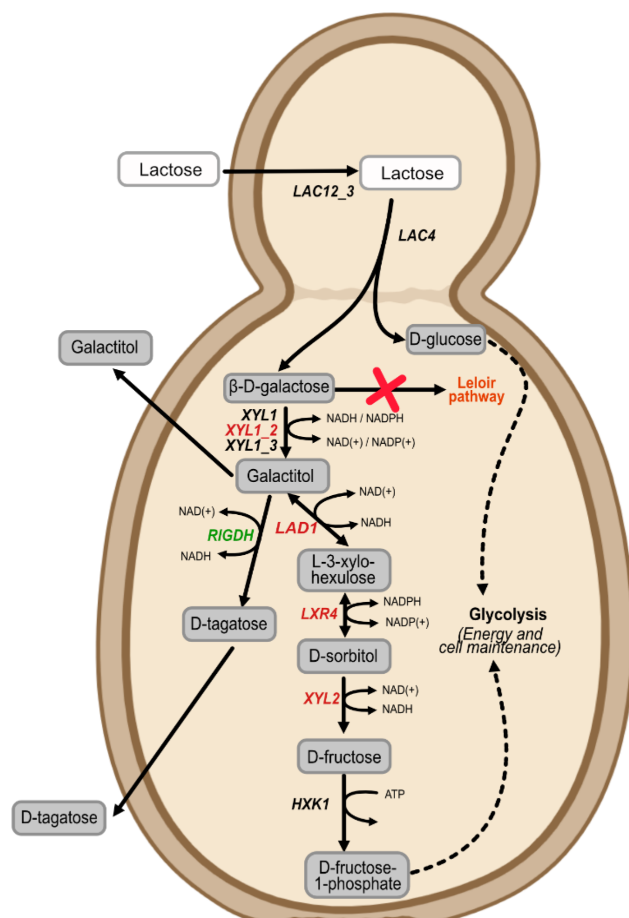

Figure 16. Illustration of "Push-Pull-Block" strategy for strain engineering employed in *S. intermedia* to produce galactitol and tagatose. Genes deleted to change flux by "push or block" are shown in red. Heterologous gene expressed in *S. intermedia* has been shown in green. Lactose uptake and hydrolysis is enabled by the *LAC* cluster genes *LAC12_3* and *LAC4*, following which glucose is metabolized via glycolysis for energy and cell maintenance. **Push**- The galactose moiety from lactose is prevented from entering the Leloir pathway by deleting the *GAL* cluster genes (indicated by the red cross), thus channeling the flux towards aldose reductase mediated conversion of β-d-galactose to galactitol. Galactitol was detected in the media, but mechanism of its transport is unknown. **Block** – The consumption of galactitol by *S. intermedia galΔ* indicated a possible oxidoreductive pathway downstream of the galactitol production step. To circumvent the consumption of the precursor, downstream genes *LAD1*, *LXR4*, *XYL2* were deleted individually in the *galΔ* background strain. **Pull**– Heterologous expression of the galactitol-2-dehydrogenase gene from *Rhizobium leguminosarum* directed flux towards the formation of tagatose.

5.2. Process improvement for efficient bioproduction

The bioprocess design for the *S. intermedia galΔ* strain with lactose as substrate, was guided by insights from *Sint*GEM *galΔ* model (Leloir pathway was blocked) (**Paper V**). Flux sampling under oxygen limitation also for *Sint*GEM *galΔ* model, consistent with findings from the wild-type model, suggested that oxygenation is a key factor in the secretion of redox-associated metabolites, particularly galactitol and ethanol. Experimental validation in batch reactors under low oxygen (2%

oxygen in inlet air) confirmed model predictions, showing that ethanol secretion was linked to low growth rates alongside galactitol production (**Paper V: S7**). On the other hand, aerobic cultivation of the *galΔ* strain did not show any ethanol secretion (**Paper V: S7**). These results suggested that the process of galactitol production should be aerobic, ensuring that as much carbon as possible is channeled into galactitol.

5.2.1 Titer

Titer represents the final concentration of the product and is one of the most widely used metrics to evaluate the performance of a microbial fermentation process [37]. The titer is determined by how much product is formed per unit of fermentation volume (g product/l cultivation medium). High titers are desired because they enhance the efficiency of downstream processing, where product purification takes place, and have a significant impact on the overall economics of the production process. Some of the ways to improve titers are: increasing yeast biomass (the number of cell factories available to produce the product), prolonging fermentation time, and increasing the substrate concentration [37]. It is important to note that an increase in the final titer is not only dependent on the strain's tolerance to high substrate concentrations, but also to high concentrations of the product. In the case of galactitol, this discriminates many yeast species as production hosts, as galactitol is toxic to several yeasts [179].

To improve galactitol titers, growth of the chassis strain was tested with substrate concentration (50 g/L lactose), resembling the composition of CW (Fig. 18). The experiments were conducted under aerobic conditions in a 1L bioreactor with a 500ml culture volume and resulted in a drastic increase in titer from the previously observed 9g/L to 21 g/L (**Paper V: Fig.7**), which indicates that *S. intermedia* can tolerate high galactitol concentrations.

5.2.2 Rate

Rate represents the productivity of the cell factory and can be measured as volumetric (product per volume and time) or biomass-specific productivity (product per biomass and time) [37]. It was observed that when precultured in glucose, *galΔ* grew after a long lag phase (~120 h) in minimal lactose media (20 g/L) (Fig. 13C). Since it has been observed that the lag phase shortens when cells have previously encountered the same carbon source due to rapid metabolic adaptation, we

hypothesized that preculturing in lactose would enhance productivity [180]. Preculturing the *galΔ* strain in lactose reduced the lag phase to ~32 h, improving productivity in shake flasks from 0.015 g/L h^{-1} to 0.042 g/L h^{-1} (**Paper V: Fig. 7**). Thus, preculture schemes using lactose as the sole carbon source were adopted for all experiments in bioreactors.

5.2.3 Yield

Yield represents the amount of product generated per substrate, either available or consumed, reflecting the efficiency of carbon conversion into the desired product. The many ways to increase the overall yield of the final product include optimization of strain, media and process conditions [181]. Using a stoichiometric model, it is possible to calculate the maximum yield in a cell that represents an upper boundary for the yield [37].

In addition to the secretion predictions under carbon and oxygen-limitation (as explored in **Chapter 4, Section 4.3.4**), flux sampling for *Sint*GEM wild-type model was also tested on galactose or lactose under nitrogen-limited condition. The results indicated that nitrogen limitation favored galactitol accumulation over all the different metabolites associated to the oxidoreductive pathway (Table 2).

Table 2: Percentage of simulations *Sint*GEM wild-type model secreting oxidoreductive pathway intermediates in FBA random sampling using *Sint*GEM at different limiting conditions. 10,000 simulations were run for each combination of carbon source and nutrient limitation.

Metabolite	**Galactose**			**Lactose**		
	C-limited	O2-limited	N-limited	C-limited	O2-limited	N-limited
Galactitol	0.8%	60.51%	0.33%	0.11%	100%	42.93%
D-tagatose	0.19%	0%	0%	0.13%	0%	0
D-sorbitol	0.04%	0%	0%	0.09%	0%	0
L-sorbose	0.03%	0%	0%	0.13%	0%	0

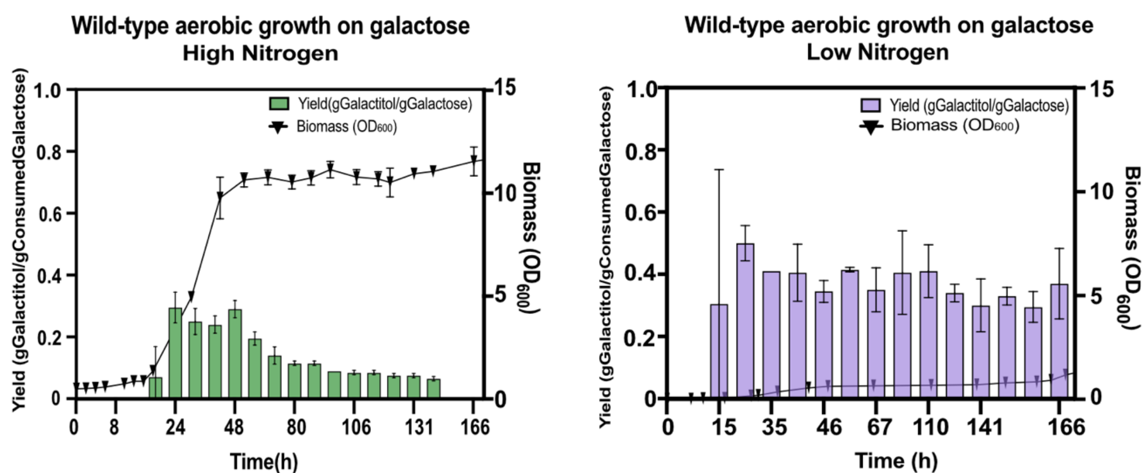

Figure 17. Experimental validation of *Sint*GEM flux sampling results in nitrogen limited condition. Graph depicts the yields of galactitol and the biomass of the wild-type strain in minimal media containing two different concentrations of ammonium sulphate (0.5% = high and 0.001% = low). The graphs represent Yield in $g_{Galactitol}/g_{consumedGalactose}$ on the left y-axis and the biomass (OD_{600}) formation on the right y-axis, plotted against Time(h) on the x-axis. Figure taken from **Paper V.**

5

Culturing the *S. intermedia* wild-type strain in both high and low nitrogen conditions (0.5% vs 0.001% nitrogen concentrations in minimal Verduyn media) revealed significant differences in galactitol yield (Fig. 17). Despite a notable growth deficit, galactitol yields were approximately three times higher in the low nitrogen condition compared to the high nitrogen condition. These findings aligned well with model results for predictions under nitrogen limitation condition where galactitol secretion becomes the preferred pathway for excess carbon disposal.

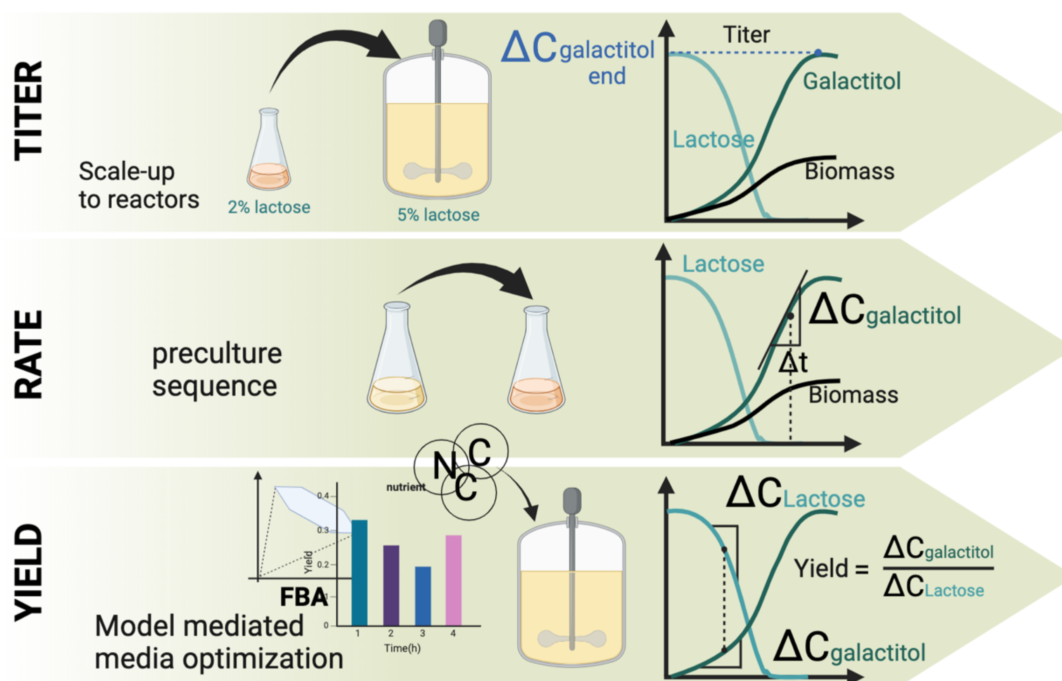

Figure 18: Strategies to improve TRY metrices of galactitol production process. TITER is determined by the total concentration of galactitol at the end of fermentation. Galactitol end titer was improved by using higher concentration of lactose (5% instead of 2%) and scale up from shake flask to reactor. RATE is defined as the change in galactitol produced over time. The total rate of production of the process was improved by changing the preculture sequence resulting in shorter lag phase and earlier production of galactitol. YIELD was estimated by the ratio of the change in galactitol produced over time by the change in lactose consumed. Rate for galactitol production was improved by reducing nitrogen content in the minimal media.

It is desirable to be able to increase all three TRY metrices for a sustainable bioprocess. However, often the impact of optimization for one parameter may affect another parameter as seen in the production process for galactitol. Although the increased starting lactose concentration in 1L bioreactor (from 2% to 5%) resulted in drastic increase in galactitol titer, it was observed with a decrease in yield as compared to that observed in the shake flasks (from >90% to 85%). Similarly, the change in preculturing sequence resulted in an improved productivity in shake flasks but accompanied by a drop in the yield to 85% of the theoretical maximum. Thus, careful assessment of each of the TRY metrices is necessary to establish a successful bioprocess.

5.3. Production of galactitol and tagatose using industrial side-stream as raw material

Galactitol production was assessed by growing the *galΔ* strain in cheese whey permeate with 60 g/L of lactose in fully aerated 1L bioreactors (Fig. 19). After 150h, only 5 g/L galactitol was obtained,

although the galactitol yield was high (~0.54 $g_{galactitol}/g_{consumed\ lactose}$). The yield remained stable throughout the duration of fermentation. This result was in accordance with the results from the previous experiment in low nitrogen reactors where the low nitrogen content corresponded to the range of available nitrogen observed in cheese whey permeate [27]. Upon addition of ammonium sulfate to the culture (dotted line in Fig. 19), growth rate, lactose uptake, and galactitol production rate increased significantly, reaching a final titer of 10.2 g/L. However, the galactitol yield remained unchanged, possibly due to the low final concentration of ammonium sulfate in the CWP and the limited cultivation duration post-addition. Combined, these results are compatible with model predictions about the effect of nitrogen on the production of galactitol and demonstrate the cell-factory potential of the *S. intermedia galΔ* strain.

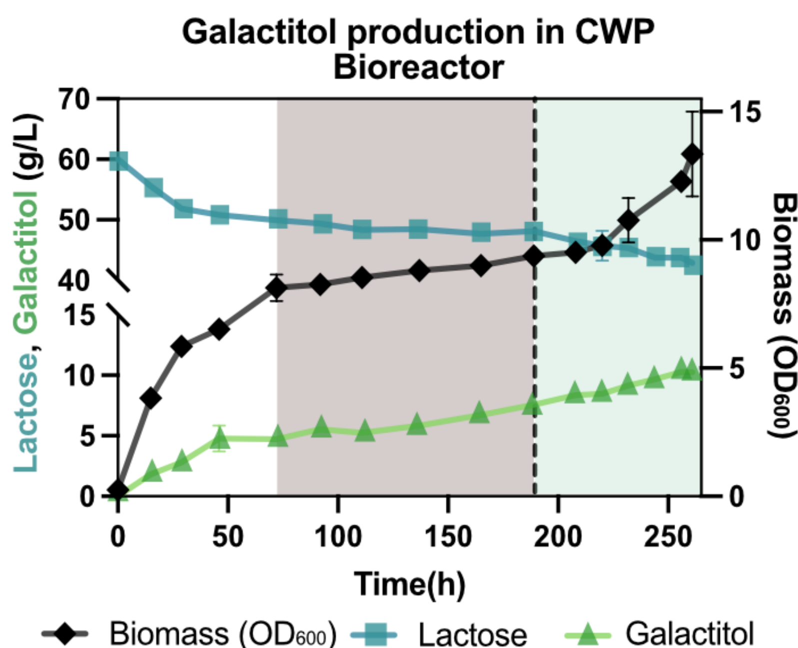

Figure 19. Galactitol production in 1L bioreactor containing CWP with starting lactose concentration of 60 g/L. Dotted line depicts the addition of ammonium sulphate to the culture. Shaded regions have been compared for change in galactitol yield, growth rate and lactose uptake rates to determine the effect of ammonium sulphate addition. Biomass (OD) is shown in black and plotted on right y-axis, lactose (blue) and galactitol (green) plotted on left y-axis, against time(h) on x-axis. Error bars represent standard deviation for two biological replicates. Figure taken from **Paper V**.

Similarly, tagatose production in the *galΔade1*::GDH strain was analyzed in shake flasks containing CWP with 6% lactose and supplemented with adenine (Fig. 20). The strain was able to produce 4.1 g/L galactitol and 1.3 g/L of tagatose, both detected extracellularly. These values amount to 60% of the theoretical maximum yield of galactitol and 30% of the theoretical maximum yield of tagatose based on lactose consumed. Though galactose-to-galactitol conversion is still relatively high, the low conversion efficiency of galactitol-to-tagatose indicates inefficient activity of the heterologous *Rl*GDH enzyme, necessitating further optimization.

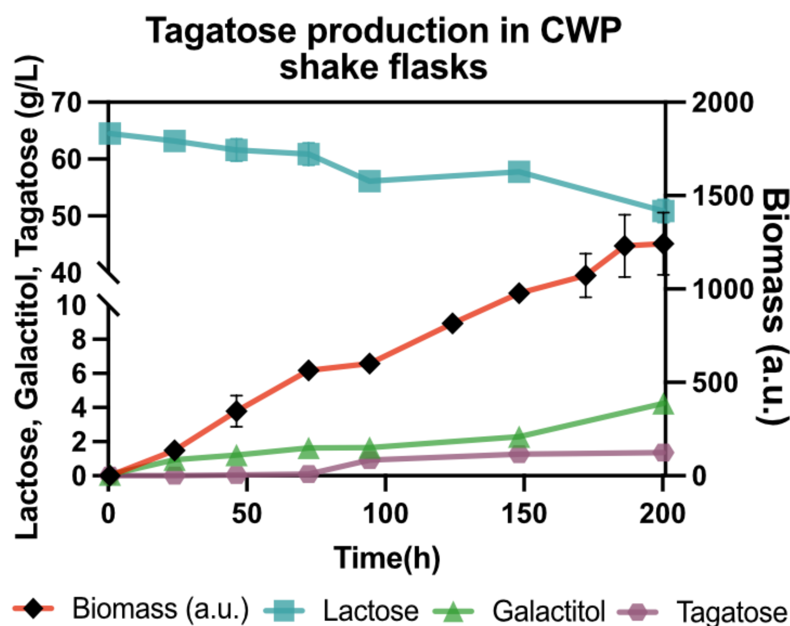

Figure 20. Tagatose production by *galΔade1*::*GDH* strain in shake flasks containing CWP with a starting lactose concentration of 60g/L. Biomass (a.u) is shown in red and plotted on right y-axis, lactose (blue), galactitol (light-green) and tagatose (purple) plotted on left y-axis, against time(h) on x-axis. Error bars represent standard deviation for two biological replicates. Figure taken from **Paper V**.

Key take-away point from Chapter 5:

1. The *S. intermedia* strain was engineered by deleting genes in the Leloir pathway, which resulted in galactitol overproduction. A heterologous galactitol-2-dehydrogenase gene was introduced to enable tagatose production from galactitol.

2. The chassis strain was assessed for galactitol production using the TRY metrices.

3. Aerobic conditions were found to be crucial for galactitol production, while nitrogen limitation significantly increased galactitol yield by promoting carbon flow towards galactitol secretion rather than growth.

4. The *S. intermedia galΔade1*::GDH strain demonstrated its potential in producing galactitol and tagatose from industrial dairy byproduct (cheese whey permeate), making the bioprocess a sustainable and circular solution for utilizing side-streams in industrial fermentation.

Chapter 6. Conclusions and outlook

6.1. Conclusions

This thesis provides a deeper understanding of yeast physiology and metabolism, with a focus on NCYs aimed at revalorizing industrial side-streams (Fig. 21). Investigating the yeast biodiversity of Africa's tropical niche revealed lactose-metabolizing yeasts with promising traits for industrial biotechnology applications. The experimental and computational tools developed in this thesis enabled a detailed exploration of *S. intermedia*, one of the few lactose-utilizing yeasts, revealing genetic and metabolic insights which paved way to its potential use as a cell factory for dairy industry side-streams.

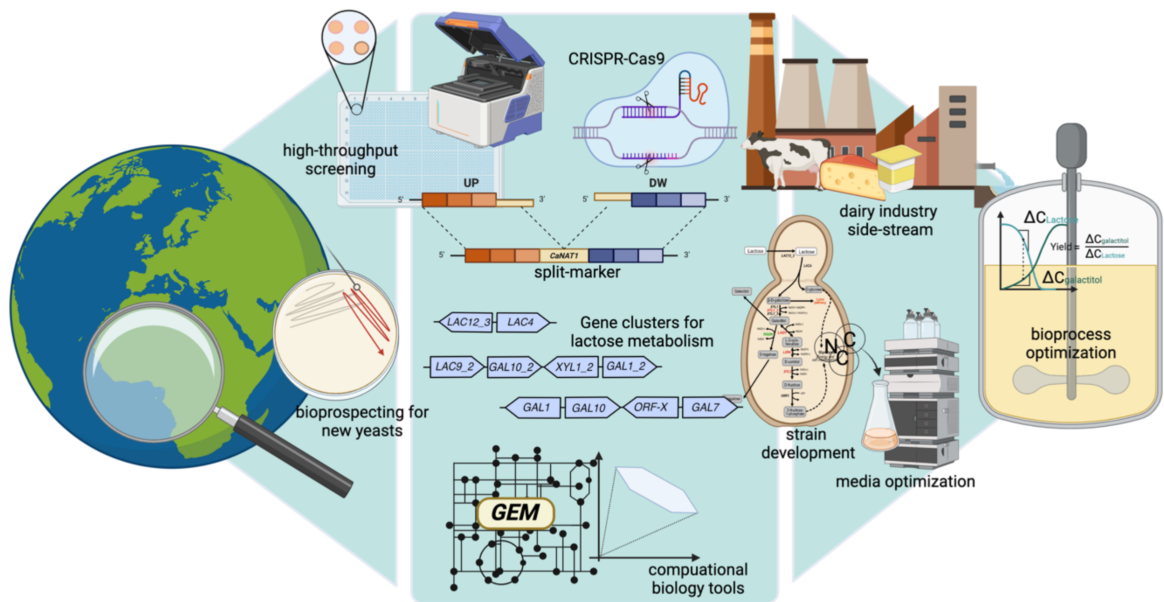

Figure 21: The illustration demonstrates the collective steps taken in this thesis for the exploration, characterization and application of a lactose metabolizing yeast (from left to right). Exploration of new yeasts is performed by bioprospecting for yeasts and subsequent screening by isolation for single colonies and their screening using HTP technology. Steps to characterize one of the lactose metabolizing yeasts involves development of genome editing tools (split-marker and CRISPR-Cas9 techniques) and computation biology tool (GEM) to elucidate the genetic (gene clusters) and metabolic determinants (pathway). Knowledge gained is applied in strain development, media and process optimization to use dairy industry side-stream for production of value-added products.

Overall, the work addressed the research questions posed in this thesis, as outlined below:

- *Which lactose-metabolizing yeasts, with traits suitable for cell-factory application, can be discovered from untargeted bioprospecting?*

Exploring the diversity of yeasts in different ecological niches can enable discovering candidates with interesting prospects for application in industrial biotechnology. The characterization of yeast diversity from Nigeria with a tropical ecosystem, provided an opportunity to explore the metabolic capabilities of 1,996 yeast strains. As assessed by HTP screening, approx. 27% of the total isolates could grow fast on lactose, with a doubling time shorter than 3 h. The bioprospecting approach was untargeted and dairy-rich niches were not sampled, indicating that yeasts isolated from lactose-free niches can have the genetic machinery to metabolize this disaccharide.

Identification of the lactose-metabolizing yeasts by ITS sequencing revealed that the two predominant species among the lactose-metabolizing yeast isolates were *S. intermedia* and *S. pseudointermedia*. Both strains were among the top growers on lab media containing lactose as well as the industrial side-stream cheese whey. In this study, the potential cell-factory applications of the lactose-growers was also assessed by testing growth and metabolite production profiles on industrial side-stream cheese whey. Lactose-metabolizing strains with a range of consumption rates belonged to both the basidiomycetous and ascomycetous yeasts, indicating their potential roles in whey valorization. Multiple strains were found to produce industrially relevant compounds, such as ethanol, organic acids, and lipids.

Overall, the work described in **Chapter 2** of this thesis highlights the potential of untargeted bioprospecting to discover known or novel yeast species from diverse ecological niches, with the potential to be developed into future cell factories for bio-based production.

- *Having discovered a new strain with interesting properties, what tools must be developed to enable genetic and metabolic characterization of a NCY like the lactose-metabolizing* S. intermedia*?*

Further exploration of newly discovered yeasts, which pose as potential model yeasts or cell-factories may require development of new genetic engineering and systems biology tools [82].

Development of these methods in *S. intermedia* required a step-by-step analysis of suitable transformation protocols and selection markers. The split-marker approach significantly improved

targeting frequency in *S. intermedia* enabling construction of deletion mutants. In addition, the CRISPR-Cas9 technique was used successfully to construct indels at high targeting efficiency as well as enable marker-free integration of heterologous genes in *S. intermedia*. However, low targeting rates for integration indicated a need to better understand double strand break repair mechanisms in NCYs. Additionally, *in silico* assessment of *S. intermedia* metabolism was enabled by the reconstruction of the first curated GEM, called *Sint*GEM. Draft *Sint*GEM based on Yeast8 model for *S. cerevisiae* was improved by manual curation and incorporation of gene expression data of *S. intermedia*. The stoichiometric model enabled the analysis of *S. intermedia* metabolism by reporter metabolite analysis, FBA and flux sampling which were crucial for the *in sillico* exploration of *S. intermedia* lactose and galactose metabolism, as well as its cell-factory potential on dairy industry side-streams. However, the limitation of GEMs to capture regulatory mechanisms restricts somewhat the usefulness of *Sint*GEM to completely understand regulation of lactose and galactose metabolism in this yeast.

- *What are the genetic and metabolic determinants of lactose metabolism in S. intermedia?*

Leveraging the genome editing and computational tools developed for *S. intermedia*, an in-depth investigation into its lactose and galactose metabolism revealed novel genetic and regulatory elements. A unique *GALLAC* gene cluster, containing homologs of genes involved in lactose and galactose metabolism in model yeasts, coordinates a distinct regulatory mechanism in *S. intermedia*. Unlike the system in *K. lactis*, *S. cerevisiae*, and *C. albicans*, lactose and galactose metabolism are primarily controlled by two key genes in the *GALLAC* cluster: *GAL1_2* and *LAC9_2*. Gal1_2 seems to function as a transcriptional regulator of the *LAC* cluster in *S. intermedia*, while *LAC9_2* regulates the *GALLAC* cluster genes. Based on the detection of galactitol in the growth medium, an alternate galactose catabolic pathway via galactitol was hypothesized. *SintGEM* predictions combined with data from continuous and batch bioreactors, revealed that *S. intermedia* uses upstream enzymes of the oxidoreductive pathway producing galactitol as a carbon overflow metabolite, linked to redox cofactor balancing during both lactose and galactose metabolism.

- *How can the knowledge gained for S. intermedia enable revalorization of lactose-rich industrial side-stream to produce added-value compounds?*

Strain engineering and process optimization steps were performed to further improve the ability of the galactitol chassis strain to overproduce the industrially relevant precursor galactitol upon growth on lactose. Process metrices titer, rate and yield for galactitol production were determined and improved by changing preculture sequence, increasing starting substrate concentration and media optimization, respectively. While determining galactitol production on industrial side-stream CWP as the substrate, the final titer of 10.2 g/L, highest rate of 0.035 g/L h^{-1} and 0.54 $g_{galactitol}/g_{consumed\ lactose}$ (>95% of the theoretical maximum) were achieved. Furthermore, the “push-pull-block” strategy of metabolic engineering further enabled the construction of a production strain for the natural sweetener tagatose. However, tagatose production using CWP resulted in 1.3 g/L of tagatose achieving yield of a 30% of the theoretical maximum but low productivity of <0.01 g/L h^{-1}, indicating further strain engineering and process optimization are necessary to boost tagatose production by *S. intermedia* on CWP.

6.2. Outlook

Advances in microbial ecology and biotechnology show promise in providing innovative and alternative methods to address the challenges of the established linear economic models and facilitate transition to a circular bioeconomy. It is imperative that efforts towards enabling the adoption of technologies for biobased production are accelerated.

The discovery of microbial species with industrially relevant traits continues to expand our understanding of global microbial diversity. While this thesis used an untargeted bioprospecting approach in terms of sampling from diverse ecological niches, also targeted sampling from rarely explored niches holds great potential for uncovering yeast species with genetic and metabolic traits that enhance host performance. As demonstrated with *S. intermedia*, discovering new microorganisms, enabled by bioprospecting, can bridge nature-based solutions with synthetic biology by uncovering novel microbial cell factories. The observation that most of the tested strains accumulated biomass on whey indicated that a better screening strategy for a particular metabolite may be achieved by optimization of cultivation media, for example increasing the C/N ratio for lipid production. Similarly, broader metabolomic analyses may uncover a wider range of molecules with industrial relevance. However, large-scale bioprospecing efforts are labour intensive and can benefit by the adoption of automation and data-driven methods as exhibited in this thesis. Advanced large-scale sequencing technologies would further enable faster and more accurate classification of microbial isolates and provide essential genetic insights for research. These approaches can facilitate the rapid identification of strains with desirable traits, a critical need in industries such as wine and craft brewing, where demand for diverse natural products is growing [50, 55]. Overall, these points show the importance of biodiversity conservation to preserve ecosystems, species, and genetic diversity that are in turn, fundamental to combat problems of climate change, ecological imbalance and depleting natural resources.

Tools developed in this thesis for both *in vivo* and *in silico* exploration of *S. intermedia* demonstrate the advantages of novel methods for new species. Improved predictability, refined models, and efficient genome-editing tools are crucial for faster strain development. Enhancing yeast identification and classification methods will enable toolboxes that leverage genetic components from related species. Efforts to create orthogonal systems, where synthetic biology tools can be easily transferred between NCYs, are already underway and will accelerate NCY exploration. Modeling NCY cellular functions at the genome scale is limited by gaps in knowledge about gene function and the availability of kinetic and physicochemical parameters for enzymes.

However, *Sint*GEM showed significant improvements in predictive accuracy when validated by experimental data, and incorporating additional omics, such as proteomics and metabolomics, will further enhance predictions. The high-quality reconstruction of *S. intermedia* using a refined Yeast8 model, based on homology curation from a wide range of fungal species, highlights advancements in systems biology tools. These tools can help overcome the commercial limitations of NCYs and enable the application of engineering strategies that are typically reserved for model organisms.

Microorganisms have evolved signaling networks to prioritize carbon source utilization, and studying gene regulation in fungi has greatly advanced our understanding of eukaryotic gene expression and metabolism. This thesis highlights the importance of understanding the regulatory and metabolic networks of NCYs, such as *S. intermedia*, to unlock their full potential as cell-factories but also potential model yeasts to understand lesser studied mechanism of fungal sugar regulation. Being phylogenetically distant from the well studied model organisms *K. lactis* and *S. cerevisiae*, *S. intermedia* puts forward an example of a novel regulatory system for lactose and galactose metabolism. Further elucidation of this unique regulatory system is possible by incorporating enzyme characterization, transcription factor identification (e.g., via DNA affinity purification sequencing), and protein interaction studies. Recent work on machine learning based elucidation of genes associated to the oxidoreductive pathway, with galactitol as a key metabolite [182], aligns well with the work discussed in this thesis. Combination of such research, whether it involves large scale exploration of 1000's of species or in depth-understanding of a single yeast strain, can lead to discovery of new metabolic pathways and filling knowledge gaps on enzyme functions and regulatory networks.

Despite significant advances in bioprocesses aimed at replacing fossil-based production of chemicals, many challenges remain before microbial-based solutions can fully replace fossil-derived methods. As shown with *S. intermedia*, strain improvement can be achieved through innovative experimental and computational techniques which ultimately can enable reaching higher TRY metrices. For example, adaptive laboratory evolution could enhance tolerance to high sugar concentrations to increase the titers by enabling the use of yet another dairy industry side-stream called de-lactose permeate, which contains a higher lactose concentration (approx. 20%) than those tested in this project. Strain engineering could also benefit from artificial intelligence to predict the catalytic activity of heterologous proteins improving productivity. Future research could explore cofactor engineering, such as using transhydrogenases to interconvert electron carriers or employing non-canonical cofactors for reactions in the oxidoreductive pathway and lead to product diversification. While this work tested only two promoter combinations to enable tagatose production in *S. intermedia*, exploration of both constitutive and inducible endogenous promoters

could be highly useful for further strain engineering. Although the focus of this project was on utilizing dairy industry side-streams, the approach of understanding yeast metabolism and applying it to specific applications can be extended to any newly discovered yeast species.

// Acknowledgements

Moving to Sweden right in the midst of a pandemic to embark on this PhD journey was not an easy decision. However, I can say for sure that these 4 years have been amazing, filled with valuable lessons and beautiful moments, thanks to the nice and loving people I have met in this part of the world.

First and foremost, my heartful thanks to my supervisor, ***Cecilia***. I am grateful for your guidance and support over these years. I still remember the first day when I moved to Olofshöjd, and you gave me a bag full of supplies to help me settle into Covid-hit-Sweden. Your support has been monumental in countless situations, whether in designing a project I could readily embark on or helping me deal with my visa issues. I highly appreciate your critical questions, which improved the quality of my work, as well as the freedom to try my hands on new technologies such as CRISPR and GEM. I am also grateful for the experience beyond the PhD research, from writing the patent to the startup journey with GALAXY! Thank you!

This thesis exemplifies of how collaborative work can bring together diverse expertise to produce good research. Thanks a lot, to my collaborators ***Lisbeth*** *(and co-supervisor)*, ***Johan*** and ***Eduard*** for the nice discussions and immense support. Thanks to ***Fabio Oliveira***, whom I have met the least often but whose contributions have been pivotal in shaping this project! Thanks also to master students ***Abril*** and ***Marta P.*** whom I had the pleasure to work with and for your respective contributions to the project.

Thanks to my examiner, ***Calle*** for not only your guidance during these years but also the timely support you have given in my panic moments of missing credits to transfer or forgotten visa application! Special thanks to ***Punchalee*** and ***Vijay***, for all the help at work and the fun outside. I highly appreciate your patience to deal with my sudden requests, whether with HPLC or Dasgip and always guiding me with a smile!

I would like to also thank the researchers with whom I got hands-on experience in learning new skills, a part of which, I could present in this thesis: ***Hanna***, ***Karl***, ***Le***, ***Tomas,*** and ofcourse, the one who is obsessed with a systems approach: ***Ivan*** - you have been a teacher and a great friend. I am more than happy that after Wageningen, our paths crossed here in Gothenburg again!

Since moving to Gothenburg, I've had the pleasure of enjoying the company of colleagues at both IndBio and Sysbio.

From the start, with the three of us in that small room in Olofshöjd cooking and drinking together: ***Chaeho*** and ***Fabio***, those days assured me that our time in Sweden would be great. Thank you *Imma*, for the good times we have shared and for the many delicious Korean meals you prepared as Chef Im. ***Fabio***, thank you for always being there with/for me (literally, in the office), through all the courses, the teaching experiences, discussing about our travel plans to Germany/Netherlands, feeding me pasta regularly and for the continuous supply of *Nduja*!

Someone whom I cannot thank enough: my dear *Hermanski*. All our conversations - amazing and ridiculous, scientific and non-scientific, uplifting and depressing, at ***Mauro***/Kamesh-gatan - are some memories I take with me. Thank you!

Also, special mention: my dear Amiga! ***Nathalia***, thank you for always being there to listen to my concerns and reassure me with your wise words that everything is/will be OK (*ta OK!*).

After Covid, the kitchen became busier, and our evenings were packed with dinner plans, Yaki-trains, Bollywood-nights, F1 & DnD evenings, and travels (Denmark, Singapore and ski trips), adding to the memories I cherish. Thanks a lot: ***Tom, Angelo, Andrea, Simo, Daniel, Arne, Sophie***, ***Mauri, Luca, Ibai, Cecio, Andrea*** *(Nuclear),* ***Punchalee, Andrés*** *(Brazillian),* ***Vero*** *(Italian1),* ***Rexy, Vero*** *(Italian2),* ***Veronica*** *(Dutch1),* ***Andrea, Giada, Marta, Oliver, Dany, Luisa, Marta C., Julie, Paula, Xiang, Xiaozhi***, ***Jiwei, Monika, Gustav, Vaskar, Amanda, Max, Emil*** for adding your energy to these moments.

There are so many of you from both IndBio and Sysbio whom even though I don't mention here have contributed to making my PhD journey fun and fulfilling. Thank you very much!

Although I left the Netherlands, I never truly left, did I? My friends there, whom I consider my family, have always supported me. Thanks a lot ***Fede*** for being there with me, I know that you know how much it means to me. My ***Parea***, thanks for encouraging me always and to my dear Shaakaal – "*Papu pass ho gaya!*". My ***Delft people***, thanks a ton for the good times and the lovely reunions.

Lastly, I cannot end this without thanking my family and friends in **India**, for their continuous support! Love you guys!

Thank you, everyone - *bahut bahut Dhanyawaad*!

References

1. Siekmann, F., H. Schlör, and S. Venghaus, *Linking sustainability and the Fourth Industrial Revolution: a monitoring framework accounting for technological development.* Energy, Sustainability and Society, 2023. **13**(1): p. 26.
2. Isbell, F., et al., *Linking the influence and dependence of people on biodiversity across scales.* Nature, 2017. **546**(7656): p. 65-72.
3. Cowan, A.E., et al., *Microbial production of fuels, commodity chemicals, and materials from sustainable sources of carbon and energy.* Current Opinion in Systems Biology, 2023. **36**: p. 100482.
4. Thursby, E. and N. Juge, *Introduction to the human gut microbiota.* Biochem J, 2017. **474**(11): p. 1823-1836.
5. Byrd, A.L., Y. Belkaid, and J.A. Segre, *The human skin microbiome.* Nature Reviews Microbiology, 2018. **16**(3): p. 143-155.
6. Nielsen, J., *Yeast Systems Biology: Model Organism and Cell Factory.* Biotechnology Journal, 2019. **14**(9): p. 1800421.
7. Nielsen, J., C.B. Tillegreen, and D. Petranovic, *Innovation trends in industrial biotechnology.* Trends in Biotechnology, 2022. **40**(10): p. 1160-1172.
8. De Wannemaeker, L., I. Bervoets, and M. De Mey, *Unlocking the bacterial domain for industrial biotechnology applications using universal parts and tools.* Biotechnology Advances, 2022. **60**: p. 108028.
9. Patra, P., et al., *Recent advances in systems and synthetic biology approaches for developing novel cell-factories in non-conventional yeasts.* Biotechnology Advances, 2021. **47**: p. 107695.
10. Füting, P., et al., *Filamentous fungal applications in biotechnology: a combined bibliometric and patentometric assessment.* Fungal Biol Biotechnol, 2021. **8**(1): p. 23.
11. Meyer, V., et al., *Current challenges of research on filamentous fungi in relation to human welfare and a sustainable bio-economy: a white paper.* Fungal Biology and Biotechnology, 2016. **3**(1): p. 6.
12. Hong, K.-K. and J. Nielsen, *Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries.* Cellular and Molecular Life Sciences, 2012. **69**(16): p. 2671-2690.
13. Nielsen, J., et al., *Metabolic engineering of yeast for production of fuels and chemicals.* Current Opinion in Biotechnology, 2013. **24**(3): p. 398-404.
14. Nielsen, J., *Systems Biology of Metabolism.* Annual Review of Biochemistry, 2017. **86**(1): p. 245-275.
15. Geijer, C., R. Ledesma-Amaro, and E. Tomas-Pejo, *Unraveling the potential of non-conventional yeasts in biotechnology.* FEMS Yeast Res, 2022. **22**(1).
16. Liu, Z., et al., *Different expression systems for production of recombinant proteins in Saccharomyces cerevisiae.* Biotechnology and Bioengineering, 2012. **109**(5): p. 1259-1268.
17. Sleep, D., G.P. Belfield, and A.R. Goodey, *The Secretion of Human Serum Albumin from the Yeast Saccharomyces cerevisiae Using Five Different Leader Sequences.* Bio/Technology, 1990. **8**(1): p. 42-46.
18. Otero, J.M., et al., *Industrial Systems Biology of Saccharomyces cerevisiae Enables Novel Succinic Acid Cell Factory.* PLOS ONE, 2013. **8**(1): p. e54144.
19. Steen, E.J., et al., *Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol.* Microbial Cell Factories, 2008. **7**(1): p. 36.

20. Zhang, J., et al., *A microbial supply chain for production of the anti-cancer drug vinblastine.* Nature, 2022. **609**(7926): p. 341-347.
21. Ro, D.-K., et al., *Production of the antimalarial drug precursor artemisinic acid in engineered yeast.* Nature, 2006. **440**(7086): p. 940-943.
22. Hittinger, C.T., J.L. Steele, and D.S. Ryder, *Diverse yeasts for diverse fermented beverages and foods.* Curr Opin Biotechnol, 2018. **49**: p. 199-206.
23. Epremyan, K.K., et al., *Altered Mitochondrial Morphology and Bioenergetics in a New Yeast Model Expressing Aβ42.* Int J Mol Sci, 2023. **24**(2).
24. Beopoulos, A., T. Chardot, and J.-M. Nicaud, *Yarrowia lipolytica: A model and a tool to understand the mechanisms implicated in lipid accumulation.* Biochimie, 2009. **91**(6): p. 692-696.
25. Epremyan, K.K., D.V. Mamaev, and R.A. Zvyagilskaya, *Alzheimer's Disease: Significant Benefit from the Yeast-Based Models.* Int J Mol Sci, 2023. **24**(12).
26. Monteiro de Oliveira, P., et al., *Screening and Growth Characterization of Non-conventional Yeasts in a Hemicellulosic Hydrolysate.* Frontiers in Bioengineering and Biotechnology, 2021. **9**.
27. Estrada, M., et al., *Utilization of salt-rich by-products from the dairy industry as feedstock for recombinant protein production by Debaryomyces hansenii.* Microbial Biotechnology, 2023. **16**(2): p. 404-417.
28. Lane, M.M. and J.P. Morrissey, *Kluyveromyces marxianus: A yeast emerging from its sister's shadow.* Fungal Biology Reviews, 2010. **24**(1): p. 17-26.
29. Ravn, J.L., et al., *Yeasts Have Evolved Divergent Enzyme Strategies To Deconstruct and Metabolize Xylan.* Microbiol Spectr, 2023. **11**(3): p. e0024523.
30. Liu, J.J., et al., *Overcoming the thermodynamic equilibrium of an isomerization reaction through oxidoreductive reactions for biotransformation.* Nat Commun, 2019. **10**(1): p. 1356.
31. Li, X., et al., *Mining natural products for advanced biofuels and sustainable bioproducts.* Current Opinion in Biotechnology, 2023. **84**: p. 103003.
32. Bernard, A., T. Rossignol, and Y.-K. Park, *Biotechnological approaches for producing natural pigments in yeasts.* Trends in Biotechnology, 2024.
33. Ma, J., et al., *Yeast proteins: The novel and sustainable alternative protein in food applications.* Trends in Food Science & Technology, 2023. **135**: p. 190-201.
34. Thomsen, P.T., et al., *Beet red food colourant can be produced more sustainably with engineered Yarrowia lipolytica.* Nature Microbiology, 2023. **8**(12): p. 2290-2303.
35. Kulagina, N., et al., *Yeasts as Biopharmaceutical Production Platforms.* Frontiers in Fungal Biology, 2021. **2**.
36. Cao, J., et al., *Versatile and on-demand biologics co-production in yeast.* Nature Communications, 2018. **9**(1): p. 77.
37. Konzock, O. and J. Nielsen, *TRYing to evaluate production costs in microbial biotechnology.* Trends in Biotechnology.
38. Mabrouki, J., et al., *The dairy biorefinery: Integrating treatment process for Tunisian cheese whey valorization.* Chemosphere, 2022. **293**: p. 133567.
39. Patel, A., et al., *Biological treatment of pulp and paper industry effluent by oleaginous yeast integrated with production of biodiesel as sustainable transportation fuel.* Journal of Cleaner Production, 2017. **142**: p. 2858-2864.
40. Zotta, T., et al., *Valorization of cheese whey using microbial fermentations.* Applied Microbiology and Biotechnology, 2020. **104**(7): p. 2749-2764.
41. Colacicco, M., et al., *Process scale-up simulation and techno-economic assessment of ethanol fermentation from cheese whey.* Biotechnology for Biofuels and Bioproducts, 2024. **17**(1): p. 124.
42. Lappa, I.K., et al., *Cheese Whey Processing: Integrated Biorefinery Concepts and Emerging Food Applications.* Foods, 2019. **8**(8): p. 347.

43. Stasinakis, A.S., P. Charalambous, and I. Vyrides, *Dairy wastewater management in EU: Produced amounts, existing legislation, applied treatment processes and future challenges.* Journal of Environmental Management, 2022. **303**: p. 114152.
44. Chandra, R., et al., *A biorefinery approach for dairy wastewater treatment and product recovery towards establishing a biorefinery complexity index.* Journal of Cleaner Production, 2018. **183**: p. 1184-1196.
45. Raak, N., et al., *Blending side streams. A potential solution to reach a resource efficient, circular, zero-waste food system.* Future Foods, 2022. **6**: p. 100207.
46. Estrada, M., et al., *Open (non-sterile) cultivations of Debaryomyces hansenii for recombinant protein production combining industrial side-streams with high salt content.* New Biotechnology, 2023. **78**: p. 105-115.
47. Shen, X.X., et al., *Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum.* Cell, 2018. **175**(6): p. 1533-1545 e20.
48. Groenewald, M., et al., *Census of Yeasts Isolated from Natural Ecosystem and Conserved in Worldwide Collections*, in *Yeasts in Natural Ecosystems: Diversity*, P. Buzzini, M.-A. Lachance, and A. Yurkov, Editors. 2017, Springer International Publishing: Cham. p. 455-476.
49. Boekhout, T., et al., *Trends in yeast diversity discovery.* Fungal Diversity, 2022. **114**(1): p. 491-537.
50. Alperstein, L., et al., *Yeast bioprospecting versus synthetic biology—which is better for innovative beverage fermentation?* Applied Microbiology and Biotechnology, 2020. **104**(5): p. 1939-1953.
51. Hosseini, H., et al., *Marine microbial bioprospecting: Exploitation of marine biodiversity towards biotechnological applications—a review.* Journal of Basic Microbiology, 2022. **62**(9): p. 1030-1043.
52. Nimsi, K.A., et al., *Diversity, distribution, and bioprospecting potentials of manglicolous yeasts: a review.* FEMS Microbiology Ecology, 2023. **99**(5).
53. Vergara Alvarez, S.C., et al., *Bioprospecting of the probiotic potential of yeasts isolated from a wine environment.* Fungal Genetics and Biology, 2023. **164**: p. 103767.
54. Aguiar-Cervera, J.E., D. Delneri, and O. Severn, *A high-throughput screening method for the discovery of Saccharomyces and non-Saccharomyces yeasts with potential in the brewing industry.* Eng Biol, 2021. **5**(3): p. 72-80.
55. Holt, S., et al., *Bioflavoring by non-conventional yeasts in sequential beer fermentations.* Food Microbiology, 2018. **72**: p. 55-66.
56. Smith, M.T. and M. Groenewald, *The treasure trove of yeast genera and species described by Johannes van der Walt (1925-2011).* IMA Fungus, 2012. **3**(2): p. 179-87.
57. Vong, W.C., et al., *High-throughput screening for aroma production in food fermentations.* Food Research International, 2024. **177**: p. 113902.
58. Wallach, I., et al., *AI is a viable alternative to high throughput screening: a 318-target study.* Scientific Reports, 2024. **14**(1): p. 7526.
59. Najah, M., et al., *Droplet-Based Microfluidics Platform for Ultra-High-Throughput Bioprospecting of Cellulolytic Microorganisms.* Chemistry & Biology, 2014. **21**(12): p. 1722-1732.
60. Zackrisson, M., et al., *Scan-o-matic: High-Resolution Microbial Phenomics at a Massive Scale.* G3 (Bethesda), 2016. **6**(9): p. 3003-14.
61. Schoch, C.L., et al., *Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for <i>Fungi</i>.* Proceedings of the National Academy of Sciences, 2012. **109**(16): p. 6241-6246.
62. Vu, D., et al., *DNA barcoding analysis of more than 9 000 yeast isolates contributes to quantitative thresholds for yeast species and genera delimitation.* Stud Mycol, 2016. **85**: p. 91-105.

63. Buzzini, P., M.-A. Lachance, and A. Yurkov, *Yeasts in natural ecosystems: diversity*. 2017: Springer.
64. Geijer, C., et al., *Genomic and transcriptomic analysis of Candida intermedia reveals the genetic determinants for its xylose-converting capacity.* Biotechnol Biofuels, 2020. **13**: p. 48.
65. Moreno, A.D., et al., *Evolutionary engineered Candida intermedia exhibits improved xylose utilization and robustness to lignocellulose-derived inhibitors and ethanol.* Appl Microbiol Biotechnol, 2019. **103**(3): p. 1405-1416.
66. Moreno, A.D., et al., *Complete Genome Sequences of the Xylose-Fermenting Candida intermedia Strains CBS 141442 and PYCC 4715.* Genome Announc, 2017. **5**(14).
67. Peri, K.V.R., et al., *Split-marker-mediated genome editing improves homologous recombination frequency in the CTG clade yeast Candida intermedia.* FEMS Yeast Res, 2023. **23**.
68. Peri, K.V.R., et al., *A unique metabolic gene cluster regulates lactose and galactose metabolism in the yeast <i>Candida intermedia</i>.* Applied and Environmental Microbiology. **0**(0): p. e01135-24.
69. Liu, F., et al., *Phylogenomic analysis of the Candida auris-Candida haemuli clade and related taxa in the Metschnikowiaceae, and proposal of thirteen new genera, fifty-five new combinations and nine new species.* Persoonia, 2024. **52**: p. 22-43.
70. Xu, G., et al., *Fumaric acid production in Saccharomyces cerevisiae by in silico aided metabolic engineering.* PLoS One, 2012. **7**(12): p. e52086.
71. Passoth, V., et al., *Oleaginous yeasts for biochemicals, biofuels and food from lignocellulose-hydrolysate and crude glycerol.* Yeast, 2023. **40**(8): p. 290-302.
72. Vuong, P., S. Chong, and P. Kaur, *The little things that matter: how bioprospecting microbial biodiversity can build towards the realization of United Nations Sustainable Development Goals.* npj Biodiversity, 2022. **1**(1): p. 4.
73. de Almeida, E.L.M., E.J. Kerkhoven, and W.B. da Silveira, *Reconstruction of genome-scale metabolic models of non-conventional yeasts: current state, challenges, and perspectives.* Biotechnology and Bioprocess Engineering, 2024. **29**(1): p. 35-67.
74. Papon, N., et al., *Deus ex Candida genetics: overcoming the hurdles for the development of a molecular toolbox in the CTG clade.* Microbiology (Reading), 2012. **158**(Pt 3): p. 585-600.
75. Cannan, W.J. and D.S. Pederson, *Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin.* Journal of Cellular Physiology, 2016. **231**(1): p. 3-14.
76. Wilson, T.E., *Nonhomologous End-Joining: Mechanisms, Conservation and Relationship toIllegitimate Recombination*, in *Molecular Genetics of Recombination*, A. Aguilera and R. Rothstein, Editors. 2007, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 487-513.
77. Decottignies, A., *Microhomology-Mediated End Joining in Fission Yeast Is Repressed by Pku70 and Relies on Genes Involved in Homologous Recombination.* Genetics, 2007. **176**(3): p. 1403-1415.
78. Sfeir, A. and L.S. Symington, *Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?* Trends Biochem Sci, 2015. **40**(11): p. 701-714.
79. Fontana, G.A., et al., *Shepherding DNA ends: Rif1 protects telomeres and chromosome breaks.* Microb Cell, 2018. **5**(7): p. 327-343.
80. Mathiasen, D.P. and M. Lisby, *Cell cycle regulation of homologous recombination in Saccharomyces cerevisiae.* FEMS Microbiology Reviews, 2014. **38**(2): p. 172-184.
81. Boulton, S.J. and S.P. Jackson, *Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways.* The EMBO Journal, 1996. **15**(18): p. 5093-5103.
82. Lobs, A.K., C. Schwartz, and I. Wheeldon, *Genome and metabolic engineering in non-conventional yeasts: Current advances and applications.* Synth Syst Biotechnol, 2017. **2**(3): p. 198-207.

83. Shen, J., W. Guo, and J.R. Kohler, *CaNAT1, a heterologous dominant selectable marker for transformation of Candida albicans and other pathogenic Candida species.* Infect Immun, 2005. **73**(2): p. 1239-42.
84. Kawai, S., W. Hashimoto, and K. Murata, *Transformation of Saccharomyces cerevisiae and other fungi: methods and possible underlying mechanism.* Bioeng Bugs, 2010. **1**(6): p. 395-403.
85. Potter, H., *[34] Application of electroporation in recombinant DNA technology*, in *Methods in Enzymology*. 1993, Academic Press. p. 461-478.
86. Ito, H., et al., *Transformation of intact yeast cells treated with alkali cations.* J Bacteriol, 1983. **153**(1): p. 163-8.
87. De Backer, M.D., et al., *Transformation of Candida albicans by electroporation.* Yeast, 1999. **15**(15): p. 1609-18.
88. Štafa, A., et al., *In Saccharomyces cerevisiae gene targeting fidelity depends on a transformation method and proportion of the overall length of the transforming and targeted DNA.* FEMS Yeast Research, 2017. **17**(4).
89. Tsakraklides, V., et al., *Improved gene targeting through cell cycle synchronization.* PLoS one, 2015. **10**(7): p. e0133434.
90. Fairhead, C., et al., *New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker'recombination.* Yeast, 1996. **12**(14): p. 1439-1457.
91. Chung, K.-R. and M.-H. Lee, *Split-Marker-Mediated Transformation and Targeted Gene Disruption in Filamentous Fungi*, in *Genetic Transformation Systems in Fungi, Volume 2*, M.A. van den Berg and K. Maruthachalam, Editors. 2015, Springer International Publishing: Cham. p. 175-180.
92. Morschhauser, J., S. Michel, and P. Staib, *Sequential gene disruption in Candida albicans by FLP-mediated site-specific recombination.* Mol Microbiol, 1999. **32**(3): p. 547-56.
93. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
94. Cong, L., et al., *Multiplex Genome Engineering Using CRISPR/Cas Systems.* Science, 2013. **339**(6121): p. 819-823.
95. Jiang, W., et al., *Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice.* Nucleic acids research, 2013. **41**(20): p. e188-e188.
96. Friedland, A.E., et al., *Heritable genome editing in C. elegans via a CRISPR-Cas9 system.* Nat Methods, 2013. **10**(8): p. 741-3.
97. Jinek, M., et al., *RNA-programmed genome editing in human cells.* eLife, 2013. **2**: p. e00471.
98. DiCarlo, J.E., et al., *Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.* Nucleic Acids Research, 2013. **41**(7): p. 4336-4343.
99. Ozaki, A., et al., *Metabolic engineering of Schizosaccharomyces pombe via CRISPR-Cas9 genome editing for lactic acid production from glucose and cellobiose.* Metabolic Engineering Communications, 2017. **5**: p. 60-67.
100. Liu, Q., et al., *CRISPR–Cas9-mediated genomic multiloci integration in Pichia pastoris.* Microbial Cell Factories, 2019. **18**(1): p. 144.
101. Vyas, V.K., M.I. Barrasa, and G.R. Fink, *A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families.* Sci Adv, 2015. **1**(3): p. e1500248.
102. Schwartz, C., et al., *Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast Yarrowia lipolytica.* Metab Eng, 2019. **55**: p. 102-110.
103. Shi, Y., et al., *A CRISPR-Cas9 System-Mediated Genetic Disruption and Multi-fragment Assembly in Starmerella bombicola.* ACS Synth Biol, 2022. **11**(4): p. 1497-1509.
104. Strucko, T., et al., *A CRISPR/Cas9 method facilitates efficient oligo-mediated gene editing in Debaryomyces hansenii.* Synth Biol (Oxf), 2021. **6**(1): p. ysab031.

105. Weninger, A., et al., *Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris.* J Biotechnol, 2016. **235**: p. 139-49.
106. Norton, E.L., R.K. Sherwood, and R.J. Bennett, *Development of a CRISPR-Cas9 System for Efficient Genome Editing of Candida lusitaniae.* mSphere, 2017. **2**(3).
107. Liachko, I. and M.J. Dunham, *An autonomously replicating sequence for use in a wide range of budding yeasts.* FEMS Yeast Res, 2014. **14**(2): p. 364-7.
108. Sharp, P.M. and W.H. Li, *The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.* Nucleic Acids Res, 1987. **15**(3): p. 1281-95.
109. Antony, J.S., J.M. Hinz, and J.J. Wyrick, *Tips, Tricks, and Potential Pitfalls of CRISPR Genome Editing in Saccharomyces cerevisiae.* Front Bioeng Biotechnol, 2022. **10**: p. 924914.
110. Ng, H., N. Dean, and M. Lorenz, *Dramatic improvement of CRISPR/Cas9 editing in Candida albicans by increased single guide RNA expression. mSphere 2: e00385-16*. 2017.
111. Ryan, O.W., et al., *Selection of chromosomal DNA libraries using a multiplex CRISPR system.* eLife, 2014. **3**: p. e03703.
112. Horwitz, J.P., et al., *Substrates for Cytochemical Demonstration of Enzyme Activity. I. Some Substituted 3-Indolyl-β-D-glycopyranosides1a.* Journal of Medicinal Chemistry, 1964. **7**(4): p. 574-575.
113. Domenzain, I., et al., *Evaluating accessibility, usability and interoperability of genome-scale metabolic models for diverse yeasts species.* FEMS yeast research, 2021. **21**(1).
114. Bruggeman, F.J. and H.V. Westerhoff, *The nature of systems biology.* Trends Microbiol, 2007. **15**(1): p. 45-50.
115. Suthers, P.F., et al., *Genome-scale metabolic reconstruction of the non-model yeast Issatchenkia orientalis SD108 and its application to organic acids production.* Metab Eng Commun, 2020. **11**: p. e00148.
116. Reķēna, A., et al., *Genome-scale metabolic modeling reveals metabolic trade-offs associated with lipid production in Rhodotorula toruloides.* PLOS Computational Biology, 2023. **19**(4): p. e1011009.
117. Valli, M., et al., *A subcellular proteome atlas of the yeast Komagataella phaffii.* FEMS Yeast Res, 2020. **20**(1).
118. Hilliard, M., et al., *Elucidating redox balance shift in Scheffersomyces stipitis' fermentative metabolism using a modified genome-scale metabolic model.* Microbial Cell Factories, 2018. **17**(1): p. 140.
119. Heirendt, L., et al., *Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0.* Nature Protocols, 2019. **14**(3): p. 639-702.
120. Wang, H., et al., *RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor.* PLoS Comput Biol, 2018. **14**(10): p. e1006541.
121. Lu, H., et al., *Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection.* Mol Syst Biol, 2021. **17**(10): p. e10427.
122. Thiele, I. and B.O. Palsson, *A protocol for generating a high-quality genome-scale metabolic reconstruction.* Nat Protoc, 2010. **5**(1): p. 93-121.
123. Mojzita, D., et al., *l-xylo-3-Hexulose Reductase Is the Missing Link in the Oxidoreductive Pathway for d-Galactose Catabolism in Filamentous Fungi*.* Journal of Biological Chemistry, 2012. **287**(31): p. 26010-26018.
124. Harrison, M.-C., et al., *Machine learning enables identification of an alternative yeast galactose utilization pathway.* Proceedings of the National Academy of Sciences, 2024. **121**(18): p. e2315314121.
125. Chroumpi, T., et al., *Revisiting a 'simple' fungal metabolic pathway reveals redundancy, complexity and diversity.* Microb Biotechnol, 2021. **14**(6): p. 2525-2537.

126. Cuevas, D.A., et al., *From DNA to FBA: How to Build Your Own Genome-Scale Metabolic Model.* Front Microbiol, 2016. **7**: p. 907.
127. Patil, K.R. and J. Nielsen, *Uncovering transcriptional regulation of metabolism by using metabolic network topology.* Proceedings of the National Academy of Sciences, 2005. **102**(8): p. 2685-2689.
128. Orth, J.D., I. Thiele, and B.Ø. Palsson, *What is flux balance analysis?* Nature Biotechnology, 2010. **28**(3): p. 245-248.
129. Herrmann, H.A., et al., *Flux sampling is a powerful tool to study metabolism under changing environmental conditions.* npj Systems Biology and Applications, 2019. **5**(1): p. 32.
130. Fröhlich-Wyder, M.-T., E. Arias-Roth, and E. Jakob, *Cheese yeasts.* Yeast, 2019. **36**(3): p. 129-141.
131. Prado, M.R., et al., *Milk kefir: composition, microbial cultures, biological activities, and related products.* Front Microbiol, 2015. **6**: p. 1177.
132. Opulente, D.A., et al., *Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts.* Science, 2024. **384**(6694): p. eadj4503.
133. Zachariae, W. and K.D. Breunig, *Expression of the transcriptional activator LAC9 (KlGAL4) in Kluyveromyces lactis is controlled by autoregulation.* Mol Cell Biol, 1993. **13**(5): p. 3058-66.
134. Wray, L.V., Jr., et al., *Characterization of a positive regulatory gene, LAC9, that controls induction of the lactose-galactose regulon of Kluyveromyces lactis: structural and functional relationships to GAL4 of Saccharomyces cerevisiae.* Mol Cell Biol, 1987. **7**(3): p. 1111-21.
135. Varela, J.A., et al., *Origin of Lactose Fermentation in Kluyveromyces lactis by Interspecies Transfer of a Neo-functionalized Gene Cluster during Domestication.* Curr Biol, 2019. **29**(24): p. 4284-4290 e2.
136. Schaffrath, R. and K.D. Breunig, *Genetics and molecular physiology of the yeast Kluyveromyces lactis.* Fungal Genet Biol, 2000. **30**(3): p. 173-90.
137. Salmeron, J.M., Jr and S.A. Johnston, *Analysis of the Kluyveromyces lactis positive regulatory gene LAC9 reveals functional homology to, but sequence divergence from, the Saccharomyces cerevisiae GALA gene.* Nucleic Acids Research, 1986. **14**(19): p. 7767-7781.
138. Reinhardt-Tews, A., et al., *A double role of the Gal80 N terminus in activation of transcription by Gal4p.* Life Sci Alliance, 2020. **3**(12).
139. Meyer, J., A. Walker-Jonah, and C.P. Hollenberg, *Galactokinase encoded by GAL1 is a bifunctional protein required for induction of the GAL genes in Kluyveromyces lactis and is able to suppress the gal3 phenotype in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 1991. **11**(11): p. 5454-5461.
140. Halvorsen, Y.C., K. Nandabalan, and R.C. Dickson, *LAC9 DNA-binding domain coordinates two zinc atoms per monomer and contacts DNA as a dimer.* Journal of Biological Chemistry, 1990. **265**(22): p. 13283-13289.
141. Godecke, A., et al., *Coregulation of the Kluyveromyces lactis lactose permease and beta-galactosidase genes is achieved by interaction of multiple LAC9 binding sites in a 2.6 kbp divergent promoter.* Nucleic Acids Res, 1991. **19**(19): p. 5351-8.
142. Friedrich, A., et al., *Contrasting Genomic Evolution Between Domesticated and Wild Kluyveromyces lactis Yeast Populations.* Genome Biology and Evolution, 2023. **15**(2).
143. TOKUNAGA, M., et al., *Secretion of Mouse α-Amylase from Kluyveromyces lactis.* Yeast, 1997. **13**(8): p. 699-706.
144. Van Ooyen, A.J.J., et al., *Heterologous protein production in the yeast Kluyveromyces lactis.* FEMS Yeast Research, 2006. **6**(3): p. 381-392.
145. Maullu, C., et al., *High-level production of heterologous protein by engineered yeasts grown in cottage cheese whey.* Appl Environ Microbiol, 1999. **65**(6): p. 2745-7.

146. Yonten, V. and N. Aktas, *Exploring the optimum conditions for maximizing the microbial growth of Candida intermedia by response surface methodology.* Prep Biochem Biotechnol, 2014. **44**(1): p. 26-39.
147. Nützmann, H.-W., C. Scazzocchio, and A. Osbourn, *Metabolic Gene Clusters in Eukaryotes.* Annual Review of Genetics, 2018. **52**(Volume 52, 2018): p. 159-183.
148. Slot, J.C. and A. Rokas, *Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi.* Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10136-41.
149. Dalal, C.K., et al., *Transcriptional rewiring over evolutionary timescales changes quantitative and qualitative properties of gene expression.* Elife, 2016. **5**.
150. Chaffey, N., *Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. Molecular biology of the cell. 4th edn.* (0305-7364 (Print)).
151. Bhat, P.J. and T.V. Murthy, *Transcriptional control of the GAL/MEL regulon of yeast Saccharomyces cerevisiae: mechanism of galactose-mediated signal transduction.* Mol Microbiol, 2001. **40**(5): p. 1059-66.
152. Leonardo, J.M., S.M. Bhairi, and R.C. Dickson, *Identification of upstream activator sequences that regulate induction of the beta-galactosidase gene in Kluyveromyces lactis.* Mol Cell Biol, 1987. **7**(12): p. 4369-76.
153. Riley, M.I. and R.C. Dickson, *Genetic and biochemical characterization of the galactose gene cluster in Kluyveromyces lactis.* J Bacteriol, 1984. **158**(2): p. 705-12.
154. Sun, X., et al., *Recognition of galactose by a scaffold protein recruits a transcriptional activator for the GAL regulon induction in Candida albicans.* Elife, 2023. **12**.
155. Martchenko, M., et al., *Transcriptional rewiring of fungal galactose-metabolism circuitry.* Curr Biol, 2007. **17**(12): p. 1007-13.
156. Harrison, M.C., et al., *The evolution of the GALactose utilization pathway in budding yeasts.* Trends Genet, 2022. **38**(1): p. 97-106.
157. Yano, K. and T. Fukasawa, *Galactose-dependent reversible interaction of Gal3p with Gal80p in the induction pathway of Gal4p-activated genes of Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1721-6.
158. Frey, P.A., *The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose.* Faseb j, 1996. **10**(4): p. 461-70.
159. Fekete, E., et al., *Regulation of formation of the intracellular beta-galactosidase activity of Aspergillus nidulans.* Arch Microbiol, 2002. **179**(1): p. 7-14.
160. Chroumpi, T., et al., *Detailed analysis of the D-galactose catabolic pathways in Aspergillus niger reveals complexity at both metabolic and regulatory level.* Fungal Genetics and Biology, 2022. **159**: p. 103670.
161. Lu, H., et al., *Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection.* Molecular Systems Biology, 2021. **17**(10).
162. Bernstein, D.B., et al., *Addressing uncertainty in genome-scale metabolic model reconstruction and analysis.* Genome Biology, 2021. **22**(1): p. 64.
163. !!! INVALID CITATION !!! [118].
164. Fekete, E., et al., *The alternative D-galactose degrading pathway of Aspergillus nidulans proceeds via L-sorbose.* Arch Microbiol, 2004. **181**(1): p. 35-44.
165. Verhagen, K.J.A., W.M. van Gulik, and S.A. Wahl, *Dynamics in redox metabolism, from stoichiometry towards kinetics.* Current Opinion in Biotechnology, 2020. **64**: p. 116-123.
166. Mendes-Ferreira, A., et al., *Accumulation of non-superoxide anion reactive oxygen species mediates nitrogen-limited alcoholic fermentation by Saccharomyces cerevisiae.* Appl Environ Microbiol, 2010. **76**(24): p. 7918-24.
167. Herrero, E., et al., *Redox control and oxidative stress in yeast cells.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2008. **1780**(11): p. 1217-1235.

168. Barros, K.O., et al., *Oxygenation influences xylose fermentation and gene expression in the yeast genera Spathaspora and Scheffersomyces.* Biotechnology for Biofuels and Bioproducts, 2024. **17**(1): p. 20.
169. Sibirny, A.A., *Metabolic engineering of non-conventional yeasts for construction of the advanced producers of biofuels and high-value chemicals.* BBA Advances, 2023. **3**: p. 100071.
170. Buchanan, D., et al., *Recent advances in whey processing and valorisation: Technological and environmental perspectives.* International Journal of Dairy Technology, 2023. **76**(2): p. 291-312.
171. Zhai, B., et al., *Dianhydrogalactitol induces replication-dependent DNA damage in tumor cells preferentially resolved by homologous recombination.* Cell Death & Disease, 2018. **9**(10): p. 1016.
172. Zhang, G., et al., *Biocatalytic conversion of a lactose-rich dairy waste into D-tagatose, D-arabitol and galactitol using sequential whole cell and fermentation technologies.* Bioresource Technology, 2022. **358**: p. 127422.
173. Jagtap, S.S., et al., *Production of galactitol from galactose by the oleaginous yeast Rhodosporidium toruloides IFO0880.* Biotechnol Biofuels, 2019. **12**: p. 250.
174. Ortiz, A.C., et al., *D-Tagatose: A Rare Sugar with Functional Properties and Antimicrobial Potential against Oral Species.* Nutrients, 2024. **16**(12).
175. Bertelsen, H., H. Andersen, and M. Tvede, *Fermentation of D-tagatose by human intestinal bacteria and dairy lactic acid bacteria.* Microbial Ecology in Health and Disease, 2001. **13**(2): p. 87-95.
176. Jeon, E.J., et al., *Production of Tagatose by Whole-cell Bioconversion from Fructose Using Corynebacterium glutamicum.* Biotechnology and Bioprocess Engineering, 2023. **28**(3): p. 419-427.
177. Liu, J., et al., *Chassis engineering for microbial production of chemicals: from natural microbes to synthetic organisms.* Current Opinion in Biotechnology, 2020. **66**: p. 105-112.
178. Jagtap, S.S., et al., *Cloning and characterization of a galactitol 2-dehydrogenase from Rhizobium legumenosarum and its application in d-tagatose production.* Enzyme and Microbial Technology, 2014. **58-59**: p. 44-51.
179. de Jongh, W.A., et al., *The roles of galactitol, galactose-1-phosphate, and phosphoglucomutase in galactose-induced toxicity in Saccharomyces cerevisiae.* Biotechnol Bioeng, 2008. **101**(2): p. 317-26.
180. Vermeersch, L., et al., *Do microbes have a memory? History-dependent behavior in the adaptation to variable environments.* Frontiers in Microbiology, 2022. **13**.
181. Carruthers, D.N. and T.S. Lee, *Translating advances in microbial bioproduction to sustainable biotechnology.* Front Bioeng Biotechnol, 2022. **10**: p. 968437.
182. Harrison, M.-C., et al., *Machine learning illuminates how diet influences the evolution of yeast galactose metabolism.* bioRxiv, 2023: p. 2023.07.20.549758.