THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Characterisation of two novel oleaginous yeast strains for efficient and sustainable production of acetyl-CoA derived products

Friederike Mierke

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Life Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2023

Characterisation of two novel oleaginous yeast strains for efficient and sustainable production of acetyl-CoA derived products

FRIEDERIKE MIERKE

ISBN 978-91-7905-875-3

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5341
ISSN 0346-718X

Division of Food and Nutrition Science and Division of Systems and Synthetic Biology
Department of Life Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
A closer look at yeasts in nature.

Printed by Chalmers Digitaltryck
Gothenburg, Sweden 2023

“Don't you know I'm still standin' better than I ever did?
Lookin' like a true survivor, feelin' like a little kid”

- Elton John

Preface

This dissertation serves as partial fulfilment of 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 grants from the following funders: the Swedish Research Council Formas (grant number 2018-01875), and the Novo Nordisk Foundation (grant number NNF20CC0035580). The PhD studies were carried out between June 2019 and August 2023 at the Division of Food and Nutrition Science and the Division of Systems and Synthetic Biology, in close collaboration with Lund University. The PhD studies were supervised by Verena Siewers and Thomas Andlid, and co-supervised by Joakim Norbeck and Eduard Kerkhoven. The thesis was examined by Elin Esbjörner.

Friederike Mierke
August 2023

Characterisation of two novel oleaginous yeast strains for efficient and sustainable production of acetyl-CoA derived products

Friederike Mierke

Division of Food and Nutrition Science and Division of Systems and Synthetic Biology
Department of Life Sciences
Chalmers University of Technology

Abstract

Yeasts are major players in industry for production of valuable compounds. The market for yeast bioproduction is constantly growing, and sustainable bioproduction from renewable feedstocks is becoming increasingly important. Previously, *Sacccharomyces cerevisiae* has been the key player in research and industrial applications, and engineered strains have been created to have many features already present in other yeast species (production of valuable products, and utilisation of pentose sugars as carbon sources, etc.). The efficient use of pentoses is a highly investigated topic due to their abundance in renewable feedstocks such as lignocellulosic hydrolysates. Numerous yeast species exist that are naturally capable of assimilating pentose sugars with greater efficiency than engineered *Saccharomyces*. These species can also produce industrially relevant products like terpenoids, enzymes, and bioactive substances. Using the biodiversity naturally available among yeasts is an important step towards the improvement of sustainable bioproduction.

This thesis focuses on and discusses the importance of non-conventional yeasts, their isolation, characterization, and potential bioproduction of acetyl-CoA derived compounds. Two novel, non-conventional basidiomycete oleaginous yeast strains were investigated in detail: BOT-O, a *Pseudozyma hubeiensis* strain with high levels of storage lipid and biosurfactant production, and BOT-A2, a *Rhodotorula toruloides* strain, which produces of high-value carotenoids and storage lipids. Both were isolated from plant material and characterised regarding physiology, genomics and transcriptomics. The strains were *de novo* sequenced, their genomes assembled and annotated. RNA sequencing was conducted to investigate the differences in transcriptional levels between growth on xylose and glucose, or during nitrogen starvation, which is an important condition for high lipid production. Both BOT-O and BOT-A2 produced high amounts of lipids on glucose or xylose. BOT-O showed identical growth rates on both sugars. Both were also somewhat tolerant towards lignocellulosic inhibitors. Interestingly, BOT-O was more tolerant towards furanic aldehydes, while BOT-A2 was more tolerant towards weak acids. Six new promoters that are regulated during nitrogen starvation were identified for BOT-A2. They are a valuable addition to the *R. toruloides* genetic engineering toolbox. Overall, these two new isolates have highly interesting characteristics and the potential to become hosts for bioproduction of native and non-native acetyl-CoA derived compounds.

Keywords: *Pseudozyma hubeiensis*, *Rhodotorula toruloides*, genome sequencing, RNA sequencing, promoters, xylose utilisation, oleaginous yeasts, non-conventional yeasts, lignocellulosic hydrolysate, inhibitor tolerance

List of Publications

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Qvirist L, **Mierke F**, Vazquez Juarez R, and Andlid T (2022) Screening of xylose utilizing and high lipid producing yeast strains as a potential candidate for industrial application. *BMC Microbiology*, 22(1):173. DOI: https://doi.org/ 10.1186/s12866-022-02586-y

II **Mierke F***, Brink DP*, Norbeck J, Siewers V, and Andlid T (2023). Functional genome annotation and transcriptome analysis of *Pseudozyma hubeiensis* BOT-O, an oleaginous yeast that utilizes glucose and xylose at equal rates. *Fungal Genetics and Biology*, 166:103783. DOI: 103783. 10.1016/j.fgb.2023.103783

III Brink DP*, **Mierke F***, Norbeck J, Siewers V, and Andlid T (2023). Expanding the genetic toolbox of *Rhodotorula toruloides* by identification and validation of six novel promoters induced or repressed under nitrogen starvation. *Submitted.*

IV **Mierke F**, Stucchi D, Norbeck J, Andlid T, and Siewers V (2023). Tolerance assessment of two non-conventional oleaginous yeast isolates towards lignocellulosic inhibitors. *Manuscript*.

*These authors contributed equally.
Reprints were made with permission from the respective publishers

Contribution Report

I Performed part of the wet lab experiments (shake flask cultivations, lipid extraction, HPLC) and data analysis. Drafted part of the manuscript and edited it.

II Performed all the wet lab experiments (cultivations, lipid extraction and analysis, HPLC, RNA extraction, DNA extraction and MinION genome sequencing). Contributed to the design and performed part of the bioinformatics workflow and to genome assembly, differential gene expression, gene set analysis and data analysis. Shared in writing and editing the manuscript.

III Performed part of the wet lab experiments (cultivations, lipid extraction and analysis, HPLC, RNA extraction, DNA extraction and MinION genome sequencing, promoter strain screening and part of the flow cytometry experiments). Contributed to the design and performed part of the bioinformatics workflow. Shared in performing of genome assembly, differential gene expression and data analysis. Shared in writing and editing of the manuscript.

IV Shared in study conception. Initiated and designed all experimental work, performed some of the experiments and drafted the manuscript. Shared in editing of the manuscript.

Table of Contents

Abbreviations

ANI – Average nucleotide identity
CAZymes – Carbohydrate-active enzymes
cDNA – Complementary DNA
FAS – Fatty acid synthase
gDNA – Genomic DNA
GFP – Green fluorescent protein
GRAS – Generally Regarded As Safe
HMF – Hydroxymethylfurfural
HPLC – High performance liquid chromatography
LiAc – Lithium acetate
MEL – Mannosylerythritol lipid
NGS – Next-Generation Sequencing
OD_{600nm} – Optical density at 600 nm
PCR – Polymerase chain reaction
PHA – Polyhydroxyalkanoate
PHB – Poly(3-hydroxybutyrate)
RIN – RNA integrity number
PPP – Pentose phosphate pathway
ROS – Reactive oxygen species
qPCR – Quantitative polymerase chain reaction
TAG – Triacylglyceride
TCA – tricarboxylic acid
TGS – Third-Generation Sequencing
YPD – Yeast extract Peptone Dextrose

Chapter 1: Introduction

Biotechnological applications have been used for millennia, although unknowingly for the most part, for example in wine and beer production, and are important tools in the fight against climate change and in efforts towards a more sustainable future for our planet. Despite decades of knowledge of the existence of climate change, its causes and available tools, not enough actions have been taken to fulfil the suggested climate goals. However, immediate and efficient actions are more and more necessary, and science aims to solve some of the most pressing issues. For example, the replacement of fossil resources with natural and sustainable ones such as bioethanol and biodiesel instead of conventional petrol, or recyclable and biodegradable plastics instead of fossil plastics. Microorganisms play a vital role in the advancement of sustainable production.

1.1 Microorganisms and Bioproducts

Microorganisms are omnipresent in our world and have a large impact on human civilisation. The earliest biotechnological products included wine, beer, bread, cheese, and other fermented foods. As centuries have progressed, knowledge and awareness about production mechanisms and microorganisms have increased substantially. Fermented alcoholic beverages, first produced in neolithic times, were created without understanding of the processes involved [1-3]. The 17th century saw the invention of the microscope, and thus the first visual observations of microorganisms by Robert Hooke and Antonie van Leeuwenhoek [4]. It wasn't until the mid-19th century that the presence of microorganisms was finally shown to be responsible for fermentations [5, 6], and it took another century for James Watson, Francis Crick, and Rosalind Franklin to make their breakthrough revelation of the double-helix structure of DNA and usher in the age of genetics [7]. The invention of the polymerase chain reaction (PCR), a method used to amplify small amounts of DNA, just a few decades ago by Kary Mullis is an indispensable tool that revolutionised genetic engineering [8]. All of these discoveries are important milestones which have led to today's targeted use, engineering and manipulation of microorganisms to produce complex biotechnological products such as industrial enzymes, pharmaceutical biomolecules, and bioplastics. These modern production processes utilize a variety of microorganisms, among which, yeasts are some of the most important and frequently used.

The most commonly used yeast is *Saccharomyces cerevisiae*, the so-called baker's yeast. Archaeological evidence of this species' existence and usage in production of fermented foods and beverages reaches back thousands of years in all areas of the world [1-3, 9]. It is also the primary model organism in yeast biotechnology. Consequently, a wide variety of

ways to engineer it for improved production of native and non-native products have already been successfully developed [10]. However, *S. cerevisiae* does have some drawbacks, especially with regards to the efficient utilisation of plant waste materials as feedstocks: it is not naturally capable of assimilating pentose sugars such as xylose, and it also is not very tolerant towards inhibitors found in these feedstocks, especially when engineered for xylose assimilation [11-13]. These so-called lignocellulosic hydrolysates are widely investigated as a renewable feedstock for sustainable bioproduction because they contain high amounts of sugars like glucose and xylose. In addition to *S. cerevisiae*, there are a wide variety of other yeast species that have been isolated from nature for biotechnological purposes. Some of these are natural xylose utilisers such as *Scheffersomyces stipitis* [14] or *S. shehatae* [15]. In addition, other yeasts can be natural producers of a wide range of high-value bioproducts such as carotenoids [16], lipids [17], and biosurfactants [18]. Thus, the discovery and investigation of new and non-conventional yeast species and strains is important and can add value for both yeast research and the biotechnology industry.

A variety of bioproducts share one common precursor – acetyl-CoA. This is a metabolic intermediate and serves as a pivotal point for the biological production of lipids, carotenoids and biosurfactants, which will also be shown in this thesis. Yeasts that naturally accumulate a high amount of lipids during stress conditions such as nitrogen limitation [19], so-called oleaginous yeasts, typically have a high flux towards cytosolic acetyl-CoA, and are thus interesting candidate cell factories for the production of acetyl-CoA derived products.

1.2 Renewable Raw Materials for Bioproduction

For sustainability aspects, the feedstock used for the production of high-value bioproducts is important to consider. Lignocellulosic hydrolysates have gained importance over the past few decades as a sustainable bioproduction feedstock [20]. So-called first-generation lignocellulosic hydrolysates, made from primary materials such as corn, wheat and sugarcane, have recently been primarily used for the production of compounds such as bioethanol [21, 22]. However, due to the competition of these raw materials with the growing food demand, those first-generation materials are not considered sustainable, and are also more costly than alternative lignocellulosic materials derived from plant waste, such as corn stover, wheat straw or bagasse. The origin of the raw material generally depends on its abundance in specific parts of the world. In the Nordics, for example, materials such as spruce wood and other softwoods play a major role for plant waste valorisation as these are abundant in this climate [23]. Bagasse, a waste product of sugarcane plantations, is however highly abundant in South America, whereas corn stover is abundant in the U.S. [24]. Chemically, lignocellulose from plant materials is composed of three major components: cellulose, hemicellulose and lignin, with slight differences in ratio depending on the raw material [25]. Very few microorganisms can directly utilise plant material, and thus the three components are often broken down by chemical pretreatment and hydrolysis of the biomass, releasing sugars such as glucose and xylose, but also potential growth inhibitors like weak acids and furanic aldehydes [26, 27]. In order

to efficiently use lignocellulosic materials, microorganisms would have to be able to somewhat withstand the growth inhibitors present, as well as being capable of utilising most of the available sugars, namely glucose but also xylose, which are generally the most abundant carbon sources in this type of material.

1.3 Isolation and Characterisation of New Yeast Strains for Bioproduction

Although a variety of yeasts, dominated by *S. cerevisiae*, are already successfully being used in industry and have been well-established as production hosts, the search for new species and strains is important. It has been shown that different strains of the same species can differ substantially depending on their isolation environment [28]. These differences, which are likely due to the adaptation of the specific strain to its ecological niche, can affect a variety of phenotypic responses, and might reveal more desirable traits in new isolates compared to established strains. Thus, increasing knowledge on yeast biodiversity is beneficial for both science and industry.

New yeast strains can be isolated from basically anywhere in nature [29-31]. The presence of certain species can, among other things, be attributed to the specific sample medium, e.g. soil, or temperature [32]. After successful isolation, phenotypic and molecular characterisation of the isolates can be performed to taxonomically identify the isolated species. The ever-growing availability and decreasing costs of genome sequencing aid substantially in the correct classification of new isolates and give much better and faster insights into the biology of the strain.

1.4 Aims and Structure of this Thesis

The overall aim of this thesis was the identification and characterisation of novel, non-conventional oleaginous yeast strains and the assessment of their potential as production hosts for acetyl-CoA derived products from renewable feedstocks.

First, new yeast isolates were identified and preliminarily characterised in **Paper I**. The aims for the isolation and screening in **Paper I** were to find new yeast strains that are capable of xylose utilisation and of producing high amounts of lipids intracellularly. Xylose utilisation is an important characteristic for efficient utilisation of renewable feedstocks such as lignocellulosic hydrolysates. High lipid producing yeasts generally have a high flux towards cytosolic acetyl-CoA, which is an important precursor for a variety of high-value bioproducts.

To further analyse and assess the potential of the new yeast isolates as production hosts, the two most promising strains in terms of high lipid production and xylose utilisation were selected for further characterisation using shake flask cultivations, genomics and RNA sequencing. Findings for BOT-O, a non-conventional oleaginous yeast of the *Pseudozyma*

hubeiensis species, are presented in **Paper II** and for BOT-A2, a non-conventional oleaginous yeast of the *Rhodotorula toruloides* species, are presented in **Paper III**.

Functional genetic engineering strategies are important assets for production host candidates. The possibility of genetically engineering BOT-O and BOT-A2 were therefore investigated. While it has not yet been possible to engineer BOT-O, despite attempts to do so, engineering strategies were successful for BOT-A2. In **Paper III**, RNA and genome sequencing results were used to identify potential promoter candidates for differential regulation during nitrogen starvation for BOT-A2.

Xylose utilisation, which was ensured through the selection during isolation of the strains in **Paper I**, is an important factor for efficient use of plant waste materials in the form of lignocellulosic hydrolysates. While these renewable feedstocks provide usable carbon sources, they also contain inhibitors that affect microbial growth and production yields. Thus, the two yeast isolates needed to be tested for their inhibitor tolerance. This was investigated on spruce wood hydrolysate for BOT-O and BOT-A2, as well as synthetic hydrolysate and single lignocellulosic inhibitors in **Paper IV**.

Of course, this thesis does not just discuss all the papers, but also aims to give a more detailed overview and discuss the results of this work in a different light. Please view this thesis as sort of a yeast enthusiast's guide to the discovery and characterisation of novel yeast production strains, as it will guide through all the important aspects of finding new production strains: from the importance and diversity of yeast, and the isolation and screening of novel strains from nature in **Chapter 2**, via the characterisation of new isolates and available methods in **Chapter 3**, to a deep-dive into bioproducts, their precursors and potential ways of flux redirection in **Chapter 4**, and finally, in the name of sustainability, the usage of renewable feedstocks with all its advantages and drawbacks in **Chapter 5**.

Results from additional experiments will also be discussed throughout the chapters, along with results from the papers. Discussions from all paper**s** will be expanded on, and comparisons between the two investigated oleaginous yeast species will be made, with the aim to reveal interesting new information and insights.

Chapter 2: Let's Talk About Yeasts

The word "yeast" is derived from old English and Germanic words meaning "foam" and "to ferment", referencing fermentation processes for food and beverage production that some, but not all, yeasts are part of [33]. In fact, there are more than 1,500 known yeast species, and yet only a very small fraction of them, even today, is being used for research, food and pharmaceutical production [34]. Yeasts have played an important role in the advancement in society for thousands of years, with or without the knowledge of humankind. Nowadays, we understand a lot more of what yeasts are, what they do and how we can utilise them.

Yeasts have existed for hundreds of millions of years [35]. They are eukaryotes, and part of the fungi kingdom. Yeasts, in contrast to other fungi, can reproduce asexually through budding and fission. While some of them are human pathogens, with infections being caused mainly by *Candida* and *Cryptococcus* species, there is a smorgasbord of beneficial applications in fundamental research and biotechnology for these organisms (Figure 2.1) [36].

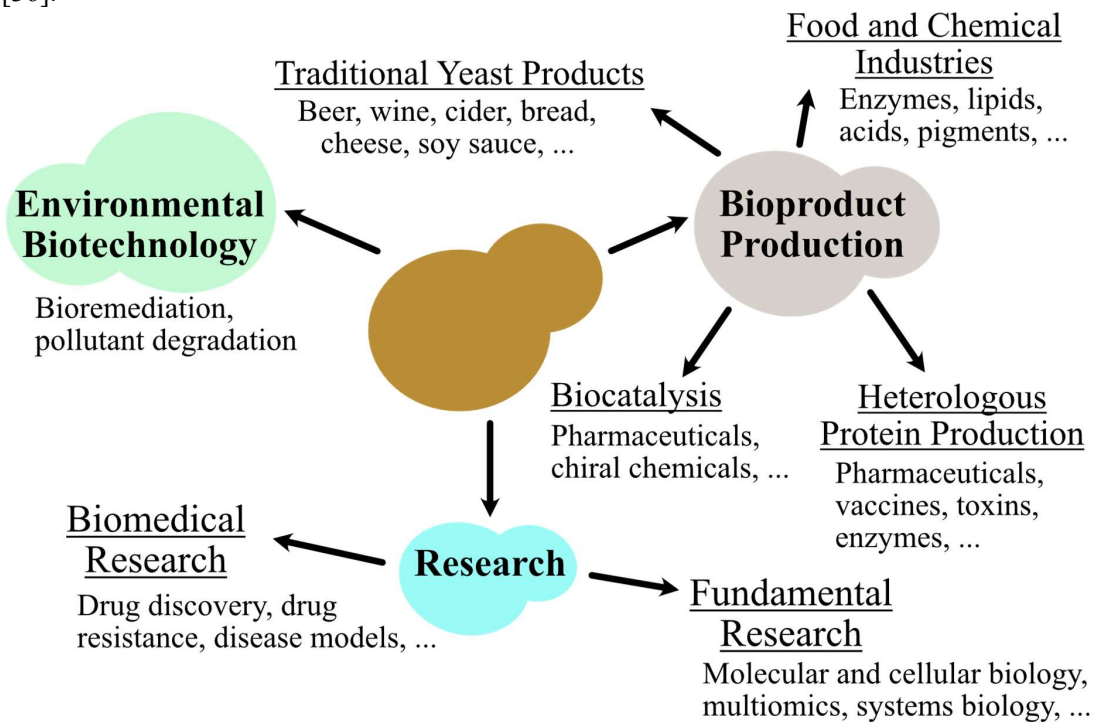

Figure 2.1: Overview of research and biotechnological applications of yeasts (adapted from [36]).

Yeasts are versatile, robust and efficient microbes, which adds to their importance for research and industrial applications. In general, yeast applications can be divided into three major groups: research, bioproduct production (i.e. industrial applications) and

environmental biotechnology (Figure 2.1). The oldest applications, such as wine and bread production, are considered to be traditional yeast applications. Biomedical research is necessary for studying of diseases and drug discovery. Fundamental research into understanding the basic working mechanisms of yeast cells began a few centuries ago and is still ongoing today. Industry-wise, yeast biotechnology contributes to a trillion-dollar market across different industries [37-40]. The biggest yeast utilisation area in industry is bioethanol production for fuel [41]. Additionally, pharmaceuticals and bio-based chemicals make up a big part of the market share [42]. It is an industry with high growth historically, and according to market research, the yeast and yeast extract market is projected to have a compound annual growth rate of around 10% between 2023 and 2028 [43] .

In general, yeasts can be found in two phyla: Ascomycota, which *S. cerevisiae* belongs to, and Basidiomycota, which the two strains further discussed and characterised in this work belong to. In order to understand the diversity present in yeast and the importance and advantages of certain species, it is important to discuss the different phyla and yeast species. The two phyla are distinguished by their way of sexual reproduction. Ascomycetes form specialised sacs, so-called "asci", for holding the sexual spores [44], whereas basidiomycetes form basidia, which are specific cellular structures for haploid, sexual spore production [45]. In research, there is a clear bias towards Ascomycota, due to the preference for *S. cerevisiae*, with over 11,000 genomes available on NCBI, compared to only around 2,000 for Basidiomycota in June 2023 [46]. Genome sequence availability plays a major role for the advancement of research into a certain species. The most commonly used yeasts in the biotechnology industry are generally also the best researched ones with most information and tools available. Some examples of common industrially used species and their applications are listed in Table 2.1. Ascomycota yeasts are certainly more widely used in industry and research.

Table 2.1. Examples of important industrial yeasts and some of their applications.

Species	Phylum	Industrial application
Saccharomyces cerevisiae	Ascomycota	Recombinant proteins [47], lactic acid [48], succinic acid [47], bioethanol [49], food fermentations [50]
Yarrowia lipolytica	Ascomycota	Recombinant proteins [47], single-cell oil [51, 52]
Kluyveromyces marxianus	Ascomycota	Aromatic compounds and enzymes [53, 54], food fermentations [55]
Komagataella pastoris **(f.k.a. *Pichia pastoris*)**	Ascomycota	Recombinant proteins [47], vaccines and other pharmaceutical biomolecules [56]
Ogataea polymorpha **(f.k.a. *Hansenula polymorpha*)**	Ascomycota	Recombinant proteins [47], vaccines and biopharmaceuticals [57]
Scheffersomyces stipitis **(f.k.a. *Pichia stipitis*)**	Ascomycota	Lactic acid [58], bioethanol [59]
Phaffia rhodozyma **(f.k.a. *Xanthophyllomyces dendrorhous*)**	Basidiomycota	Carotenoids [16, 60]
Meyerozyma guilliermondii **(f.k.a. *Pichia guilliermondii*)**	Ascomycota	Enzyme production, metabolite synthesis, biocontrol [61]

2.1 Yeasts from the Ascomycota Phylum

The Ascomycota phylum is the most species rich phylum in the Fungi kingdom, and also the most diverse [62]. It can be divided further into three subphyla: Taphrinomycotina, Saccharomycotina and Pezizomycotina. As mentioned above, despite the large number of species, there is one major species that dominates research and industry: *S. cerevisiae*, baker's yeast. The same way yeast research is skewed towards Ascomycota with over five times more genomes available on NCBI as for Basidiomycota as mentioned above, research is also clearly skewed towards *S. cerevisiae*: More than 10% of the available genomes for Ascomycota on NCBI are made up of genomes for this particular species [46].

2.1.1 *Saccharomyces cerevisiae – The* Yeast?

S. cerevisiae is the primary model for biotechnological applications, due to intensive research and advances in terms of sequencing, and genetic engineering tools for this species. *S. cerevisiae* is probably most well-known for its ethanol production capabilities. It is a good natural producer of ethanol, and is also the most ethanol-tolerant yeast to date [63], with ethanol-tolerance of up to 12% reported for some strains [64]. Additionally, it is remarkably tolerant towards very high sugar concentrations, and a known producer of aromatic, volatile compounds. It is classified as a GRAS organism (Generally Regarded As Safe), which is an important factor for an industrial strain. It is, however, not naturally

capable of utilising pentose sugars, which are abundant in renewable plant materials that can be used as sustainable feedstocks.

Evidence for the presence of *S. cerevisiae* in fermented alcoholic beverages goes back to Neolithic times in various places around the world [1-3]. Thus, the connection between humankind and *S. cerevisiae* is deeply rooted. Yet, ancient history is not the only reason for the establishment of *S. cerevisiae* as the biotechnological model organism. Foremost, the first successful complete genome sequence of this species in the 1990s, which was also the first eukaryotic genome sequence to be presented, affected the importance of *S. cerevisiae* in research and industry in the following decades even more than before. The genome sequence was a combined effort of hundreds of different researchers from various universities and research institutes [65]. As sequencing techniques and methods progressed, the sequences have been updated and expanded. Together with a variety of available sequences, RNA sequences and proteomics data, a good basis for intensive research was laid. For this reason, *S. cerevisiae* will likely continue to be the most popular species for research and industry for the foreseeable future.

Not much is known about the origins and the actual natural environment of domesticated *S. cerevisiae* strains, although recent research suggests Far East Asia as a place of origin [66, 67]. Interestingly, while wild *S. cerevisiae* strains can be found in a variety of environments, grapes are surprisingly not considered a native environment [68], although it is a typical wine yeast. Investigation into the microbiome of grapes has shown that other microorganisms are much more commonly found on grapes than *S. cerevisiae* [68]. The abundance of this yeast species in natural settings appear reasonably high; Wang and colleagues collected more than 2,000 samples from natural habitats and found evidence of *S. cerevisiae* in around 11% of the samples [29]. Genetic diversity between different yeast isolates was much higher than in domesticated strains, and diversity in domesticated strains in China has been found to be much higher than on any other continent [66], supporting the theory of this area being a potential origin of the *S. cerevisiae* species.

The *Saccharomycetaceae* family itself has been present on our planet for over 200 million years [35]. *S. cerevisiae* diverged from other species in its family some time after that. Then, about 100 million years ago, *S. cerevisiae* and some other yeast species in that family are thought to have undergone a whole-genome duplication event [69]. It was also around that time that the so-called Crabtree effect most likely emerged [35]. The Crabtree effect describes the repression of respiration by the fermentation pathway during the presence of high concentrations of a substrate such as glucose [70]. This is also called overflow metabolism, and is characterised by high growth rates and high glycolytic fluxes, as well as an increase of carbon channelled towards by-products of fermentation [71], such as ethanol.

Interestingly, there are significant differences between industrial, laboratory and wild strains isolated from nature. One big notable difference is ploidy. While laboratory strains are generally haploid or diploid, industrial or wild strains are typically diploid or aneuploid,

and specifically the industrial strains sporulate poorly [36, 72]. This leads to difficulties for classic mating procedures and genetic engineering strategies. Furthermore, differences in phenotypic responses between laboratory strains and industrial strains can be observed [28, 36, 73-75]. These differences are obvious when it comes to environmental changes, nutrient limitations, desiccation and cold stresses. Comparisons between wild strains isolated from the environment and lab and industrial strains also reveal the presence of additional survival strategies in the natural strains [50].

2.1.2 *Kluyveromyces marxianus*, *Yarrowia lipolytica* and Co.

As mentioned above, there is a variety of biotechnologically important Ascomycota species other than *S. cerevisiae*, each with their own advantages and characteristics. Traditionally, in applied research, these yeasts, along with the Basidiomycota yeasts, have been referred to as "non-conventional" yeasts. While this may still be an appropriate phrase for some species, a few yeast species have been increasingly investigated and utilised, with increasing numbers of available genome sequences, annotations, and tools for genetic engineering, turning them into more established species.

Some species that have gained significant scientific interest include *Yarrowia lipolytica*, *Kluyveromyces marxianus*, and *Scheffersomyces stipitis* [47]. *Y. lipolytica* is known for its high lipid production (for more details, see Chapter 2.3.1), which has been heavily researched in recent years [17, 76], but also for the production of lipolytic enzymes and lactones, and as a host for single-cell protein production. It can utilise polyalcohols, organic acids and long-chain hydrocarbons as feedstocks [36, 77]. This yeast is, however, not naturally capable of utilising pentose sugars like xylose as a sole carbon source, though metabolic engineering has been used to introduce this trait [78]. *K. marxianus*, along with its relative *K. lactis*, has gained interest due to its ability to utilise lactose as a carbon source [79]. *K. lactis* was moreover the second yeast species to be generally regarded as safe (GRAS) [80]. *K. marxianus* has recently piqued interest as a result of its desirable traits for biotechnological application. Among other traits, it can assimilate lactose, inulin and xylose, shows great secretory capacities and a high thermotolerance, as well as a rapid growth rate even in comparison to other *Kluyveromyces* species [79, 81]. Methods for metabolic engineering of *K. marxianus* exist, and has been applied to construct production hosts for various compounds including carotenoids, polyketides or short-chain fatty acids [82-84]. *S. stipitis*, also known as *Pichia stipitis*, has been extensively studied for its xylose metabolism and is a source for xylose metabolism genes that are used to engineer other yeasts such as *S. cerevisiae* [85]. It is a known lactic acid and ethanol producer and can produce high amounts of these compounds even from xylose under well-controlled oxygen conditions, and with low xylitol formation [58, 85, 86] , which is beneficial for efficient use of xylose-rich feedstocks such as lignocellulosic hydrolysates. Formation of high amounts of xylitol would indicate inefficient xylose utilisation. In the oxidoreductase pathway, which can be found for example in *P. stipitis*, xylose is first reduced to xylitol by xylose reductase which is NAD(P)H-dependent, and then oxidised to xylulose by the NAD^+-dependent xylitol dehydrogenase. While the conversion of xylose to xylulose is redox-neutral, different redox factors are used, which can cause imbalance, especially in

non-native xylose utilising strains [87]. As *P. stipitis* only produces low amounts of xylitol, it is considered to be an efficient xylose utiliser. In fact, the first xylose-utilising *S. cerevisiae* strains were generated by expression of the *XYL1* and *XYL2* genes from the oxidoreductive pathway for xylose metabolism from this yeast [14, 85, 88, 89]. *S. stipitis*, despite all its interesting characteristics, is however not a popular ethanol production strain due to low ethanol tolerance [90].

Of course, there are a large number of other important Ascomycota yeast species with useful characteristics. *Debaromyces hansenii*, for example, is very osmotolerant and able to grow in media with much higher sodium chloride contents than *S. cerevisiae*. Osmotic tolerance can be a desirable trade in industry as high salt media can supress the growth of other microorganisms and thus reduce risks of contamination and process costs [91]. *D. hansenii* has also shown tolerance towards chlorine dioxide, which functions as a disinfectant [92], and it has also been shown to produce toxins active against other yeast species [93]. Furthermore, this yeast produces high amounts of lipids and can also naturally utilise xylose [91]. *Ogataea polymorpha*, also known as *Hansenula polymorpha*, is another natural xylose utiliser, one of the most thermotolerant yeasts [94], and can produce ethanol. *Schizosaccharomyces pombe* is a widely used model organism for the investigation of molecular and cellular processes [95]. Important *Candida* strains for industrial production in general are *C. maltosa* and *C. tropicalis*, which are known for their alkane and phenol utilisation [36]. The *Candida* genus in itself is interesting as it is very heterogenous. A variety of characteristics can be observed in *Candida* yeasts, in part due to the fact that classification of a species as belonging to the *Candida* genus is largely based on having similar morphology to previously-known *Candida* yeasts, along with a lack of a defined teleomorph [96, 97], making this genus somewhat of a “melting pot” [36]. The adaptation of molecular studies for utilisation in taxonomic classification led to changes in nomenclature. Genome sequence analyses are more reliable for classification than phenotypic analyses. The nomenclature changes due to this development happen throughout the entire Fungi kingdom, but the *Candida* genus is arguably the most affected one [98].

2.2 Yeasts from the Basidiomycota Phylum

There are three subphyla in the Basidiomycota phylum: Puccinimycotina, Agaricomycotina, and Ustilaginomycotina. It is the second most species-rich phylum in the Fungi kingdom after the Ascomycota phylum but has been largely under-researched with exception of a few species despite the many interesting native products synthesised by yeasts in this phylum. Not much is known about basidiomycete yeast ecology in general, and species have been shown to occur in a plethora of environments ranging from marine and freshwater habitats to soil and plant surfaces [99]. Species in the *Trichosporon* genus are a great example of this phenomenon, as they have been isolated from a wide variety of different environments [36]. Thus, it is no wonder that they can also use a wide variety of substrates, like many other Basidiomycota yeasts.

2.2.1 "Red Yeasts" or Carotenoid-Producing Yeasts

Carotenoids are some of the most common pigments found in nature [100]. They function as antioxidants acting against reactive oxygen species (ROS) [101, 102] and provide photoprotection [103, 104]. They can be produced by chemical synthesis, but in the light of sustainable production, interest has grown in carotenoids as bioproducts. The carotenoid market is rather large, as this metabolite is used in the food and feed industry, as well as pharmaceutical, chemical and cosmetic industries [105]. While yeast species from both phyla are capable of carotenoid production [106], it is mainly some basidiomycete yeasts that are known to produce and accumulate carotenoids. Those yeasts often occur on plant material surfaces [107]. The most widely known basidiomycete carotenoid producer is *Phaffia rhodozyma* (also known as *Xanthophyllomyces dendrorhous*). This yeast produces astaxanthin [16, 60], which is an oxygenated carotenoid and only appears to be synthesised, among yeasts, by this species [108]. However, yeasts of the *Rhodotorula* and *Sporobolomyces* genera are also known carotenoid producers, and have been shown to produce β-carotene, γ-carotene, torulene and torularhodin [108]. One of the two new yeast isolates discussed in more detail in this thesis, a strain of the *Rhodotorula toruloides* species, is also classified as a "red yeast".

2.2.2 *Ustilago maydis* and its Relatives

Ustilago maydis is a dimorphic basidiomycete fungus and a known plant pathogen. Depending on its life cycle and on environmental conditions, this organism undergoes a transition from a yeast form to a filamentous form [109-111]. In its haploid yeast form, a large cell body and asexual reproduction through budding can be observed [112]. The filamentous state has two different forms, the infectious one that results from fusion of basidiospores, and the parasitic proliferative filament stage [113]. *U. maydis* is the cause of corn smut disease in *Zea mays*, and has long been used and established as a model system in plant pathology [114-116]. It first grows along the epidermis of the host plant, and then forms a structure which is specialised to infect the host. The so-called appressorium then penetrates the epidermal cells, using cell wall degrading enzymes [117]. Despite being a plant pathogen, *U. maydis* is also an organism used in traditional food production in Latin America [118]. Production of killer toxins beneficial for biocontrol in wine production has also been observed [119], as well as production of organic acids, glycogen, lipids, and biosurfactants [120-122]. Furthermore, the *U. maydis* genome revealed the presence of interesting enzymes such as xylanases [123].

Efforts are being made to establish *U. maydis* also as a model organism for higher eukaryotes for the investigation of more basic cellular biological process such as mitosis [116]. Genome-wide analysis suggests that *U. maydis* is actually more closely related to humans than the ascomycete yeasts *S. cerevisiae* and *S. pombe* that are widely used as models. In fact, a variety of proteins is only shared between humans and *U. maydis* [116] as compared to *S. cerevisiae*. In the 1960s, *U. maydis* was used to investigate the mechanisms behind homologous recombination [124] before *S. cerevisiae* took off as the prime model organism.

There are additional interesting species to be found in the family of *Ustilaginaceae*, also with some cases of nomenclature changes. The two yeasts *Moeszyomces antarcticus* (formerly known as *Pseudozyma antarctica*) and *Moeszyomyces aphidis* (formerly known as *Pseudozyma aphidis*) are known for their excellent glycolipid production [125]. Specifically, similarly, to the *Pseudozyma hubeiensis* species, to which one of the two yeast isolates discussed in this thesis belongs to, they produce mannosylerythritol lipids (MELs), which can be used as biosurfactants in various industries [125]. Both *M. antarcticus* and *M. aphidis* have even been shown to produce these lipids from waste stream materials [126]. *M. antarcticus* also possesses a highly interesting enzyme for bioplastic degradation [127]. They are species of high interest for the production of biodiesel, heterologous proteins and lipase enzymes [36, 128].

2.3 Oleaginous Yeasts

Oleaginous yeasts are defined by the naturally high production of intracellular storage lipids. A microorganism is considered oleaginous if it accumulates more lipids than 20% of its cell dry weight [129, 130]. Only about 70 oleaginous yeast species, out of the over 1,500 existing yeasts, are known [131]. Generally, high storage lipid production, for example in the form of triacyl glycerides (TAGs), requires a stress factor such as nutrient limitations, and an excess of carbon [129, 132]. Commonly used nutrient limitations include nitrogen, sulphur or phosphorous starvation. Especially nitrogen starvation has been a major focus in lipid accumulation condition optimisation, and high carbon to nitrogen ratios, also called C/N ratios, have been shown to be beneficial for high lipid accumulation [19]. Oleaginous yeasts are gaining more and more interest, especially in the food and feed industry, as yeast storage lipids could be used to substitute plant oils such as palm oil or cocoa butter [17].

Oleaginous yeasts can be found in both discussed yeast phyla. A few oleaginous yeasts have already been touched upon in the previous subchapters. Some well-known ascomycete oleaginous yeasts include *D. hansenii*, *Y. lipolytica*, and some *Candida*, *Lipomyces*, *Cryptococcus* and *Pichia* species [36]. Among basidiomycete yeasts, a lot of "red yeasts" have been described as oleaginous, for example *P. rhodozyma*, *Rhodotorula toruloides*, *Rhodotorula diabovata* (also known as *Rhodosporidium diabovatum*), *Sporidiobolus salmonicolor*, *Cystofilobasidium infirmominiatum*, and *Sporobolomyces carnicolor* [131]. During isolation of yeasts in **Paper I** a high incidence of pigment producing yeasts among the high storage lipid producers was observed [133]. Furthermore, more oleaginous yeasts were isolated from warmer environments than from colder ones in **Paper I**.

In **Paper I**, it was also discovered that the strain BOT-O, belonging to the *Pseudozyma hubeiensis* species, produced high amounts of TAGs on both glucose and xylose, but had not often been described as oleaginous previously, potentially due to the focus of earlier research on its biosurfactant production capabilities.

2.3.1 *Yarrowia lipolytica* – The Best Characterised Oleaginous Yeast

Yarrowia lipolytica is possibly the best characterised oleaginous yeast used in research, and thus somewhat of an oleaginous yeast model organism. It has been well described, and genome sequences, as well as genetic engineering strategies exist for this species. Genome sizes are around 20 Mb, which is much larger than the genome of *S. cerevisiae* at 12 Mb [134]. Several different strategies have been employed to improve lipid production in this yeast [135], resulting in an increase from the lipid accumulation from up to 30 to 50% of the cell dry weight in some natural isolate strains to 90% of the cell dry weight in so-called "obese" strains, with the help of genetic engineering and process optimisation [136]. To induce high lipid accumulation, nitrogen starvation is the most commonly used growth condition [137], and glucose and other sugars are used as carbon sources. Lipid production has been shown to be significantly lower on substrates such as crude glycerol, with cells focusing on metabolising other products [138]. *Y. lipolytica* has been a candidate for food oil production, and genetic engineering has started to aim not only at increasing the TAG formation, but also change the fatty acid profile to match commonly used plant oils [17, 139].

However, this yeast also does have some other interesting properties: Strains of the *Y. lipolytica* species are widely distributed in nature and can be isolated from various different habitats such as food products like cheese and meat, salterns and marine algae [140-142], including salty environments and foods. Therefore, tolerance towards higher salt concentrations can be observed, with most strains growing well in NaCl concentrations of up to 7.5%. Some strains have even been shown to tolerate up to 15% NaCl [143]. This yeast can also grow in a wide pH range and is able to adsorb metallic atoms such as Fe, Cu, and Ni, which is a useful characteristic for bioremediation of heavy metal containing waste [143]. Natural isolates are, in contrast to *S. cerevisiae*, usually haploid [144, 145], and colony morphology can be highly diverse. *Y. lipolytica* can grow, depending on the growth conditions, as a round budding yeast, forming pseudohyphae, or as mycelium with septate hyphae [145-147]. It is also a popular host for heterologous production due to its efficient secretion pathway, and has been used for production of more than 130 different proteins from over 80 different species of organisms [148].

This oleaginous yeast can grow naturally on sugars such as glucose, fructose and mannose as a sole carbon source, but natural xylose utilisation is very strain dependent [136, 149]. However, omics studies have proven that *Y. lipolytica* possesses a somewhat dormant pentose pathway [150]. Xylose can be consumed in mixed sugar cultivations after glucose depletion, and enzymes in the functional xylose pathway are simply poorly expressed, with xylitol dehydrogenase and xylose kinase presenting limiting steps [150, 151]. Variability is high in xylose consumption after glucose depletion among different strains [152]. Generally, xylose utilisation capabilities are improved using genetic engineering and overexpression of the native xylose genes [136, 151], or genes from well-known xylose utilisers such as *S. stipitis* [78, 151], which has already been discussed in the case of *S. cerevisiae*. Xylose utilisation can also be condition-dependent, with adaptation or starvation periods increasing xylose uptake [149]. Nevertheless, successful engineering

endeavours have overcome this issue, and *Y. lipolytica* strains have been developed that can produce high amounts of lipids on pure xylose.

2.3.2 *Pseudozyma hubeiensis* – A Non-Conventional, Underexplored Yeast with Great Potential

Another yeast capable of high storage lipid production on pure xylose, albeit naturally and without the need for genetic engineering, is *Pseudozyma hubeiensis*. In **Paper I** it was shown that *P. hubeiensis* BOT-O can accumulate up to 46% lipids of its cell dry weight in nitrogen-limited medium and 50 g/L of glucose or xylose [133], without any process optimisation and only minimal differences in yield compared between the two different carbon sources. This lipid content is on par with that found in this species in general, and also with the average lipid content reported for *Y. lipolytica* [153]. *P. hubeiensis* is a basidiomycete yeast that is more closely related to *Ustilago maydis* than to other, more commonly used yeasts such as *Rhodotorula* spp. or *Moeszyomyces antarcticus* (Figure 1, **Paper II**) [154]. The species was first isolated from plant leaves by Wang and colleagues [155]. In fact, many new *Pseudozyma* strains have been isolated from plant parts [133, 156, 157], although some new strains were also isolated from saltwater or human blood [158, 159]. Another very interesting feature of this species was the finding that they have equal growth on glucose and xylose in **Paper II** (Figure 2 and Figure 8, **Paper II**). Growth rates and sugar accumulation rates were similar for growth on either glucose or xylose, which is a highly desirable trait, especially as it was combined with no xylitol formation during growth on xylose. When grown on a combination of glucose and xylose, the growth rate did not change, however glucose was still taken up faster than xylose, so a preference for glucose somewhat exists. Tanimura and colleagues were able to show a similar effect in a mix of sugars [160] including not only glucose and xylose, but also arabinose. Additionally, I have been able to show that *P. hubeiensis* BOT-O can grow on a selection of xylans (Supplemental Figure S8 in Supplemental File S1 of **Paper II**).

P. hubeiensis is not just an excellent xylose utilising and lipid producing yeast but is also a producer of MELs and cellobiose lipids, which are biosurfactants important for various industries. In fact, before publication of **Paper I** and **Paper II**, *P. hubeiensis* had mainly been researched with a focus on its biosurfactant production [125, 161, 162]. MELs are produced in a similar fashion as in the already discussed model organism *U. maydis*, and I was able to reconstruct the entire MEL pathway from the genome sequence, based on *U. maydis* genes, in **Paper II** (Figure 3, **Paper II**).

Despite all its interesting features, there are still some downsides to using *P. hubeiensis* as a production hosts. Foremost, genetic engineering have not been established for this species, and limited attempts have been published; only one electroporation protocol exists for the *P. hubeiensis* strain SY62 [163]. In this thesis, attempts to genetically engineer BOT-O using a similar transformation method as for *M. antarcticus* (which was formerly classified as belonging to the *Pseudozyma* genus) [164, 165] were made but have so far not been fruitful. However, after phylogenetic analysis presented in **Paper II**, we theorise that

genetic engineering strategies could be adapted from *U. maydis* instead. Different approaches, including a CRISPR-Cas9 system, exist already for *U. maydis* [166].

2.3.3 *Rhodotorula toruloides* – Not So Novel But Quite Underexplored

Rhodotorula toruloides is a carotenoid producing oleaginous yeast with a salmon pink colour (Figure 2.2) that has been shown to be able to accumulate high amounts of storage lipids during growth on different carbon sources, including pentose sugars like xylose. In **Paper I**, a lipid content of over 40% of cell dry weight on glucose or xylose in a medium with a high C/N ratio was achieved, albeit with a higher lipid content in the glucose medium. Growth and sugar assimilation rates differ between growth on glucose and growth on xylose, as demonstrated for *R. toruloides* strain BOT-A2 in **Paper III**.

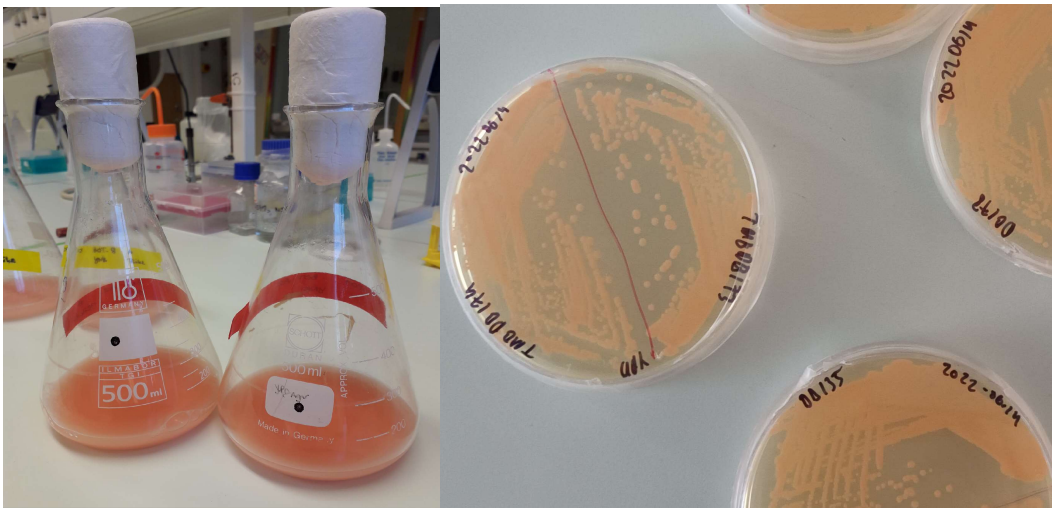

Figure 2.2. *R. toruloides* BOT-A2 and derivative strains in liquid culture (a) and on YPD plates (b).

Rhodotorula toruloides is also known as *Rhodosporidium toruloides* [167], and both species names are still quite frequently used in scientific literature. Yeasts can have two different states, a teleomorph or sexual state, and an anamorph or asexual state. The genus name *Rhodotorula* has historically been used for the anamorphic state, whereas the name *Rhodosporidium* has been used for the respective teleomorph [36]. However, under the Amsterdam declaration on fungal nomenclature which was adapted in 2013 [168], having multiple valid species names caused by distinction of teleomorph and anamorph was abandoned, and, overall, the anamorph name was adapted. Therefore, *Rhodotorula toruloides* is the currently correct name for the *R. toruloides* species.

R. toruloides can be isolated from a variety of backgrounds, but has mainly been isolated from soil or plant parts. This includes the strain BOT-A2 which was isolated from plant material as described in **Paper I** and further investigated in **Paper III** [133, 169-171]. One *R. toruloides* strain has recently been found in flooding water of an old uranium mine [172], suggesting the presence of bioremediation qualities. Analysis of all available *R. toruloides* genome sequences in **Paper III** revealed large differences in the average nucleotide identity (ANI) (ANI analysis: Figure 1 in **Paper II)**, and a formation of two distinct

clusters. Coincidentally, the cluster distinction corresponds to differences in mating types. *R. toruloides* has a bipolar mating type system, dependent on only one locus consisting of multiple genes, hence only two different mating types exist: MAT-A1 and MAT-A2 [173]. Most of the *R. toruloides* strains have genetic information for just one of the mating types. Accordingly, mating type switching, as is presumed to happen in *S. cerevisiae,* does not occur in *R. toruloides* [174]. However, one strain seemed to contain loci for both mating types, suggesting a diploid strain: the *R. toruloides* strain CCT 0783, which still clustered with the other MAT-A1 strains in the ANI analysis. Different mating types exist to decrease occurrences of self-mating and promote outbreeding [175]. Investigation into mating types is considered interesting as mating type sequences are thought to have evolved at a higher speed than other genes that encode for more conserved metabolic functions [176, 177]. However, many species have not yet been characterised with respect to mating types and their loci. The mating type system of *S. cerevisiae* has, of course, been researched in detail, and some basidiomycete species have also been investigated more closely in terms of their mating types. In general, two mating type systems can be found in the Basidiomycota: the bipolar mating system that can be found in *R. toruloides*, and the tetrapolar mating system found in *U. maydis.* The latter is also suspected to be the system found in *P. hubeiensis*. The tetrapolar mating system, where the mating type is determined by two loci, is even more complex [175]. For *U. maydis*, mating type loci are designated a and b, and while there are only two alleles for the a locus, there are up to 25 for the b locus [178-180]. The clustering that was discovered in the ANI analysis is not just due to the mating types however, as the mating type ORFs do not make up a large enough percentage of the genome sequence.

2.4 Isolation of New Yeast Species and Strains

2.4.1 Yeast Isolation Strategies, and The Importance of Yeast Ecology and Evolution

Yeasts are widely distributed and can be found anywhere in the world. As described above, they are a diverse group with different characteristics dependent on their natural habitats. They can be found in obligate relationships with other organisms such as beetles, and differ in their means of propagation [181]. Hence, their habitats range from the upper levels of the atmosphere to the very deepest parts of the ocean [181]. Isolation of new species and strains can be done in a variety of ways and is also highly dependent on the research question. Some researchers devote their entire careers to the discovery of new yeast species, while other researchers may simply just be looking for new strains of a known species with more beneficial characteristics for their goals. Either way, the knowledge of the biodiversity will be expanded upon. This expansion is important for research and industry, as there are still struggles with the few species that are commonly used: A large effort goes into engineering them to utilise certain feedstocks or produce certain compounds efficiently; using yeasts that can already naturally for example utilise xylose or produce carotenoids could prove to be more efficient.

Generally, there are two different approaches to the isolation of new yeast species and strains from the environment: (i) an exploratory approach towards isolation, to broaden the understanding of yeast ecology, or (ii) a goal-oriented approach to isolate strains and species with specific properties. When there is a specific goal, this goal shapes the questions of where to sample for yeasts, and how to select for the desired yeasts. In **Paper I**, the goal was to isolate xylose utilising and high storage lipid producing yeasts. A variety of oleaginous and xylose utilising yeast strains had previously been isolated from plant material [182-188], thus plants were also chosen for sampling in this work. However, if one were to look for very salt tolerant species, for example, sampling in salt rich environments like sea water or salt lakes could prove more beneficial [141]. Yeasts that might be beneficial for bioremediation have been isolated from unusual places such as car washes [189]. If this is due to niche adaptation, as is a suggested underlying reason for yeast diversity, is hard to tell. Researchers have long argued about how and why yeasts, along with other microorganisms, are so widespread in nature. Baas-Becking suggested in 1934 that "everything is everywhere – the environment selects", which is also called the ubiquity theory [190]. Essentially, it suggests that geography has no impact on the distribution of microorganisms. This theory has been heavily debated by yeast researchers in the field of ecology [181, 191-193], and the debate persists even today with more easily available sequencing methods. Of course, arguments are not made easier by the fact that biases are introduced into research: Domesticated yeasts behave very differently when it comes to sporulation for example [194], evolution experiments are generally carried out in a human-made environment which has been found to promote faster adaptation [195], and during the strain isolation process a bias is introduced as isolation usually is performed with a goal or an assumption in mind [196]. For example, with the goal of isolating xylose utilising yeasts from plant material in mind, a xylose-based medium would be used. Therefore, non-xylose utilisers would not be able to grow, which does not mean they did not grow on the plant material in general. Or, when *S. cerevisiae* strains are to be isolated from nature, the isolation medium might be supplemented with ethanol as this yeast is known to be quite tolerant towards this compound and capable of utilising it. However, broad variations regarding those capabilities exist within the *S. cerevisiae* species, thus a bias towards more tolerant strains would be introduced, potentially obscuring the range of *S. cerevisiae* strains that we are aware of [196].

Nevertheless, researchers are eager to understand yeast evolution and uncover its mechanisms as this will add valuable insights into the species. The newly widespread availability of genomic information for a plethora of species has certainly been of help and has revealed ecological and evolutionary aspects that would have been undetectable in a laboratory setting using cultivation-based methods [197]. However, descriptions of natural habitats and their history, as well as potential interactions between yeasts and their environments remain to be collected and investigated [193, 197]. To further complicate matters, evolutionary mechanics might also be species dependent. *S. cerevisiae,* for example, is suggested to be somewhat niche-less and very broadly distributed [196], *Y. lipolytica* has been suggested to thrive in a wide environment range instead of increasing

their competitiveness in just a few substrates [135]. For basidiomycete yeasts, not much is known about their evolution in general [193].

2.4.2 Yeast Isolation Methods

Isolation of yeast is perhaps the easiest and least time-consuming task in the workflow of discovering and characterising a new strain. After initial sampling, with potential goals in mind as discussed above, samples need to be carefully treated, ideally with sterile equipment, and packed into sterile sample containers to avoid contamination from other sources as much as possible. Samples should also be used as fast as possible to avoid changes in the microbiome of the sample that could negatively affect the present yeast population. Of course, not only yeasts are present in the sample, and thus the first medium used should at least contain one antibiotic to suppress bacterial growth. In **Paper I**, chloramphenicol was used. If too much mould is present in the samples, a simple sediment method can be used to separate the yeast from the mould [133]. Strategies of microorganism transfer from the sample onto agar plates depends on the form of the sample. In **Paper I,** plant leaves and twigs were either directly placed on the agar surface, or repeatedly brought in contact with the agar to allow for transfer. Solid samples could, however, also be suspended in small amounts of sterile water, which is then plated. Liquid samples can be directly plated onto the agar, and different dilutions could be used [141, 189]. Plates can then be incubated at different temperatures; for example, if thermotolerant species are to be isolated, a higher temperature could lead to the desired goal. After growth, different colonies can be picked and re-streaked on agar plates to achieve pure colonies. Then, characterisation of the new isolates follows.

Chapter 2: Summary

- Yeast biotechnology is a multi-billion-dollar market. Yeasts can naturally produce a wide variety of products and can also be used as non-native production hosts.
- While *S. cerevisiae* is the most commonly studied and industrially relevant yeast species, it has some drawbacks and investigation of other yeasts can lead to improved cell factories.
- Both the Ascomycota and the Basidiomycota phyla offer interesting yeast species, but research on Basidiomycota is still lagging behind.
- Oleaginous yeasts are yeasts that produce high amounts of storage lipids intracellularly in the form triacylglycerides and can be found in both phyla.
- Isolation of new species is an important task to broaden the knowledge of the biodiversity and find novel species and strains with excellent industrial potential.
- The two novel isolates further discussed in this thesis, *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2, are oleaginous basidiomycete yeasts, that produce other valuable products in addition to high amounts of storage lipids.

Chapter 3: Characterisation of New Yeast Isolates – Gaining In-Depth Knowledge

After the isolation of a new strain the work has only just begun. In order to learn what species has been isolated, sequencing of just a few key genes is generally sufficient. However, thanks to recent technological advances, even whole-genome sequencing is now easily available and at a lower price point. Achieving high quality *de novo* sequences requires reliable methods, which are now also made available through advances in the field. Furthermore, in order to determine the specific characteristics of the isolated strain, extensive and in-depth investigation into physiology and omics data analysis will be necessary.

Historically, physiology and morphology were the only ways to characterise and classify new yeast isolates. Macroscopic and microscopic descriptions for morphological characterisation, as well as physiological characterisation can be employed [198]. Such phenotypic description has, however, led to a variety of identifications that have been questioned and subsequently changed due to the use of molecular, higher precision methods. This has caused big waves of reclassifications of yeast species [98]. Nonetheless, "fundamental" characterisations of physiology and morphology are still important. While genomes from different strains belonging to the same species can be nearly identical, the strains can still have broad phenotypic spectra, with differences in factors such as carbon source assimilation, tolerance towards certain inhibitors, or production capacities.

3.1 Morphology

Microscopic and macroscopic morphological descriptions were historically the first characteristics noted about new yeast isolates [198]. These descriptions can be obtained with the help of some agar plates and a microscope. When considering the cell morphology of yeast, one often thinks of the typical *S. cerevisiae* shape: relatively large (around 3.5-11 µm) ellipsoidal cells, that can sometimes be seen budding [199]. However, cell appearance, shape and size changes throughout the cell cycle or due to changes in environmental conditions [200]. *S. cerevisiae* cell size changes when yeasts are budding, and can even form pseudohyphae, which is defined as filamentous growth in the form of chains of elongated cells [201, 202]. Sporulating cells can undergo endobudding [201]. Single cells of the *R. toruloides* yeast strain have been observed to have a similar size as *S. cerevisiae*, and a subglobose to ovoid shape [36], which is in agreement with what was observed for BOT-A2 (Figure 3.1a). Budding cells can also be observed during growth as indicated by the blue arrow in Figure 3.1a. Diploid strains of the *R. toruloides* species have been reported to have larger cell sizes [203], ranging in the upper cell size range observed for *S. cerevisiae*, which allows for more space for lipid bodies. Lipid bodies were observed at a 1000x magnification without any lipid staining in BOT-A2 (indicated by a red arrow in

Figure 3.1a), a strain that is currently presumed to be haploid. The less explored *P. hubeiensis* BOT-O, however, differs drastically in cell size and shape (Figure 3.1b). Cells of BOT-O display a much less yeast-like morphology with a smaller cell size, which is described to be about 1.5-3 x 5.2-10 µm for the *P. hubeiensis* species in general, and a more cylindrical cell shape [155]. They, however, still do form lipid bodies during nitrogen starvation that are observable at a 1000x magnification without the need for staining (indicated by a red arrow in Figure 3.1b). Budding and pseudohyphae formation can also be observed [155].

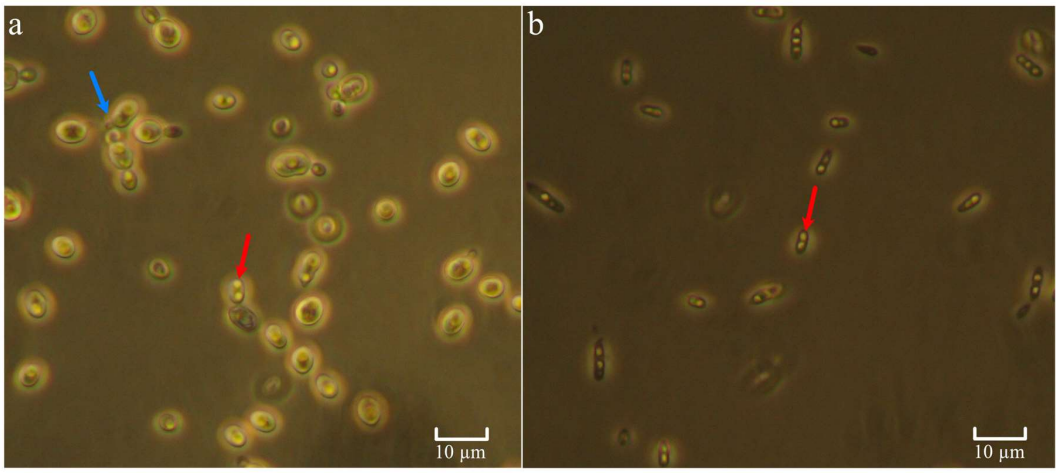

Figure 3.1. *R. toruloides* BOT-A2 (a) and *P. hubeiensis* BOT-O (b), both at the same 1000x magnification after a 96 h cultivation in high carbon, low nitrogen medium with ammonium sulphate as the nitrogen source.

Macroscopically, *R. toruloides* colonies have been described as butyrous, shiny and smooth, with a round shape and an entire edge or margin. *R. toruloides* BOT-A2 colonies are pink to orange, although the colour changes over time and starts off more cream-coloured, similar to those of other *Rhodotorula* strains, with round and smooth colonies (Figure 3.2a). For *P. hubeiensis*, colonies have been described as whitish to cream in colour, butyrous, dull, and smooth or somewhat wrinkled on yeast malt agar. The margin was entire or partially eroded. For BOT-O on Yeast extract Peptone Dextrose (YPD) agar, the colour was indeed whitish to cream, with a wrinkled surface, but not smooth (Figure 3.2b).

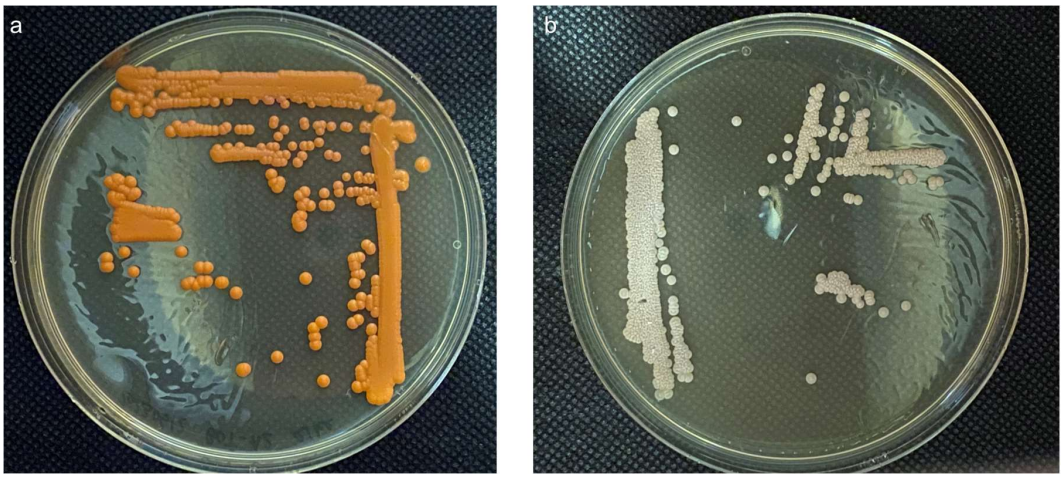

Figure 3.2. *R. toruloides* BOT-A2 (a) and *P. hubeiensis* BOT-O (b) on YPD plates after 48 h at 30°C.

On different xylans, colonies of BOT-O seem to be white around the edges, but more brownish towards the middle, with wrinkled surface and an irregular shape (Figure 3.3). These plates, however, were incubated for a longer period of time (about 5-7 days), which may have contributed to the change in macroscopic morphology.

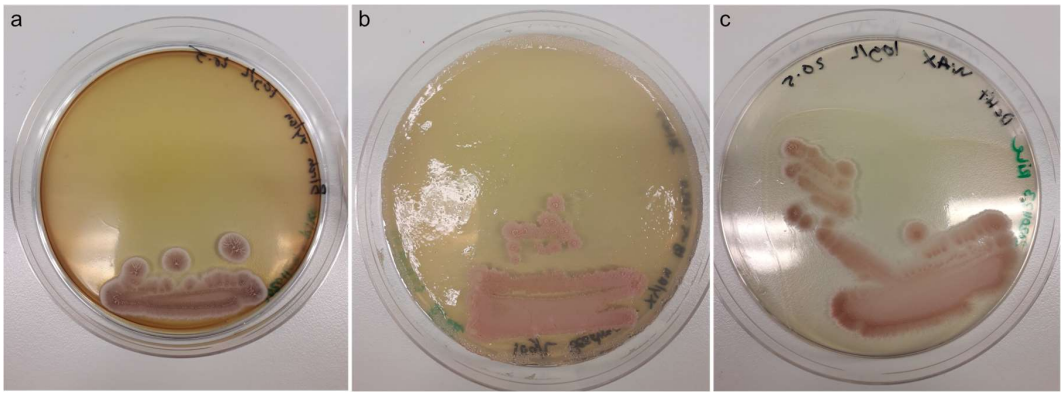

Figure 3.3. *P. hubeiensis* BOT-O on YP plates with three different xylans as the sole carbon source: birchwood xylan (a), beechwood xylan (b), and wheat arabino xylan (c).

3.2 Physiology

In addition to morphological parameters, new isolates are further investigated using growth tests, which can also narrow down the potential species a new isolate could belong to. A wide variety of carbon and nitrogen source assimilation tests and fermentation tests can be performed [36], and databases are available to identify the microbe in question depending on their metabolic activity [36]. In clinical settings, where the type of yeast causing an infection needs to be determined for proper treatment, faster and easy-to-use test kits are preferred, since the aforementioned methods are somewhat time-consuming [204]. However, those test kits can also be more specific towards certain species because there are only a few species that commonly cause infections in humans. Regardless, identification is nowadays mainly based on the sequencing of a few key genes, without the need for whole-genome sequencing. Satisfactory identification of new isolates based on

partial sequencing of large ribosomal DNA subunits and PCR, as presented in **Paper I**, can be achieved [205, 206]. However, carbon and nitrogen source assimilation tests, as well as fermentation tests are still important as there can be subtle differences even between strains of the same species. Even within the *S. cerevisiae* species there is a wide range of fermentation performances observable, depending on the origin of the strains [28].

Both strains used in this work, *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2, were initially tested in **Paper I** for growth on, and lipid production from, different carbon sources. They were found to utilise hexose and pentose sugars, as well as glycerol for example. Carbon source assimilation was measured using high performance liquid chromatography (HPLC), a very versatile method able to detect a variety of different carbon sources in the same sample. While growth on glucose or xylose in shake flasks did not lead to a significant difference in terms of growth rates, sugar and ammonium utilisation, or lipid production for *P. hubeiensis* BOT-O (Figure 3.4; Figure 2 in **Paper II**), a significant difference could be detected for *R. toruloides* BOT-A2. Growth rates on xylose were about half to a third of those on glucose (Figure 3.4), and ammonia and sugar were depleted at a slower rate (**Paper I** and **Paper III**). Lipid production was less affected by the difference in sugars (Figure 3.4; Figure 2 in **Paper III**). During fermenter cultivations on pure xylose in **Paper I**, BOT-A2 growth was better in terms of final OD_{600nm} and growth rate than in shake flask experiments on pure xylose, suggesting that xylose utilisation can be further improved with optimisation of cultivation parameters. Growth on hydrolysates was also investigated, in **Paper I** and in **Paper IV**, however in more detail in the latter. *R. toruloides* BOT-A2 was found to be more tolerant towards growth on spruce wood hydrolysate compared to *P. hubeiensis* BOT-O, which was used in **Paper IV**. Growth on solid medium was poor for *P. hubeiensis* BOT-O in the work done in **Paper I** but was later on found to be sufficient at least on YPD and YP medium with different xylans as the sole carbon source. Another potentially interesting physiology aspects that was however not investigated for either is growth temperature. Both yeasts, *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2, were grown at 30°C, a growth temperature optimum was however not determined.

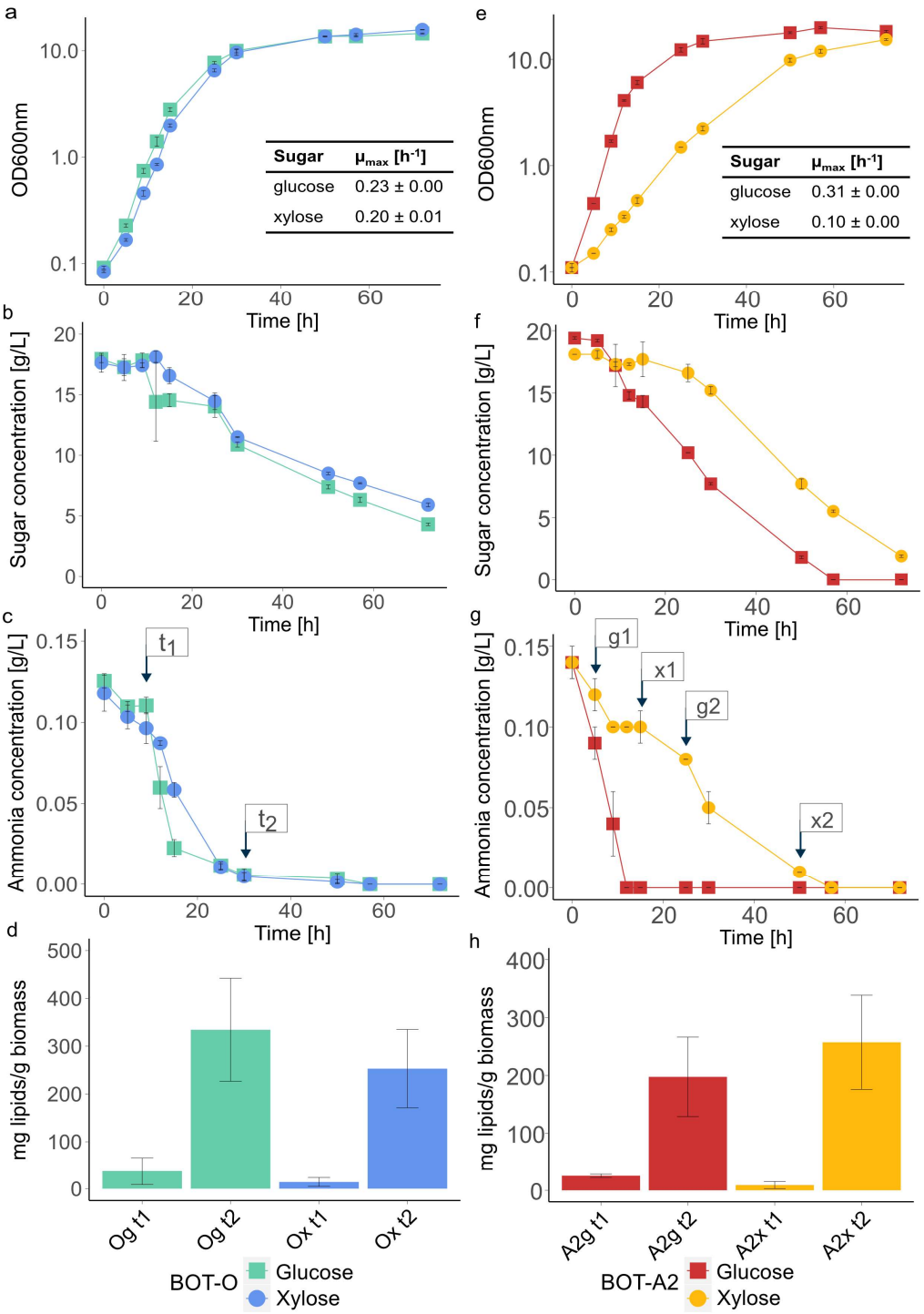

Figure 3.4. Shake flask cultivations of *P. hubeiensis* BOT-O (a-d) and *R. toruloides* BOT-A2 (e-h) on glucose and on xylose for investigation of growth parameters and sampling for RNA sequencing. a and e: Growth measured as OD600nm. b and f: Sugar concentration for glucose and xylose. c and g: Ammonia concentration, measured using an enzymatic kit. d and h: Lipid concentration at the two RNA sequencing time sampling points for growth on both glucose (g) and xylose (x). Adapted from **Paper II** and **Paper III**.

In addition to different carbon sources, two different nitrogen sources were compared, as nitrogen is an important nutrient. The type of nitrogen source used has been shown to impact cellular behaviour: lipid accumulation in oleaginous yeasts can be affected [207], as well as product formation in *S. cerevisiae* [208]. Both ammonium sulphate and urea were tested. Ammonium sulphate is perhaps the most commonly used nitrogen source in research settings and was thus an obvious choice. It is produced industrially by treatment of ammonia with sulphuric acid. This process does require a large amount of energy and produces greenhouse gas emissions [209]. Urea however can be produced from municipal waste, and is generally more affordable [210, 211]. It is thus perhaps a more industrially relevant nitrogen source. Furthermore, urea possesses buffering capabilities, while buffer addition is generally necessary when using ammonium sulphate, adding another cost factor. Different pH and pH optima were not tested for these yeasts.

The two nitrogen sources were tested for both yeasts and together with either glucose or xylose as carbon source (at a concentration of 20 g/L). Ammonium sulphate was added at 0.47 g/L, approximately a tenth of a typical ammonium sulphate concentration used in microbiological media, to achieve a C/N ratio of 80 (g/g) on glucose, while urea was added at a concentration of 0.21 g/L, again approximately a tenth of what is typically added, to achieve a high C/N ratio of 80. This ratio has been optimised during my work and was found to promote storage lipid formation while providing time for exponential growth for both BOT-O and BOT-A2. An enzymatic kit was used to measure nitrogen content in the form of urea and ammonia (Urea/Ammonia Assay Kit, Megazyme, Brae, Ireland).

There was no effect of nitrogen source on the growth curves of *R. toruloides* BOT-A2 (Figure 3.5a and 3.5b) but higher OD_{600nm}s and higher growth rates were consistently observed on glucose compared to xylose. Urea depletion was slightly slower than the depletion of ammonium sulphate, however without a significant difference between the two sugars. Depletion of the nitrogen source was, furthermore, slower on xylose than on glucose. This is related to the significantly slower growth on xylose. For *P. hubeiensis* BOT-O (Figure 3.5c and 3.5d), growth on urea was slightly slower than growth on ammonium sulphate suggesting a dependence on nitrogen source. The OD_{600nm}s were lower on urea than on ammonium sulphate on both glucose and xylose. Urea was never fully depleted in this experiment for BOT-O, whereas ammonium sulphate depletion occurred after 30 h. Due to the faster depletion of the nitrogen and the higher OD_{600nm}s reached for growth on ammonium sulphate, it was decided to proceed with ammonium sulphate as nitrogen source in this project. However, since the difference was small, urea utilisation could still be considered and optimised for future applications, providing more cost-efficient cultivations that may be necessary for industrial applications. It has, however, been reported that cellobiose lipid accumulation in *Ustilago maydis*, a relative of *P. hubeiensis*, is significantly affected by the source of nitrogen [162], resulting in differences in cellobiose lipid composition. Since cellobiose lipids are suggested to be derived from the same precursors, an effect on TAG formation could possibly be deducted. It would therefore be important to explore how urea affects lipid accumulation in *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2 in such context.

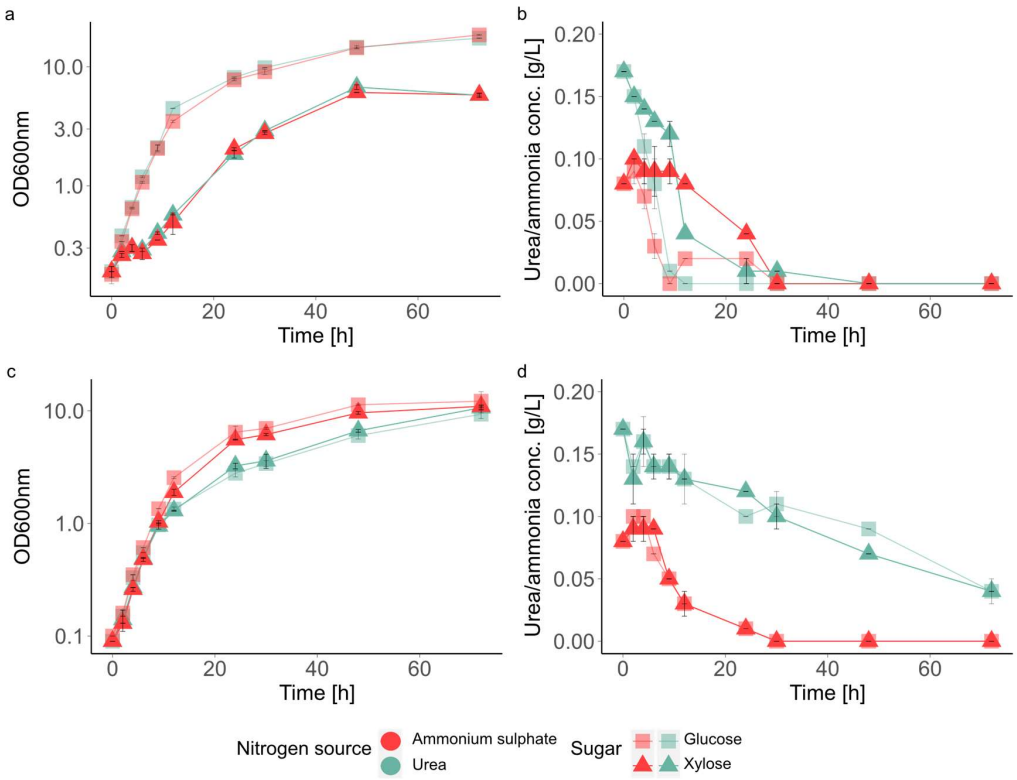

Figure 3.5. Shake flask cultivations of *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2 for the investigation of the effect of different nitrogen sources on growth and nitrogen assimilation. Growth was measured in the form of OD_{600nm} for BOT-A2 (a) and BOT-O (c), and nitrogen assimilation was measured for both urea and ammonium sulphate using an enzymatic kit for samples for BOT-A2 (b) and BOT-O (d). Cultivations were performed in duplicates on either glucose or xylose in 500-mL shake flasks with a working volume of 100 mL in a medium with yeast nitrogen base without ammonium sulphate and without amino acids. For cultivations with ammonium sulphate, 1 M potassium buffer was added.

While little is known about preferred nitrogen sources in *P. hubeiensis*, it has been suggested that ammonium sulphate is the primary source of nitrogen for *R. toruloides* [212]. This is not in agreement with my results as presented in Figure 3.5 and in the text above. Urea is a "non-primary" nitrogen source. This necessitates additional steps in the organism's nitrogen metabolism, which can explain a lesser efficiency. Zhu et al. detected an upregulation of genes associated with both the central nitrogen metabolism and the urate/urea utilisation pathway, during nitrogen starvation after ammonium sulphate depletion for the *R. toruloides* strain NP11 [212]. Those genes, including the NADP(+)-dependent glutamate dehydrogenase *GDH3*, the NAD(+)-dependent glutamate dehydrogenase *GDH2*, the glutamine synthesis *GLN1*, and the NAD(+)-dependent glutamate synthase *GLT1*, as well as the urea amidolyase *DUR1,2*, were not found to be upregulated during nitrogen starvation for BOT-A2 in **Paper III**, suggesting minimal differences between transcriptomic expression of these genes during nitrogen starvation compared to exponential growth. Upregulation of the ammonium transmembrane transporter *MEP2* in the BOT-A2 strain was also not significant on glucose during nitrogen

starvation (**Paper III**). However, upregulation was significant on xylose during nitrogen starvation with a $\log_2$ fold change of 5.86, indicating a potential effect of the choice of carbon source on ammonium transport.

3.3 Omics – High-Throughput Molecular Methods

Genomics, transcriptomics, and proteomics are gaining more and more importance in research due to easier and more affordable access to sequencing facilities, development of easy-to-use sample preparation methods, as well as sequencing or mass spectrometry methods, and availability of extraction kits, and more user-friendly and readily available bioinformatics tools. Together with metabolomics, they can give an overview over many of the ongoing mechanisms in the cell. Genomics and transcriptomics have come a long way since the first methods for DNA sequencing were developed in the 1970s [213, 214]. These first-generation sequencing approaches have since been improved, and new technologies, referred to as Next-Generation Sequencing, and Third-Generation Sequencing have been developed, along with new bioinformatics approaches to tackle the increase in data, as well as the change in type of data and their quality. Specifically, widely available genome sequencing allows for easy species identification on a genotypic level, while offering definite answers about the classification. This in turn also leads to more accurate classifications, potentially leading to less reassignments and minimising confusion within the yeast community for the time being.

Whole-genome sequencing can also deliver data for more reliable calculation of genetic distances and construction of phylogenetic trees, as well as help in correlating phenotype and genotype [215]. This is the case also for non-conventional yeast strains, which in turn can make these yeasts more interesting for research and industry and offers more options in the future: Instead of using the few current platform host strains and trying to continuously improve them and engineer them to produce non-native products, native producers of valuable compounds can be used for efficient bioproduction [47]. Certainly, physiological characterisation is of great importance in order to determine the interesting phenotypes of new strains. However, having an available genome sequence is generally a prerequisite for genetic engineering projects, which could be used to improve the production capacities of native production hosts. While genome sequences for other *R. toruloides* and *P. hubeiensis* strains already exist, *de novo* sequencing of new strains can add valuable information and insight into biodiversity of these species, while also setting the base for potential genetic engineering endeavours of these specific strains. In **Paper II** and **Paper III**, a combination of long- and short-read sequencing was used to obtain high quality genome sequences for BOT-O and BOT-A2.

3.3.1 Genome Sequencing and Annotation

First-generation sequencing methods like the Sanger method based on utilising DNA polymerase [213, 214] and the chemical sequencing method developed by Maxam and Gilbert [216], which were a breakthrough at the time of their invention in the 1970s, are quite low throughput in terms of samples and cost-intensive methods. Even improved approaches adjacent to these first-generation sequencing methods, like shotgun DNA

sequencing, do not add significant improvements [217, 218]. After a 13-year collaborative project to accurately and completely sequence the human genome, which started in the 1990s [219, 220], efforts were made to improve sequencing approaches to make sequencing faster, more high-throughput and more economically efficient. This led to the development of Next-Generation Sequencing (NGS), which started to be commercially introduced in 2005 [221].

3.3.1.1 Next-Generation Sequencing – Short-Read Sequencing

The first widely used and accepted NGS method introduced a highly parallel sequencing system, which increased the throughput by 100-fold compared to the previously commonly used Sanger method [221]. This led to a significant decrease in sequencing costs and running times. This pyrosequencing approach was then adapted by Roche, which provided sequencing with their 454 Life Science Sequencer. While there were also other companies with NGS approaches, the only other serious contender was the Illumina Genome Analyzer Platform, which was introduced by Illumina in 2007 and is still widely used today. In 2012, it took over the market with a high throughput of 20 Gb per run compared to only 400 Mb per run on the 454 Sequencer [222], which ultimately led to the discontinuation of the Roche 454 Life Science Sequencer in 2016. Now, over 90% of sequencing data worldwide is generated using the Illumina sequencing platform [223], as it is useful for a wide variety of projects, including whole-genome sequencing but also sequencing of complementary DNA (cDNA) from RNA for transcriptomic analysis.

After extraction of the DNA, or RNA for RNA sequencing, a sequencing library is prepared which ensures that the extracts are compatible with the sequencing platform. RNA has to be transformed into cDNA. The purified DNA (or cDNA) is then fragmented into smaller pieces, and adapters are added to the end of the fragments. The adapters are specialised to allow binding to the flow cells of the sequencing instrument, as well as to provide binding sites for primers [223, 224]. Barcoding is also possible by addition of a sample-specific label, i.e. a barcode, to the adapters. The library, which consists of the fragmented reads with adapters, is loaded onto a flow cell, and then placed into a sequencer. The prepared library then binds to the flow cell (Figure 3.6), and subsequently to the primers, where each fragment gets amplified through bridge amplification via PCR. Bridge amplification is repeated for multiple cycles, which eventually leads to cluster formation. Sequencing itself is done by synthesis. A primer is attached to the binding site of the adapter of the to be sequences strand. A polymerase then adds nucleotides to the DNA strand, however, only one base is added per round. All nucleotides are fluorescently tagged and give off different emissions when excited by a laser, which allows step by step sequencing of each single base of a fragment. Once a signal is detected, the fluorophore is washed away, and the sequencing process is repeated [223, 224]. This whole process is done in parallel.

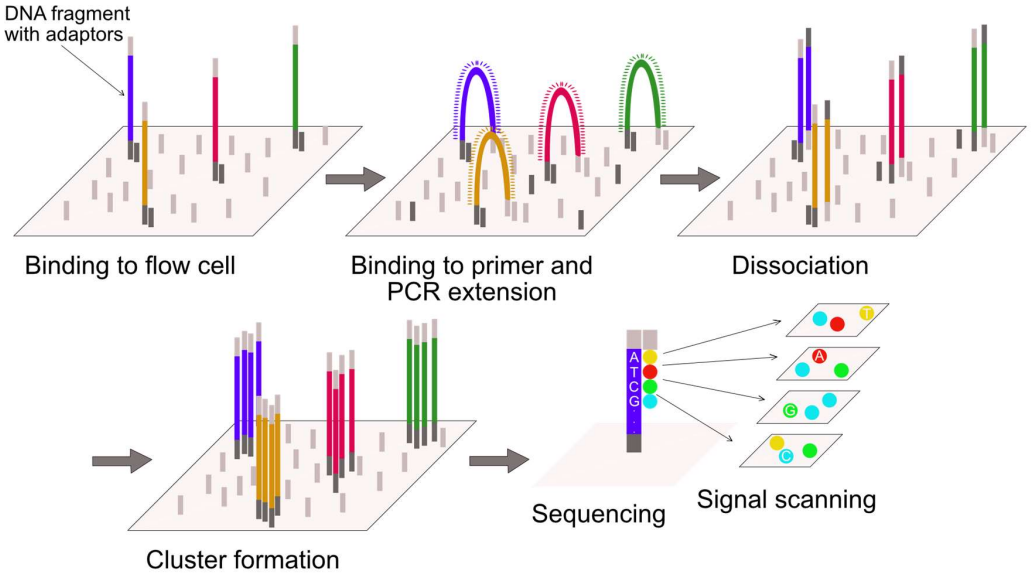

Figure 3.6. Working principle of Illumina sequencing starting with the binding of fragmented purified DNA, with adaptors, to the flow cell. Adapted from [224].

Updates to Illumina sequencing have been made throughout the years to further improve quality of the sequencing, as well as throughput. PCR-free library kits now exist, which improves the coverage of traditionally difficult to sequence areas such as regions high in AT/GC, promoters or homopolymeric regions [225]. The development of paired-end sequencing, where both ends of the library DNA fragments are sequenced and the forward and reverse reads are aligned as read pairs, helped improve NGS approaches drastically as it led to twice the amount of reads for the same amount of time spent on library preparation, while also leading to more accurate read alignment [226]. Single-read sequencing is still a preferred method for applications such as small RNA sequencing, while otherwise pair-end sequencing dominates [227]. While NGS offers high-throughput and affordable sequencing, there are still some improvements to be made. Illumina sequencing produces short reads of only up to 150 to 300 bp, which can be an issue for *de novo* sequencing approaches, as well as other projects that require longer reads [222]. Repetitive reads create uncertainties when it comes to alignment and genome assembly when using short reads [228]. Furthermore, assembling a genome with short reads can be difficult when there are no reference sequences available for the strain or even the species. However, Illumina sequencing is still somewhat more reliable and more robust, and can even be used in addition to the new Third Generation Sequencing (TGS) methods that have been developed over the past decade.

3.3.1.2 Third-Generation Sequencing – Long-Read Sequencing

Due to the challenges imposed by the short read length of NGS, advances towards high-throughput methods with longer reads were made. TGS is the result of these advances, with longer reads that make *de novo* genome assemblies easier as this sequencing provides bigger "puzzle pieces". Read lengths have gone from tens of bases to over tens of thousands of bases, while the time frame for sequencing has also been reduced to mere minutes or

hours, depending on the desired coverage and sequencing depth, as well as the genome size of the target species. The longer reads also aid in identification of repeat sequences, which, for example, makes it possible to measure telomere length [229] or identify tandem repeats that are important in certain genetic diseases [230]. Besides providing long reads and being more time efficient, the new methods are more high throughput and cost efficient than NGS. This applies not only to the sequencing process itself, but also to the actual sample preparation. This makes the slight disadvantages that still exist, namely higher sequencing error rates of 8-15% compared to 0.5% for Illumina sequencing, acceptable [231]. Furthermore, constant improvements are made on this end as well, reducing errors through improved sequencing tools as well as improved bioinformatics approaches, along with more and more companies presenting their long-read sequencing approaches, including Illumina [232].

Two companies driving TGS are Oxford Nanopore, which even offers portable devices to be used with any personal computer or laptop, and Pacific Biosciences, which offers a so-called Sequel platform that still requires sending samples off for sequencing as is done for the Illumina sequencing. The Pacific Biosciences sequencing platform is the size of a laminar flow bench, and acquisition is still quite cost-intensive. Oxford Nanopore offers a smartphone-sized sequencing platform called MinION, as well as benchtop solutions to run more sequencing projects simultaneously. The Oxford Nanopore devices are meant to be purchased for direct use in-house in the respective laboratories, meaning sample preparation after genomic DNA (gDNA) extraction is also performed in the laboratory directly.

The working principle of TGS differs vastly from the sequencing-by-synthesis approach employed by NGS. While NGS relies on PCR to amplify clusters of a target DNA template, TGS does not depend on PCR at all. Here instead, molecules are directly targeted leading to so-called real-time sequencing. For Oxford Nanopore technologies, after loading of the sample onto a flow cell and the start of the sequencing, reads are available as soon as the molecules pass through the sequencer or nanopore. A constant ionic current passes through this nanopore (of which there are hundreds on a flow cell) as constant voltage is applied across the sequencing membrane. For DNA sequencing, first a double-stranded DNA molecule is unwound, and the resulting single-stranded DNA passes through the nanopore driven by voltage. Each nucleotide that passes through the nanopore results in a current change, with distinct current changes corresponding to a specific base. The current changes are measured parallel throughout the entire flow cell, and nucleotides can be determined at a rate of around 450 bases per s [233]. This enables immediate analysis and, for Oxford Nanopore sequencing, a detailed, live overview over the amount of data generated, meaning that one can adjust the running time of the sequencing dynamically. An overview about read lengths and quality can also be accessed immediately. Mean read lengths of around 9 kb have been reported for various yeast species [234]. However, this number is constantly improving with improvements in the flow cells and library preparation kits, and is also dependent on the quality and size of the input DNA, as well as the species to be sequenced. In both **Paper II** and **Paper III** higher mean read lengths were achieved.

Sample preparation kits are targeted towards specific applications, making the sequencing process also highly aimed at a specific objective. However, prior to that, specific extraction kits need to be used that are more gentle, leading to longer DNA fragments. Most of these kits have been tested on *S. cerevisiae*, but not on many non-conventional yeasts, making it more difficult to find suitable extraction kits, and leading to a longer time period needed to optimise the workflow. This is a problem in general for working with non-conventional yeasts, as microbiological tools for yeast are often tailored towards *S. cerevisiae*, which is a distant relative to any basidiomycete.

For the long-read sequencing used in **Paper II** and **Paper III**, Oxford Nanopore sequencing using their portable MinION platform was conducted. Two different gDNA extraction kits were tested and troubleshooting was attempted in order to find a good extraction method that yielded long DNA fragments as well as high amounts of good quality DNA. While the QIAGEN Genomic Tip Kit seemed to be a favourite among people seeking high molecular weight DNA, the extraction process (including all incubation steps) took about 10 hours and did not yield DNA at all for one of the two non-conventional yeasts used in this work, while only producing very little DNA for the other. Even after troubleshooting and process optimisation, results were not satisfactory. The Invitrogen Easy-DNA gDNA Purification Kit in turn resulted in high amounts of high-quality DNA, which was tested using a low-percentage agarose gel, in merely half the time needed for the Genomic Tip Kit, and only slight adjustments of an existing extraction protocols were necessary. High-quality high molecular weight DNA was generated for both BOT-O and BOT-A2 using the same parameters. Fragments were determined to be around 48,000 bp in length or higher, and were then prepared for long-read sequencing using the Ligation Sequencing Kit from Oxford Nanopore. The instructions of the kit specify a range of gDNA of about 5-50 fmol for loading onto the flow cell for sequencing. It was found that while within the range for sequencing, aiming for the higher end of the range as was done for BOT-O led to a significant increase in data and decrease in sequencing time, as well as preservation of the flow cell. For sequencing of BOT-O, 33 fmol of gDNA were used, resulting in over 1 Gb of sequenced bases in only about 4 h. For BOT-A2, 12.8 fmol of gDNA were loaded onto the flow cell, resulting in about 830 Mb of data in about 24 h. Quality-wise, as assessed by using a Qubit v3 fluorometer and the 0.5% agarose gel, the gDNA used was similar, therefore it can be assumed that the amount of gDNA loaded onto the flow cell must have played a role in the sequencing outcome. Nevertheless, high-quality data was generated for both yeasts, with more coverage for the BOT-O sequence.

3.3.1.3 Genome Assembly

Signals from the sequencing of the MinION device can be base-called immediately. During base-calling, raw signals corresponding to the different bases are translated into nucleotide sequences and then fed to downstream analysis, usually via fastq reads [235]. These can be read by a computer and contain information about the specific base itself and the quality of the read. There are a variety of tools available for base-calling, however, some are more

specific towards certain sequencing platforms or species; Guppy (Oxford Nanopore Technologies) was specifically created to base-call reads generated using the Oxford Nanopore MinION or related platforms. The base-called data are then used to continue on in the bioinformatic workflow to assemble (Figure 3.7) and annotate the genome.

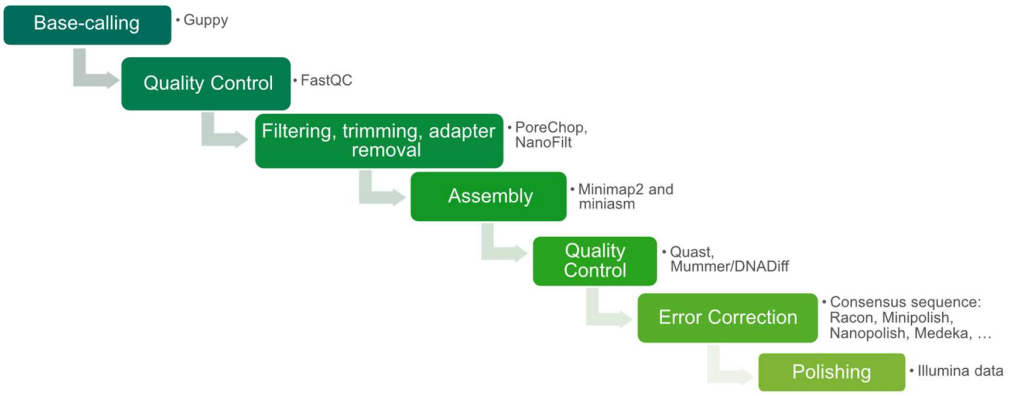

Figure 3.7. Bioinformatics workflow from base-calling to genome assembly for non-conventional yeasts, with the main steps and suggested tools used in this work. More details about the exact tools and settings used can be found in the method sections of **Paper II** and **Paper III**.

Assessment of quality of the data and the assembly is done continuously throughout all the bioinformatic steps needed. Tools like FastQC [236], which can be used for this task, are universal and can be used for data from a wide variety of microorganisms and any sequencing platform. After quality control of the base-called data, reads are filtered and trimmed, and adapters can be removed. This cleans up bad-quality reads and generally improves the quality of the subsequent assembly. The quality of single reads also generally differs throughout, with decreases in quality towards their end. Thus, trimming can be used to discard low-quality parts of reads. Overall low-quality reads can be filtered out using the information contained in the fastq files of each read. Here again, a variety of tools is available, however some tailored ones for specific sequencing platforms, such as Nanofilt [237], exist. Once reads have been trimmed and filtered, they can be assembled and mapped. Once again, diverse tools exist, among them minimap2 and miniasm [238, 239], which are specifically tailored towards long-read Oxford Nanopore data and can be used for mapping and assembling of the genome. Error correction of the preliminary genome assembly is necessary to obtain a high-quality assembly, as it, as indicated by the name, minimises errors. Tools like Nanopolish have been specifically developed for long-read data, however generally a combination of multiple tools is used for this step. A polishing step can be added if, for example short-read Illumina data exists for a *de novo* long-read assembly. This step has been shown to increase the assembly quality of long-read sequences previously [240], but did not improve the qualities of the BOT-O and the BOT-A2 genome assemblies in **Paper II** and **Paper III** drastically. In general, however, genome assembly, and specifically for non-Illumina data, is still somewhat difficult as tools only work in limited contexts and choosing the right tool for the right type of data is crucial in order to produce a high-quality genome assembly and annotation.

For the genome assembly of both BOT-O and BOT-A2, different assemblers were tested and benchmarked (**Paper II** for BOT-O, Table 3.1 for BOT-A2), as it had been noted previously that outcomes can differ greatly when *de novo* sequencing non-conventional yeasts [241]. All assemblers were used with their default settings, and even though assembly outcomes could potentially have been improved by further optimisation of certain parameters, this was judged as sufficient based on published assemblies for other strains o the same species. During the benchmarking of both yeasts, miniasm (v0.3_r179; [238]) was found to result in the most contiguous and high-quality assembly based on the sequencing parameters shown in Table 3.1. These assemblies were then further error corrected and subsequently polished with Illumina short-read sequencing data generated from the same gDNA as the long-read sequencing data. During a revision for **Paper II**, SMARTdenovo (v20180219-5 cc1356; [242]) was tested again, and instead of running it in the default setting with its wrapper script, the individual underlying sub-algorithms were called on with previously suggested parameters [243]. This in turn led to a very contiguous assembly in contrast to the default setting SMARTdenovo assembly, which was able to compete with the chosen miniasm assembly.

Table 3.1. Assembly benchmarking using five different assemblers for reads from *R. toruloides* BOT-A2.

Assembly metrics	**miniasm**	**Flye**	**Canu**	**Shasta**	**SMARTdenovo**
Number of contigs	20	26	32	1131	34622
Largest contig [bp]	2,177,041	2,200,678	2,199,517	2,201,743	168,146
Contigs ≥ 50,000 bp	19	23	26	25	3043
Total length [Mb]	20.28	20.89	21.43	22.26	854.1
GC content [%]	61.52	61.74	61.47	60.72	59.32
N50 [bp]	1,290,001	1,308,257	1,323,734	1,026,168	34,096
N75 [bp]	927,593	971,490	937,874	778,932	21,431
L50	7	7	7	8	8883
L75	12	11	12	14	16684

Only three genome sequences are currently available on NCBI for *P. hubeiensis* (June 2023) [46], with the BOT-O strain sequence from **Paper II** being the most contiguous sequence to date. Only two sequences have been annotated, including the BOT-O sequence from my work and the SY62 genome. They have a similar genome size and GC content, with 18.95 Mb and 18.44 Mb, and 56.27% and 56.50%, respectively. Interestingly, *P. hubeiensis* is a relatively intron-poor yeast species [244], and the BOT-O genome sequence revealed about 0.5 introns per gene. For *R. toruloides*, more genome sequences are

available, however, the BOT-A2 sequence is also one of the most contiguous *R. toruloides* ones to date.

3.3.1.4 Functional Genome Annotation

Genome annotation for novel assemblies, especially of eukaryotes, is a challenging computational task that involves many steps and iterative training of the annotation pipeline. Annotating a genome is also arguably more difficult than assembling one, leading to the fact that there are many more genome assemblies available than fully annotated ones. The annotation process consists of three main steps, including the identification of noncoding regions, the identification of coding regions, which is the actual gene prediction, and functional annotation by attachment of the biological information of these identified elements [245]. However, due to many genes receiving a functional annotation based on homology rather than experimental evidence, and the likelihood of the homologue also being based on homology, errors and uncertainties can occur. Traditionally, functional annotation was done manually, which was quite labour-intense. Protein coding sequences were predicted using a computer method based on codon usage, and the assumption that the frequency of a specific codon is correlated with the abundance of its corresponding tRNA [246, 247]. The whole annotation process is now generally performed using a pipeline, with very few incidences of manual curation [248].

An overview over the annotation process applied for genome annotation of the BOT-O and BOT-A2 assemblies can be found in Figure S5 in Supplemental File S1 in **Paper II**. As for genome assembly, many tools are available and choosing the right one for the type of data that is dealt with requires careful consideration. Some tools are available online, e.g. a genome annotation server by NCBI or Ensembl and have the advantage of offering an automatic pipeline and no requirements for computational power on the user side. They generally offer a user-friendly interface as well [245]. However, these may only accept smaller assemblies and might not be feasible for assemblies of bigger organisms such as yeasts. Furthermore, the command line interface annotation tools provide the possibility for tweaking and tailoring the annotation process to the individual project. In **Paper II** and **Paper III**, the MAKER pipeline (v3.01.2-beta; [249]), a command line interface tool, was run iteratively in three rounds to build gene models for both *R. toruloides* BOT-A2 and *P. hubeiensis* BOT-O. To improve the *de novo* genome annotation, optional extrinsic evidence from RNA sequencing and proteomics data were used. As input for the genome annotation of *R. toruloides* BOT-A2 and *P. hubeiensis* BOT-O, the final genome assemblies were used, as well as RNA sequencing data for BOT-A2 and BOT-O, which will be addressed in the next chapter, available proteome data for other strains of the same species, and filtered repeat sequences that were identified in the BOT-O and BOT-A2 assemblies. Specifically, RNA sequencing data have been shown to be of great value for the genome annotation [245], aiding in identification exons and splice sites, as well as alternatively spliced exons [248].

The functional annotation was performed using a separate pipeline from the gene model generation using MAKER. While assigning gene names according to previous annotations

of strains from the same species might seem natural, it was decided that assigning *S. cerevisiae* gene names would be beneficial for comparison of results between different species. Therefore, identified proteins were first queried against *S. cerevisiae* proteins (Uniprot proteome: UP000002311), and then against other proteins of the same species and other basidiomycetes. Details for the specific proteins queried can be found in **Paper II** and **Paper III**, as all of the bioinformatics workflows were carefully documented. A table with all predicted proteins and their functional annotation can be found in Supplemental File S2 of both papers. Results were also validated using Orthofinder (v 2.5.2; [250]).

Manual curation of genes is unavoidable and was performed for a few genes of interest. For example, *ERG13* was not found in the original functional annotation for *R. toruloides* BOT-A2, but was certain to be present in the genome as it is necessary for the production of carotenoids, which were metabolised by this strain as evidenced by their colour in the physiological characterisation (see above). It was detected that it had originally been missing from the functional annotation, as well as the list of predicted genes. A look at the genome assembly revealed that the protein sequence for *ERG13* overlapped with that of another protein, which was annotated as being hypothetical, explaining the missing gene prediction. Physiological characterisation therefore continues to be an important tool and can also be used to corroborate bioinformatics results. Other examples for manual curation include the genes involved in the production of MELs in BOT-O. Since genes were first assigned functional annotations based *S. cerevisiae* proteins, and only then based on basidiomycete proteins, some of the functional annotations for the MEL pathway were missing. They were therefore added manually using blast to find hits for the *U. maydis* MEL genes in the annotation for BOT-O. Interestingly, most of these genes were found to be located back-to-back in a cluster and sequences were highly identical to those of the *U. maydis* MEL genes. Potentially, some of the manual curations could have been avoided if a different strategy would have been employed for functional annotations. Some genes were perhaps falsely functionally annotated with a *S. cerevisiae* annotation at a lower confidence, when they could have received a functional annotation based on a more closely related basidiomycete. However, once a protein received a functional annotation at the confidence parameters set (set for query coverage and p-value), the annotation for already annotated proteins was not updated as they were not queried again. This was preferred as it resulted in a higher number of *S. cerevisiae* names in the annotation, and more were also added manually after the reconstruction of the pathways in the Figure 3 of both **Paper II** and **Paper III**. For research purposes, working with those established gene names helps with comparability between different species. Unfortunately, gene names are not standardized at all when it comes to non-conventional yeasts, and a collectively agreed upon gene naming convention for all species would be of great help and able to avoid confusion among the researchers working with non-conventional yeasts.

3.3.2 RNA Sequencing

Besides being a great aid for the genome annotation, RNA sequencing is a valuable method in its own rights, allowing an insight into the mechanisms within the cell in terms of transcription pattern during specific conditions and developmental stages. It can provide

extensive information on, for example, the abundance of mRNA, nucleotide variation, and alternative splicing, at chosen conditions and times [251]. During work for this thesis samples for RNA sequencing were, for example, taken during the exponential growth phase for both yeasts, and during nitrogen-limited conditions. Both time points were considered for growth on either glucose or xylose. The research interest here was to investigate transcriptome differences between exponential growth and nitrogen-limited cultivation, the differences between growth on different carbon sources, particularly glucose and xylose, and the differences between nitrogen-starved cultivation between the two different sugars. Differential gene expression analysis was used to compare the expression of genes across different conditions, and is in general one of the most common applications for RNA sequencing data [252]. RNA sequencing itself is a relatively new method for transcriptomic analysis, having been developed in 2008 [253], and was described as a revolutionary tool soon after its introduction in 2009 [254]. Before the rise of RNA sequencing for the analysis of the transcriptome, microarrays were state of the art. However, after development of RNA sequencing, it became clear that the yeast transcriptome was more complex than was initially detected when using microarrays. Furthermore, it could already be shown in 2008 that RNA sequencing had a wider dynamic range than the previously used microarrays [253]. Transcription levels can, furthermore, be more accurately determined using RNA sequencing [254], with lower background noise than microarrays.

Principally, after extraction of the RNA and conversion of the RNA into cDNA, the sequencing of cDNA is performed similar to other DNA samples. Nanopore sequencing actually allows for direct sequencing of RNA, without conversion to cDNA, removing bias introduced due to library amplification via PCR [255]. After sequencing, the transcriptomic reads are then mapped to the genome assembly, meaning that an assembly, ideally from the same strain and not just the same species, is available. In contrast to genome sequencing where only one sample is sequenced, more replicates are needed for RNA sequencing. This is due to the transcriptome being never exactly the same and very dependent on the cultivation conditions and the sampling time points, thus more replicates are necessary to achieve robust and reproducible transcriptome results. A minimum of three replicates is generally required for reliable and reproducible RNA sequencing results [256]. Some of the most used tools perhaps are edgeR [257] and DESeq2 (v1.26.0; [258]), with the latter having been used during the RNA sequencing results analysis in **Paper II** and **Paper III**. This was partially due to the fact that edgeR has previously been described as oversensitive and leading to results with high variability when dealing with fewer than five replicates [259].

Before sequencing and data analysis of the RNA can be performed, RNA has to be extracted from the cells. This is considerably more demanding than extraction of DNA and even the extraction of high molecular weight gDNA due to the lower stability of the single-stranded RNA. RNA degrades much more easily, and RNA-degrading enzymes are omnipresent in the environment. The use of low-quality, i.e. (partially) degraded, RNA samples for RNA sequencing affects the expression levels measured [260]. For measuring the level of RNA degradation, a few approaches have been suggested, with the RNA

Integrity Number (RIN) playing a major role. The RIN is based on a predictive model to determine the integrity of RNA, which in turn is based on multiple features including the height of the 28S peak and the total RNA ratio [261]. Unfortunately, the RIN value is generally calculated based on a few model organisms, causing inaccurate RIN calculations when working with non-conventional yeasts. Determination of the level of degradation is not the only issue, as RNA extraction protocols and kits are once again tailored towards more conventionally used yeasts such as *S. cerevisiae*. In this work, two different approaches for RNA extraction needed to be developed for the two yeasts *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2. The extraction methods differed especially in the steps leading up to the cell lysis and in the lysis parameters themselves. This is potentially due to differences in cell wall composition or morphology, that differs more between these two yeasts and *S. cerevisiae*, but perhaps also between these two different basidiomycetes. The exact protocols are described in **Paper II** for BOT-O, where the QIAGEN RNeasy kit was used with experimentally determined mechanical lysis settings, and in **Paper III** for BOT-A2, where TRIzol was used in the beginning of the extraction, and mechanical lysis settings differed. RNA concentrations and quality were determined using a Qubit v3 fluorometer and the Bioanalyzer, which calculates the previously mentioned RIN values.

For RNA used in **Paper II** and **Paper III**, RIN values seemed to be lower for both of the two non-conventional yeasts than recommended values for RNA sequencing, but since RIN values are calculated based on *S. cerevisiae* as a standard an no other yeast option is available in the analysis program for RIN calculations, this was determined to not be concerning, and ascribed as a species-specific feature rather than actual RNA degradation. Differences between non-conventional yeasts and *S. cerevisiae* can be expected. RIN values were also observed to be generally lower during the nitrogen starvation conditions, which is potentially due to lower levels of ribosomes, and thus less ribosomal RNA. It has also been shown that a sensing of nutrient depletion is related to downregulation of ribosomes in *S. cerevisiae* [262], and specifically nitrogen starvation has been shown to have a significant effect [263]. This downregulation was also observed by other studies in *R. toruloides* and other yeasts [212, 264, 265].

While **Paper II** presents the first transcriptomics study for a strain of the *P. hubeiensis* species, there are already a few transcriptomic studies available for *R. toruloides*, even with a focus on nitrogen starvation as presented in **Paper III** for BOT-A2 [212, 266]. However, in these studies, sampling time points were set at random intervals, whereas they were tied to the actual nitrogen concentration in the different media in this work. Due to this, and other factors such as different sampling strategies and the use of different bioinformatics tools for the analysis, comparing exact numbers for differential gene expression across different studies is difficult.

3.3.3 Proteomics

While transcriptomics and RNA sequencing can give answers as to how much a certain gene is expressed at a certain time, it does not directly translate to how much corresponding protein is actually present. Proteomics is needed to do that. Due to the increasing

affordability of both of these techniques, applying both for a given project is becoming increasingly popular. In fact, proteomics can be a great addition to RNA sequencing, validating novel findings from RNA sequencing data such as especially high or low expression of genes of interest during certain conditions [251]. It has also been suggested that posttranscriptional regulation plays a major role in controlling gene expression as the overall protein levels do not highly correlate with the overall mRNA levels [251]. Protein degradation plays a role, too. However, a correlation has been found for differentially expressed proteins and their mRNA levels [267]. Proteomics data can also be used for genome annotation, which was done in this work, as previously described. However, previously existing data, also from other strains, were used for this purpose for the annotation of the BOT-O and BOT-A2 genome.

Proteomics sampling can present similar difficulties as genomics and transcriptomics sampling when working with non-conventional yeasts, and sampling and extraction protocols might have to be adapted from the standard *S. cerevisiae* methods. Utilised methods can affect the quality of the resulting proteome data and can alter the proteome coverage and the information about post-isolation modifications [268]. In contrast to RNA sequencing and genome sequencing, samples are not sequenced but are analysed using mass spectrometry.

3.4 Comparing the Genomics and Transcriptomics Results of *Pseudozyma hubeiensis* BOT-O and *Rhodotorula toruloides* BOT-A2

While transcriptomics and genomics are discussed for each of the two non-conventional oleaginous yeast strains in the respective papers (**Paper II** for *P. hubeiensis* BOT-O and **Paper III** for *R. toruloides* BOT-A2), in this part of the thesis the two strains will be compared, and interesting differences and similarities will be discussed. They have been treated similarly so that sampling times and culture conditions for RNA sequencing are directly comparable, and have furthermore been analysed in the same manner when it comes to bioinformatics for both the genome sequencing and the RNA sequencing.

While both yeasts are non-conventional oleaginous yeasts, they do differ quite a bit. As discussed in the previous chapter, *R. toruloides* is a known natural producer of carotenoids in addition to the storage lipids, whereas *P. hubeiensis* produces MELs and can grow on glucose and xylose with similar growth rates. Furthermore, although they both are basidiomycetes, they are more closely related to other species in the phylum as they belong to different families. (Figure 2 in **Paper II**). The genome assembly of BOT-A2 is more contiguous than that of BOT-O, with only 20 contigs compared to 31, and larger with 20.28 Mb compared to the 18.95 Mb for the BOT-O assembly. The total number of chromosomes for *R. toruloides* has been predicted to be 16 for the NP11 strain by Zhu et al. [212], and estimated to be at least 18 by Martín-Hernández and colleagues for the CBS14 strain [269]. For *P. hubeiensis*, no focus has been put on determining the number of chromosomes yet.

Another significant difference, which also affects the genetic engineering possibilities, is the GC-content. Both have a high GC-content compared to *S. cerevisiae* (at around 40%) with 56.27% and 61.52% for BOT-O and BOT-A2, respectively. These numbers are comparable with other strains in their respective species, and also with other oleaginous yeast strains [270]. A high GC content can be a problem for (PCR) based cloning and gene amplification. It also makes sgRNA design for CRISPR-Cas9 engineering more complicated [271]. However, successful engineering strategies for *R. toruloides* do exist, albeit mainly with non-targeted integration as presented in **Paper III** and some other publications [272-274]. Most transformation strategies involve an *Agrobacterium tumefaciens* mediated approach, whereas here in **Paper III** an *A. tumefaciens*-free electroporation method was used. Engineering of *P. hubeiensis* BOT-O has so far been unsuccessful, but this is not necessarily due to the higher GC content. Engineering in some of its closer relatives, i.e. *U. maydis* and *Moeszyomyces antarcticus*, which also do have higher GC contents with 60.5% and 54%, respectively, has been achieved successfully as previously mentioned [164-166].

As fungi, both *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2 possess the fatty acid synthase (*FAS*) genes. Generally, there are two subunits, the fatty acid synthase subunit α encoded by *FAS2*, and the fatty acid synthase subunit β encoded by *FAS1*. However, the number of domains and which domains can be found on each subunit differs between different fungal species [275]. For *S. cerevisiae*, *FAS1* is slightly longer than *FAS2* and has more domains, as shown in Figure 3.8. The *S. cerevisiae FAS1* contains seven domains according to an HMMER Web Server [276] analysis, including the N-terminal fatty acid synthase subunit β domain (FN), the starter unit ACP transacylase domain (SAT), a domain of unknown function (DUF), the fatty acid synthase meander β sheet domain (FM), the N-terminal half of MaoC dehydratase (MCN), the C-terminal half of MaoC dehydratase (MC), and the acyl transferase domain (AT). The *S. cerevisiae FAS2* contains five domains, including the fatty acid synthase subunit α acyl carrier domain (ACP), the fatty acid synthase type I helical domain (FIH), the N-terminal ketoacyl synthase domain (KSN), the C-terminal ketoacyl synthase domain (KSC), and the phosphopantetheine transferase (PPT). According to Zhu et al. [275], *R. toruloides* has a shorter *FAS1*, with most domains contained in the *FAS2* and a doubling of the fatty acid synthase subunit α acyl carrier domain. This could be corroborated for the BOT-A2 strain (Figure 3.8), with four domains on the BOT-A2 *FAS1*, and a total of ten on *FAS2*, with FM being present on both FAS genes, and a doubling of the ACP on *FAS2*. Furthermore, the *FAS1* gene of BOT-A2 (RTBOTA2_004415) seemed to contain unrelated domains prior to the N-terminal fatty acid synthase subunit β domain which were not found in any of the other analysed *FAS* genes for the other investigated species. This could be due either to a wrong, or earlier starting annotation of this.

When it comes to *P. hubeiensis*, no separation of the fatty acid subunits was detected, and a gene encoding a complete fatty acid synthase, including all the twelve domains found on the α and β subunits, was found in the genome (PHBOTO_002928; Figure 3.8). In addition, another fatty acid synthase gene copy was found, containing all domains except for MCN

and PPT (PHBOTO_006517; Figure 3.8). A similar organisation of the fatty acid synthase was observed in the genome of *U. maydis*, which had previously been shown to possess a single-chain *FAS* gene [121, 277]. However, no publication describes the fact that there are two of these present in *U. maydis* as was discovered during work on **Paper II**. UMAG_06460 is actually listed as *FAS2* in Uniprot (Uniprot ID: A0A0D1DNX1), although it is a single-chain fatty acid synthase. Since *FAS1* and *FAS2* were considered as being misleading gene names for the two single-chain *FAS* genes in BOT-O, they were named *FASA* and *FASB*, respectively.

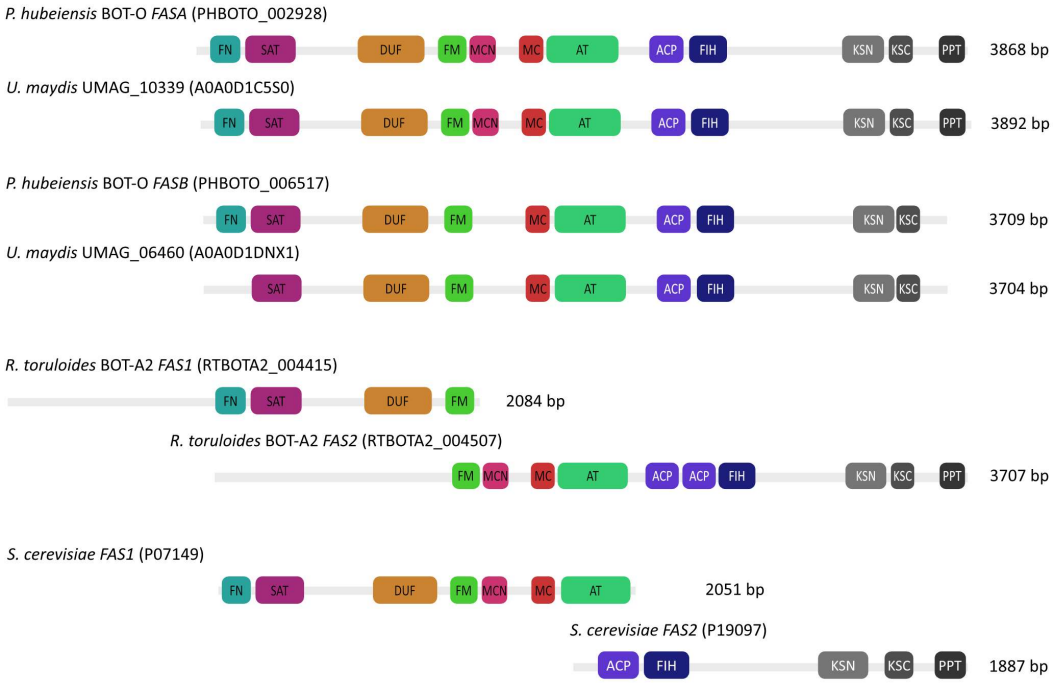

Figure 3.8 Alignment of the *FAS* genes from *P. hubeiensis* BOT-O, *R. toruloides* BOT-A2 compared to *U. maydis* and *S. cerevisiae* using the HMMER web server [276]. Domains for *FAS* candidates and the similarities in structure are presented, with full *FAS* genes for *P. hubeiensis* and *U. maydis* and split *FAS* with *FAS1* and *FAS2* in *R. toruloides* and *S. cerevisiae*. Abbreviations: FN: N-terminal fatty acid synthase subunit β domain; SAT: starter unit ACP transacylase domain; DUF: domain of unknown function; FM: fatty acid synthase meander β sheet domain; MCN: N-terminal half of MaoC dehydratase; MC: C-terminal half of MaoC dehydratase; AT: acyl transferase domain; ACP: fatty acid synthase subunit α acyl carrier domain; FIH: fatty acid synthase type I helical domain; KSN: N-terminal ketoacyl synthase domain; KSC: C-terminal ketoacyl synthase domain; PPT: phosphopantetheine transferase domain. Domains are named according to the HMMER web server.

For RNA sequencing, samples were taken in a similar manner and during similar cultivation conditions. The detailed conditions are described in **Paper II** and **Paper III**. Due to the high similarity in growth on glucose or on xylose for *P. hubeiensis* BOT-O compared to *R. toruloides* BOT-A2, samples for exponential growth and during the nitrogen-depletion phase were taken at the exact same time point, whereas those points differed for BOT-A2 depending on the sugar (Figures 3.4c and 3.4g). Bioinformatic analysis was conducted in a similar way, allowing comparisons between the two strains. Pathways were reconstructed for both BOT-O and BOT-A2 (Figure 3 in **Paper II** and

Figure 3 in **Paper III**), with added information about differential gene expression, comparing nitrogen starvation on either glucose or xylose with exponential growth on either sugar, and differences between growth on either sugar during exponential growth or the nitrogen starvation phase. Overall, more differentially expressed genes were found for BOT-A2 compared to BOT-O. The total number of differentially expressed genes in accordance with the set cut-off parameters of log2 fold change $\geq |2|$ and a p-value $\leq 10^{-6}$ in the four different comparisons g2g1, x2x1, x1g1, and x2g2 is listed in Table 3.2. A specifically high number of differentially expressed genes was noted for nitrogen starvation during growth on xylose (x2x1), and exponential growth on xylose compared to glucose (x1g1). It should also be noted that the total number of differentially expressed genes does not differ much for BOT-O when comparing the two nitrogen starvation comparisons g2g1 and x2x1 and the two sugar comparisons x1g1 and x2g2, which was expected given the similar growth on either sugar. For BOT-A2, the result was different, and in general a much higher number of genes was found to be significantly differentially expressed during nitrogen starvation compared to BOT-O. Furthermore, the number of differentially expressed genes was also higher on xylose compared to glucose (x2x1 vs. g2g1). Nitrogen starvation might have had a bigger impact on the transcriptome during growth on xylose due to the added stress of growing on a non-optimal carbon source, since *R. toruloides* BOT-A2 obviously does prefer glucose. Similarly, differences between growth on xylose and growth on glucose could be less during the nitrogen starvation phase (x2g2) compared to the exponential growth phase (x1g1) due to the fact that the cells are already subjected to the stress of limited nitrogen.

For BOT-O, few differences were detected for any of the comparisons, both during nitrogen starvation (g2g1 vs. x2x1) and the growth on different sugars (x1g1 vs. x2g2) (Table 3.2). The potentially most significantly differentially expressed genes in the pathway map are genes involved in MEL formation. These were highly upregulated during nitrogen starvation on both glucose and xylose. Furthermore, *FASB*, the fatty acid synthase containing most, but not all, domains as demonstrated in Figure 3.8, was somewhat upregulated during nitrogen starvation, whereas the expression of *FASA* was not significantly changed but seemed to be continuously expressed during all conditions. This could perhaps indicate that, in the case of BOT-O with both FAS genes containing almost all main domains, one FAS gene, *FASA*, is continuously expressed, leading to the base level production of lipids, while the other one, *FASB*, is only highly differentially expressed during certain conditions, resulting in high lipid production for example during nitrogen starvation. For genes that are part of the glycolysis, the pentose phosphate pathway (PPP), or pentose sugar utilisation, expression of only the xylitol dehydrogenase *XYL2* and the xylulokinase *XKS1* was somewhat, but not highly, increased on xylose during both exponential growth (x1g1) and nitrogen starvation (x2g2). This is unlike the differential expression noted in BOT-A2. Here, glycolysis and PPP genes seem to be more affected by the choice of carbon source, partially explaining the differences in the effect of the different carbon sources on the growth of these strains. *XYL2* and *XYL2_RT* (a *XYL2* based on homologies to another *R. toruloides* strain), as well as *XYL1_1*, were highly upregulated during growth on xylose compared to glucose. Downregulation was observed for glycolysis

genes and many of the tricarboxylic acid (TCA) cycle genes during nitrogen starvation on the different sugars. Downregulation was also observed for the first steps of the carotenoid pathway in BOT-A2. In BOT-A2, *FAS1* and *FAS2* were both slightly upregulated, in contrast to the *FAS* genes in BOT-O, where only *FASB* was found to be upregulated.

Table 3.2. Total number of differentially expressed genes during all the compared conditions for *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2. Conditions are defined as follows: g2g1 – differential gene expression during nitrogen starvation on glucose; x2x1 - differential gene expression during nitrogen starvation on xylose; x1g1 - differential gene expression during exponential growth on xylose compared to glucose; x2g2 - differential gene expression during nitrogen-starvation on xylose compared to glucose. The total number of predicted genes was 6525 for BOT-O, and 7001 for BOT-A2.

Condition	Total # of differentially expressed genes for BOT-O	Total # of differentially expressed genes for BOT-A2
g2g1	946	1192
x2x1	923	1514
x1g1	73	392
x2g2	73	94

Although BOT-O seems to be a good candidate for the investigation of efficient pentose sugar utilisation and thus also for bioproduction from renewable plant-based feedstocks, and potentially as a gene donor for improvement of pentose utilisation in other yeasts, engineering for this strain has yet to be achieved. This is an important factor for example when it comes to investigating the actual role of certain genes, for which knock-out or overexpression strategies are required. The advantage for BOT-A2 here is that engineering has already been achieved, albeit non-targeted. The new genomics and transcriptomics data provided in this work will hopefully be able to help improve knowledge and thus aid in finding and applying new genetic engineering approaches.

Chapter 3: Summary

- Morphology and physiology are necessary to investigate the phenotype of a new strain but can lead to mistakes in classification.
- BOT-O grew equally well on glucose and xylose as carbon source, but preferred ammonium sulphate over urea as nitrogen source.
- BOT-A2 growth rates were lower on xylose than on glucose, but can be improved with changes in cultivation parameters.
- The rise of third-generation sequencing improves *de novo* genome assembly and offers higher throughput, lower costs and better accessibility in some cases.
- High quality, contiguous genome sequences were assembled for both BOT-O and BOT-A2 using a combination of short- and long-read sequencing.
- Growth on xylose compared to growth on glucose leads to a higher number of differentially expressed genes in BOT-A2 than BOT-O.

Chapter 4: About Bioproducts

With the decrease in availability of fossil fuels and the increasing impact of climate change comes a rising interest in the production of sustainable bioproducts. The focus of those bioproducts is expanding from food and drink fermentations towards more high-value and complex products such as enzymes, pharmaceuticals and fine chemicals, bulk chemicals, primary and secondary metabolites, biosurfactants, heterologous proteins etc. Yeast has great potential to be a host organism to a multitude of these products and there is a large market that still grows [37, 43]. While some hosts are used for the production of native metabolites, a large variety of non-native products are also of interest. Although *S. cerevisiae* is still the prime production candidate, interest in developing non-conventional yeasts for bioproduction is steadily increasing [278-280]. The two yeasts discussed in this work, *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2, were shown to be good native producers of storage lipids, unlike *S. cerevisiae*. Storage lipids can, for example, be used as food and feed. Furthermore, MELs, produced by BOT-O, can be used as biosurfactants, and carotenoids, produced by BOT-A2, are valuable metabolites for different industries, including in the food and pharmaceutical sectors [281]. All these products share acetyl-CoA as the common precursor. Acetyl-CoA can also be an important precursor for a wide range of non-native products. Thus, the potential for non-native production in these two non-conventional yeasts is high, given that acetyl-CoA flux can be redirected towards a desired product.

4.1 Important Bioproducts for a More Sustainable Economy

One of the largest bioproduct industries is bioethanol production. While early car engines ran on fuels such as vegetable oils, no further research was conducted into biofuels due to the domination of petroleum-based fossil fuels. However, interest rose again during the global oil crisis in the mid-1970s [282]. Brazil was the first country to introduce bioethanol made from sugarcane into their fuel market in the 1980s and is still one of the main bioethanol producers worldwide [282]. With the expected decline in fossil fuels and constant rise of oil prices, combined with the need to reduce greenhouse gas emissions, more and more countries have started to follow suit and start to fund research in and production of bioethanol.

Bioethanol production is possible from a variety of crops and crop waste, making it an attractive fuel source for worldwide production [283]. So far, bioethanol has mainly been used as an additive to commonly used petroleum fuel [284]. Although it has been shown that such blends increase fuel consumption slightly compared to traditional petroleum fuel, the decrease in carbon monoxide and unburned hydrocarbon emissions is substantial, and

an increase in the addition of bioethanol leads to an even bigger decrease [285, 286]. Overall, bioethanol use leads to a decrease in greenhouse gas emissions [287], although only if the bioethanol is produced from waste products and thus is a so-called second-generation biofuel [288]. Otherwise one must account for emissions caused by land-use change [288]. the use of fertiliser for cultivation of the crop, which also is dependent on the type of plant [289] and the use of fossil fuels for the harvesting and processing of the crops for biofuel production [290]. Many of the commercially produced biofuels are first-generation biofuels [291]. They are produced from sugar- and starch-rich energy crops such as sugarcane, which is predominant in Brazil, or corn, commonly used in the U.S. [292]. Biofuel production from these sources directly competes with food and feed production, and are not as climate friendly as second-generation biofuels [288], although they are more economically competitive than their sustainable counterpart. [292, 293]. Therefore, research now focuses on second-generation biofuels from for example agricultural, forestry or municipal waste, and on the discovery of more competitive and efficient production using engineered *S. cerevisiae* strains [294]. In order to increase efficiency of second-generation fuel production, many factors need to be improved. One is the efficient utilisation of xylose. This is a native advantageous capability of both *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2, explored in this thesis. Furthermore, various oleaginous yeast strains have been shown to be excellent hosts for the production of biodiesel [295], another biofuel.

Bioplastics production, biobased from renewable feedstocks, and/or biodegradable, has also attracted recent and growing attention. Most plastic used today is based on fossil fuels and its production contributes to about 6% of all greenhouse gas emissions [296]. While laws and regulations have led to a slower increase in fossil-based plastics production in Europe, the increase worldwide is still substantial. Over 50% of plastics worldwide are produced in Asia. Packaging is the most common use for plastics, followed by construction purposes and the automotive industry [296]. While plastic recycling has increased, a lot of plastic waste still ends up in nature, and specifically oceans. Nevertheless, plastic products have become a necessity in our societies and are, moreover, important for, for example, the medical industry, for food packaging to prevent early spoilage, or as insulation in newly constructed buildings [296]. While switching to paper or reusable cotton bags for grocery purchases might be a viable option, plastics cannot easily be replaced in all sectors. Therefore, sustainable, biobased and biodegradable plastics alternatives are required. Perhaps the most common biopolymers for bioplastics production are poly(3-hydroxybutyrate) (PHB). PHB belongs to the group of polyhydroxyalkanoates (PHAs) and is a natural metabolite that was first identified in *Bacillus megaterium* [297, 298]. *Cupriavidus necator* is one of the most commonly used bacteria to produce PHB, and has been engineered for an improved production [299]. However, PHB production in bacteria is still costly, promoting an interest in PHB-producing yeasts and other microbial hosts [300]. Again, research in this area started with exploring *S. cerevisiae* as a potential production host, and the genes of the PHB pathway from *Ralstonia extorquens*, *Methylobacterium extorquens*, and from *C. necator* were expressed [300, 301]. Also other yeasts have been investigated as potential PHB cell factories [302]. While PHB production

has been shown to work in *S. cerevisiae*, its competitiveness is not up to par with that of fossil plastic production. However, the usage of oleaginous yeast confers an advantage as the precursor for PHB production, acetyl-CoA, is the same as for storage lipid accumulation. Thus, engineered *R. glutinis* has recently been investigated for PHB production and showed great potential due to high PHB concentrations [303].

Lipid production from microorganisms for use in the feed, food and other industries is also an important application and research field. Several companies have launched industrial production of microbial lipids [304]. Of particular interest has been the production of microbial oils that can mimic cocoa butter [304], which is an important oil for the food industry but also extensively used in cosmetics. The cocoa tree is cultivated across various tropical countries [305]. Pricing and availability of cocoa butter largely depends on the outcome of the harvest, which is not affected only by common weather fluctuations but also by the effects of climate change [306, 307]. Microbial lipid production circumvents these problems and makes it possible to obtain stable sources of vegetable and other food oils in general. While oleaginous yeasts seem like the obvious choice for microbial oil production, the model yeast *S. cerevisiae* has, of course, already been successfully engineered to produce oils equivalent to cocoa butter [308-310]. However, the number of genetic modifications needed to achieve this and to reach lipid yields as high as in oleaginous yeasts, makes the use of *S. cerevisiae* as a microbial oil production host a quite tedious enterprise [311]. Therefore, multiple oleaginous yeasts, including *Cutaneotrichosporon oleaginosum*, *R. toruloides* and *Yarrowia lipolytica*, have been investigated as potential microbial cocoa butter substitute sources [17, 312-314]. While similar lipid profiles could be achieved, microbial production is not yet sufficiently economically competitive for large-scale production.

Palm oil is another high-demand oil, accounting for about a third of all vegetable oils used [153]. Due to its low price, it has become increasingly popular, leading to substantial deforestation of tropical rainforest in Asia and South America [315]. This in turn has drastic effects on the worldwide climate, biodiversity, and food and water security, just to name a few [316, 317]. Some oleaginous yeasts, including *Lipomyces starkeyi*, *R. glutintis*, and *R. toruloides*, have been shown to naturally produce lipids with a composition similar to palm oil, making them a potential alternative that can even be produced from waste stream feedstocks [318-324]. The natural production of these similar oils is of advantage since engineering yeast strains to produce tailoured lipids is not only time-consuming, but can also affect the fitness of the engineered strain [17]. Although environmentally better than its vegetable oil competitor, the low price of palm oil makes the microbial alternative not yet attractive enough [325], necessitating more research in this area.

The important case of competitiveness is very different for more high-value products such as carotenoids, pharmaceuticals and enzymes. Carotenoids are natural colorants, lipid-soluble pigments, and vitamin precursors. They cannot naturally be synthesised by humans or any other animals and are therefore an important ingredient in foods and for their antioxidant functions [106, 326]. While carotenoids can be chemically synthesised,

microbial production, especially from renewable feedstocks, could increase availability and decrease costs. Some red yeasts are already used in commercial carotenoid production, along with algae and fungi species [327]. For example, *Monascus spp.*, *R. taiwanensis* and *P. rhodozyma* are used for the production of β-carotene, cryptoxanthin and astaxanthin, respectively [327]. Enzyme production from oleaginous yeast also has potential, as several strains have been identified as valuable resources of carbohydrate-active enzymes (CAZymes) for the breakdown of waste materials from forestry and agricultural residues [328]. Specifically yeasts of the Ustilaginomycetes class, part of the Basidiomycota phylum, have been identified as potential sources of CAZymes. They can break down xylan, as was also observed for *P. hubeiensis* BOT-O (Figure 3.3), which is an important trait to effectively digest major parts of the hemicellulose in plant walls. A bioinformatics approach has been used to predict CAZyme-encoding genes in oleaginous yeasts, as has previously been done for ascomycete yeasts [329].

While yeasts do play a major role in biopharmaceutical production, here *S. cerevisiae* is the clear favourite, despite common challenges such as posttranslational modifications, and specifically hypermannosylation of glycoproteins [330]. Nevertheless, yeast, and more specifically *S. cerevisiae*, is the most commonly used organism for production of insulin, human serum albumin and certain vaccines [331]. Also non-conventional yeasts are being considered as potential production hosts, or are already used as such for biopharmaceutical production in the case of *Ogataea polymorpha*, also known as *Hansenula polymorpha*, which is used for insulin and vaccine production [332], and *Pichia pastoris*, which even possesses some advantages over *S. cerevisiae* when it comes to hypermannosylation [333].

4.2 Bioproduct Precursors

Identification of the ability to produce important precursors for bioproduction can give a better understanding of the potential of certain organisms as cell factories and help focusing efforts on specific engineering endeavours, or develop new, so-called platform strains.

4.2.1 Acetyl-CoA – A pivotal metabolite for valuable products

As mentioned in previous chapters, many bioproducts share acetyl-CoA as a common precursor, making this metabolic intermediate a pivotal point for biosynthesis of high-value bioproducts, including, for example, isoprenoids (including carotenoids), PHB and 3-hydroxypropionate (Figure 4.1) [334-336]. Acetyl-CoA is, however, not only a common precursor; it can also be produced from a variety of different sources in the cell (Figure 4.1), giving many opportunities for its production and tweaking of said production.

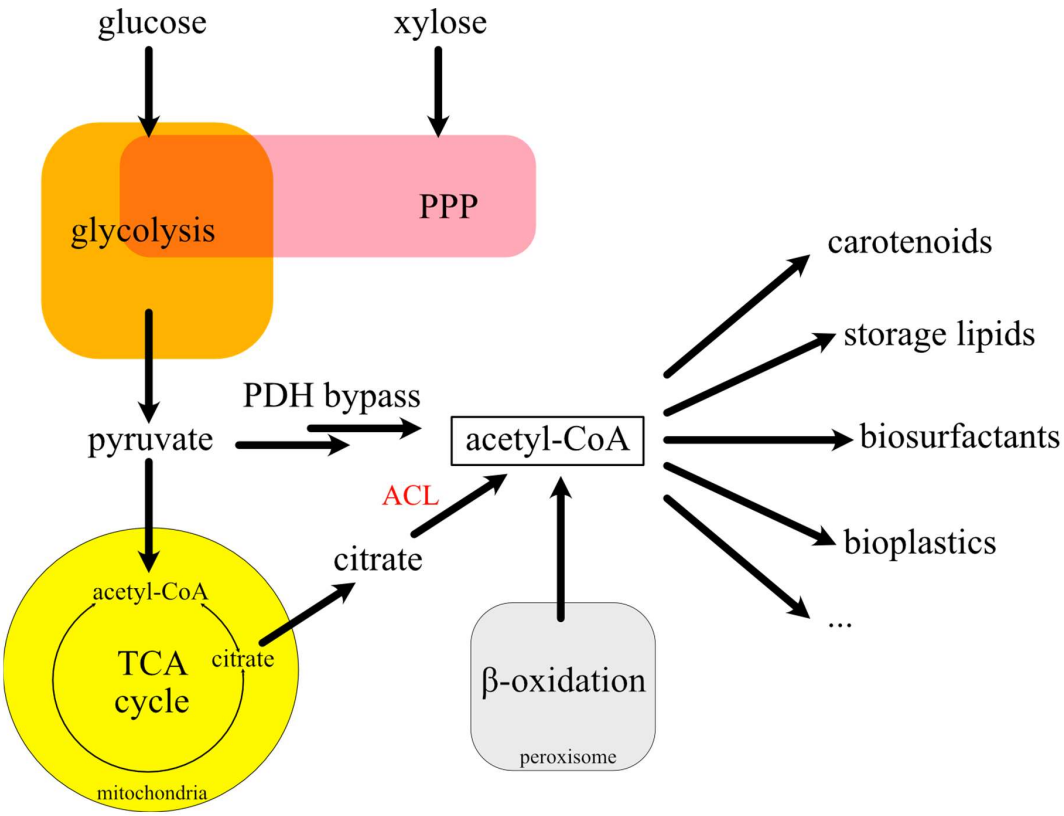

Figure 4.1. Simplified overview of the precursor acetyl-CoA with its origins within the oleaginous yeast cell, and potential bioproducts derived from it. The ATP-citrate lyase (ACL), which is naturally present in oleaginous yeasts, converts cytosolic citrate from the TCA cycle into cytosolic acetyl-CoA. The pyruvate dehydrogenase (PDH) bypass leads to cytosolic acetyl-CoA via acetaldehyde and acetate.

While many strategies such as overexpression of native enzymes and expression of enzymes from heterologous pathways have been employed to increase the cytosolic level of acetyl-CoA in *S. cerevisiae*, oleaginous yeasts do have a natural increased flux towards this intermediate metabolite during starvation conditions, when storage lipid production is increased [129]. This is due to the ATP-citrate lyase Acl1, which converts cytosolic citrate, mainly derived from the TCA cycle in the mitochondria, to cytosolic acetyl-CoA, and which has been shown to be present in the genomes and in the central metabolic pathways of both yeasts in this thesis (Figure 3, **Paper II** and Figure 3, **Paper III**). Engineering yeasts and filamentous fungi for an overexpression of this gene or introducing it to the organism of interest has been shown to significantly improve production of acetyl-CoA derived products through the increase of cytosolic acetyl-CoA [337-339]. Acetyl-CoA is present in the nucleus, the cytosol, the mitochondria and the peroxisome . In *S. cerevisiae*, there are no direct mechanisms for the transport of acetyl-CoA between the different compartments, and specifically towards the cytosol as for oleaginous yeasts with the ATP-citrate lyase [340]. Besides its role in the nucleus, acetyl-CoA is important for the central carbon metabolism in the cell. In *S. cerevisiae*, acetyl-CoA is produced in the mitochondria via the pyruvate dehydrogenase complex, and is then further oxidised via citrate in the TCA cycle [341], while in oleaginous yeasts citrate could be transported out of the mitochondria

into the cytosol, where the ATP citrate lyase then cleaves citrate for regeneration of acetyl-CoA and oxaloacetate [342]. In the cytosol, in the pyruvate dehydrogenase bypass, pyruvate is decarboxylised to acetaldehyde, which then leads to acetyl-coA via acetate. However, in *S. cerevisiae*, acetaldehyde can also be converted to ethanol by alcohol dehydrogenase, which is, due to the Crabtree-effect in *S. cerevisiae*, the main route during growth on glucose [343]. Acetyl-CoA is also the final product of β-oxidation, which occurs in the peroxisome.

4.3 Strategies for Flux Redirection

Since acetyl-CoA can be a precursor for so many different products, and some yeasts like the ones used in this work do naturally produce a multitude of such products, redirection of the acetyl-CoA flux towards the desired product may be a viable strategy to increase their production. *P. hubeiensis* BOT-O, for example, produces both storage lipids and MELs. The latter can be used as a biosurfactant. During nitrogen starvation, when storage lipid production was increased, genes in the MEL pathway were significantly upregulated. This could lead to a lower increase in storage lipids than would have been possible if the flux would have been fully redirected towards lipids. Similarly, *R. toruloides* BOT-A2 produces carotenoids and storage lipids, again both form acetyl-CoA. While carotenoid production seemed to be mostly downregulated during nitrogen starvation (Figure 3, **Paper III**), when cytosolic acetyl-CoA production and the subsequent storage lipid production is high, it has also been reported to increase at higher C/N ratios in *R. glutinis* [19], although higher C/N ratios may also enhance lipid production. Additionally, if both of these oleaginous yeasts should be considered as production hosts of other acetyl-CoA derived products than their native products, strategies for flux redirection from storage lipids towards the desired product will become important.

4.3.1 Genetic Engineering and Regulation of the Flux

Approaches for flux regulation in established cell factories such as *S. cerevisiae* or *Escherichia coli* usually involve some form of genetic engineering and modification of the metabolic pathways: overexpression of certain pathways, downregulation or deletion of competing pathways, or bypassing of specific pathway steps [344, 345]. A redirection of the metabolic flux also generally requires introduction of genes that are needed to produce the desired product, especially if it is not naturally produced by the production host. Additionally, downregulation or complete deletion of genes involved in the formation of other products that use the same precursor is needed. In order to achieve this, methods for genetic engineering must be available, including well-characterised promoters for up- or downregulation of target genes.

As previously mentioned, despite attempts, no transformation methods have been successfully applied to *P. hubeiensis* BOT-O. Therefore, no well-characterised promoters are available as testing their functionality generally requires the introduction of a construct consisting of the promoter of interest and a reporter gene. However, in **Paper III**, a genetic engineering method for *R. toruloides*, that was first applied by project collaborators in

[346], is described. In general, most engineering approaches for this species use a transformation method that is mediated by *Agrobacterium tumefaciens*, a bacterium found in soil and capable of infecting plants causing crown gall disease [347, 348]. It has, since its first discovery, been used to genetically transform not just plants, but also other eukaryotes such as yeasts, filamentous fungi and even human cells [349-351]. However, there are many drawbacks to this method of transformation, including the risk of potential contamination of other organisms in the lab itself [352]. Furthermore, transformation efficiency is highly dependent on a variety of factors, including the to be transformed species itself, the ratio of *Agrobacterium* and the recipient cells, as well as cultivation time and settings such as pH and temperature [353]. It also generally requires more time than electroporation or the lithium acetate (LiAc) methods for transformation of yeast, as it requires more transformation steps [354]. Therefore, electroporation was chosen to transform BOT-A2, and a new protocol had to be developed, tweaking, for example, the amount of DNA used, as well as adjusting buffers and changing the recovery medium, as previous electroporation methods described for other strains of the *R. toruloides* species did not lead to any colony formation. The development and optimisation of a transformation method is the first crucial step towards other metabolic engineering projects.

Another essential step is the availability of functioning and characterised promoters, that are either constitutive, inducible or repressible. As the focus of this work lies in acetyl-CoA derived products, and the flux towards cytosolic acetyl-CoA is considered to be high during nitrogen starvation in oleaginous yeasts [129], promoters regulated by the presence or absence of nitrogen would be of high interest. Surprisingly, while there are a few utilised constitutive promoters, of which many are derived from glycolytic enzymes [355-359], and even a few regulatable ones are available for *R. toruloides* [360], none of these are regulated by nitrogen starvation. Therefore, in order to facilitate regulation of production and redirection of the flux during nitrogen starvation, suitable promoters had to be identified and characterised. In the work presented in **Paper III**, promoter candidates were chosen based on RNA sequencing data from the exponential phase and the nitrogen starved phase, on either glucose or xylose. Candidates were chosen based on $\log_2$ fold changes and adjusted p-values on both sugars, as xylose is an important carbon source in renewable feedstocks. Both promoters for upregulation during nitrogen starvation with ammonium sulphate as the nitrogen source and for downregulation during the same condition were investigated. Promoters that are upregulated could be used to drive production of the acetyl-CoA derived product during nitrogen starvation, while the genes for downregulation could be used to redirect the flux away from native acetyl-CoA derived products like the carotenoids or storage lipids.

Multiple promoter candidates were investigated as oftentimes multiple promoters are needed for genetic engineering projects to be successful, as the regulation of multiple genes simultaneously is often required. Promoter candidates (listed in Table 2 in **Paper III**) were evaluated based the production of green fluorescent protein (GFP), measured via fluorescence, in response to nitrogen starvation. The GFP used here was codon-optimised for *R. toruloides* [355]. To induce expression during nitrogen starvation, a CN80 medium

was used, with YNB1x, a yeast nitrogen base (YNB) based medium with ample nitrogen (5 g/L of ammonium sulphate) as the standard medium. The media were used in reverse for the promoters repressed during nitrogen starvation. Random integration is a common, and still huge problem in the *R. toruloides* engineering community. It is thought to mainly be due to non-homologous end-joining, which, along with homologous recombination, is one of the two major DNA double-strand breaks repair pathways [361]. Due to the random integration, extensive screening of transformation candidate strains was necessary to account for the variety induced by the random integration, and find strains with desired GFP expression levels. For screening in **Paper III**, a BioLector microbioreactor was used to investigate promoters induced upon nitrogen depletion or starvation. This led to the insight that the random integration significantly affected the response of the different strains, when transformed with the same promoter, as significant differences between the fluorescence read-outs of the different strains were observed (Figure 4, **Paper III**). However, when considering the difference between the level of fluorescence measured when strains were grown in the inducing medium versus the standard medium, it was different enough for most strains with the same GFP-expressing promoter. The differences between the responses of the different transformants could be, for example, either due to differences in copy numbers, or integration sites, or both. In *R. toruloides*, product titres were found to correlate with copy numbers [362]. In *S. cerevisiae*, it has been shown that the chromosomal context around the integration site can have an effect on gene expression [363]. In order to test this, quantitative polymerase chain reaction (qPCR) can be used to determine copy numbers. A preliminary experiment was performed on a few strains with the same promoter (PGI1p, data not shown), revealing indeed differences between the copy numbers, which corresponded to fluorescence intensity for one of the tested promoters. However, results for another promoter (3877p, data not shown) did not follow the same trend, suggesting that copy number has an effect but cannot alone explain the differences in GFP expression. qPCR should be repeated with more replicates and across all strains, and potential differences in integration site should be tested. This was not possible to do within the time frame of this thesis work but could easily be investigated using whole-genome sequencing. Sequencing depth for this sequencing would not have to be as deep as for the *de novo* BOT-A2 sequence, and reads could be mapped to the BOT-A2 assembly, leading to a faster sequencing and analysis. Furthermore, the qPCR approach that was used does not inform about the completeness or the functionality of the GFP gene, thus incomplete expression cassettes could still be counted towards the copy number count although they may be non-functional. This occurrence of this could be investigated via sequencing as well.

After the initial screening, selected transformant strains were further characterised using flow cytometry. Here, due to the use of bigger culture volumes through the use of shake flasks, multiple other parameters, such as carbon assimilation and nitrogen depletion, could be measured. In total, three promoters for upregulation during nitrogen starvation were successfully identified (Figure 4.2a-c), including a *MEP2* promoter taken from *R. toruloides* NP11, a strain of the MAT-A1 mating type, whereas BOT-A2 has the MAT-A2 mating type. The functionality of the *MEP2* promoter suggests that promoters could be

used interchangeably throughout strains of this species. In order to further confirm this hypothesis, promoters presented in **Paper III** should also be tested on other *R. toruloides* strains.

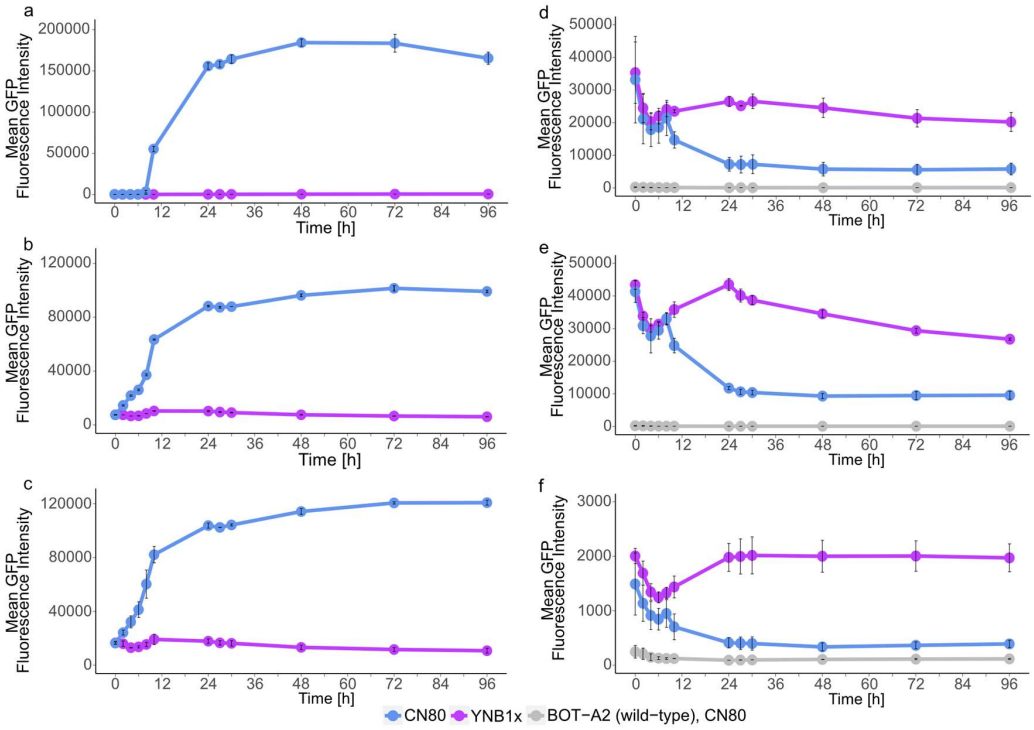

Figure 4.2. Fluorescence level profiles of transformant strains with the six novel promoters identified and characterised in **Paper III**. Upregulated promoter strains include the UP1a strain with a 3877p promoter (a) (from the gene RTBOTA2_003877 putatively encoding a putative nitrate transporter), the UP3e strain with a 0480p promoter (b) (from the gene RTBOTA2_000480 putatively encoding the Carboxypeptidase Y inhibitor Tfs1p), and the UP10a strain with a MEP2p promoter (c) from NP11. Downregulated promoter strains include the DN1a strain with a 0530p promoter (d) (from the gene RTBOTA2_000530 putatively encoding a ribosome-associated molecular chaperone Ssb2p), the DN2a strain with a 3356p promoter (e) (from the gene RTBOTA2_003356 putatively encoding a diphosphomevalonate decarboxylase Mvd1p), and the DN5c strain with a 5360p promoter (f) (from the gene RTBOTA2_005360 putatively encoding a mitochondrial 2-oxoglutarate dehydrogenase Kgd1p).

In addition to promoters upregulated during nitrogen starvation, three promoters for downregulation were identified (Figure 4.2d-f). These could be used for downregulation of pathways leading to undesired acetyl-CoA derived products. While the microbioreactor was found to be a good tool for the screening of different strains with the same upregulated promoters, the same was not seen for downregulated promotors. In general, downregulation is harder to detect using GFP as a decrease in fluorescence is not immediately detected due to the half-life of GFP. Half-life of GFP was not tested in *R. toruloides* specifically, but is known to be many hours in yeasts in general, depending on the type of GFP used [364]. While more short-lived GFPs have been developed, they meet limitations regarding a substantial loss of fluorescence [365]. Regardless, downregulation, and a significant difference in fluorescence in the strains grown in the induction and the repression medium, respectively, was not detectable in the BioLector cultivations. Therefore, a few selected strains were characterised using flow cytometry directly, which still led to satisfactory

results. Potentially, some of the more short-lived GFPs could be tested for *R. toruloides* despite their lower fluorescence intensity, or a different fluorescent protein, with a shorter half-life, could be used for characterisation of more repressed promoters. Regardless, the six promoters identified in **Paper III** are a valuable addition to be *R. toruloides* toolbox and are the first promoters regulated during nitrogen starvation or depletion.

4.3.2 Redirection Through Addition of Inhibiting Agents

As genetic engineering can sometimes be tedious or not even possible for some organisms, inhibiting agents could be used for the downregulation of certain metabolic products. This is a non-invasive approach, theoretically saving time on engineering steps such as method optimisation and screening. As storage lipid production is generally the most cytosolic acetyl-CoA consuming step, it could be valuable to downregulate it by chemical means to allow more flux towards other native and non-native products. This approach is, however, less targeted than genetic engineering.

One inhibiting agent that has been shown to affect storage lipid production is cerulenin. It is produced by *Cephalosporium caerulens*, an ascomycete fungus [366] and is an antifungal antibiotic, which was discovered in 1960 [367]. It has been shown to function as an inhibitor of *de novo* fatty acid and sterol biosynthesis from acetyl-CoA, as well as of lipid chain elongation in yeasts [368, 369]. Cerulenin specifically inhibits fatty acid synthase by binding to the active site of the ketoacyl domain of the fatty acid synthase α Fas2 [370, 371]. A quantitative correlation between cerulenin and concentrations of intracellular malonyl-CoA has been shown [372]. Although cerulenin could be added to the culture medium to reduce TAG formation and increase the production of other acetyl-CoA derived products, this is rarely done. Rather, cerulenin is used as a mutagenic agent and a selection marker. *S. cerevisiae*, for example, has been mutagenized and then spread on plates containing cerulenin to select for high ethyl caproate productivity, an important flavour component of sake with acetyl-CoA as its precursor [373]. Another *S. cerevisiae* strain has been generated using mutational screening upon cerulenin exposure during 7 days of growth [374]. This strain was shown to be producing high levels of free fatty acids, and particularly high levels of caprylic acid, which has applications in industry. The mutant strain was found to have a novel point mutation in the fatty acid synthase α gene *FAS2* [374]. Cerulenin has also been used in *R. toruloides* and other oleaginous yeasts such as *Lipomyces starkeyi* and *Yarrowia lipolytica* in similar applications as described for *S. cerevisiae*. The parent strains undergo mutagenesis and are then selected for tolerance towards cerulenin, mainly to find new strains with higher lipid production capabilities [375-378]. Cerulenin has also been supplemented to cell cultivations for resveratrol production in an engineered *R. toruloides* strain, in an attempt to increase the yield [379]. However, cerulenin addition led to a decrease in production compared to the non-supplemented cultivations even at very low cerulenin concentrations as low as 10 µM (=2.2 µg/mL) [379]. Growth inhibition was observed at 50 µM (=11µg/mL). Of course, the use of cerulenin as an agent to redirect flux could have unwanted side effects and lead to undesired selection of cerulenin resistant cells. In an attempt to use mutagenesis to achieve higher lipid producing *R. toruloides* strains performed by Yamada and colleagues [375], mutants with

a higher resistance to cerulenin also produced higher amounts of lipids. This is however counterproductive if the goal is downregulation of lipid production. Furthermore, cost is a factor for industry, and cerulenin can be an expensive additive.

Here, cerulenin was preliminarily tested as a supplement to the CN80 cultivation medium with glucose, that was used for lipid production in **Paper I**, **Paper II**, and **Paper III**. Cerulenin was added to downregulate storage lipid production, which was monitored using microscopy. Different concentrations of cerulenin were tested for both *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2, based on concentrations used for another *R. toruloides* strain [379]. The aim of this experiment was to find a cerulenin concentration that would inhibit storage lipid production, while not affecting growth significantly. As this was a tentative and explorative experiment, no replicates were taken. Measurements for OD_{600nm}, to monitor growth, and microscopy, to monitor lipid body formation, were performed. First, the addition of 1 µg/mL and 2.5 µg/mL to either BOT-O or BOT-A2 was tested. Cerulenin was added 15 h after the inoculation of the cultivations. While growth inhibition in BOT-O became obvious about 9 h after the addition of cerulenin, no such difference was detectable for BOT-A2 (Figure 4.3). In fact, growth inhibition was not observed at all over the course of the experiment. Differences between lipid body formation in the control and the inhibited cultivations were detectable for BOT-O at around 26 h after the cerulenin addition, while lipid body formation in BOT-A2 remained unchanged.

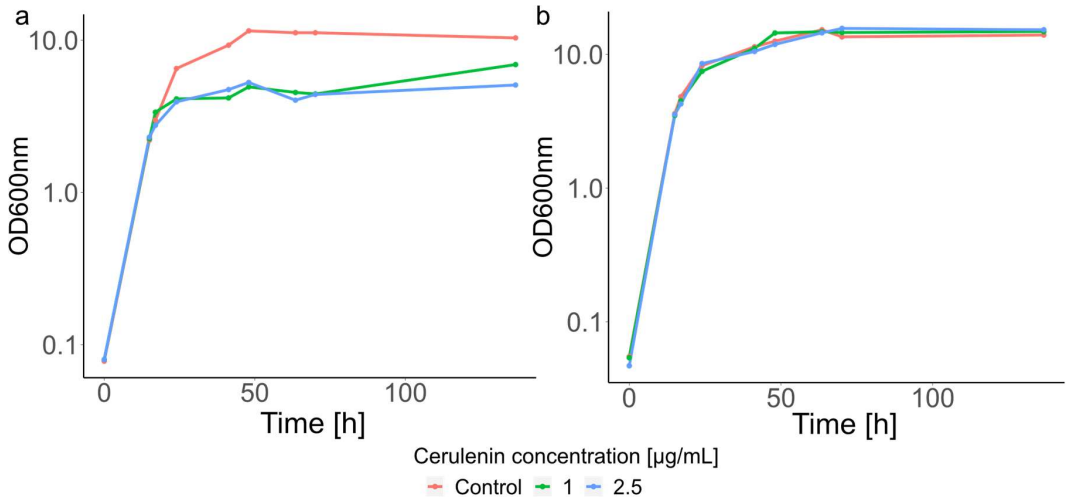

Figure 4.3. Cultivations of *P. hubeiensis* BOT-O (a) and *R. toruloides* BOT-A2 (b) in CN80 with glucose as the carbon source in shake flasks. Cerulenin was added at 15 h.

As no inhibition was detectable for BOT-A2, higher concentrations of cerulenin were tested. Concentrations were increased to up to 50 µg/mL, a concentration which has previously been described as lethal for a *R. toruloides* wild-type strain [380]. After addition of higher cerulenin concentrations at 17 h, growth seemed to be somewhat inhibited but was able to recover after only a few hours of cultivation. Similarly, no differences could be detected between the lipid body formation of the differently cultivated cells (Figure 4.4). These results were quite surprising, especially since cerulenin is known to inhibit the

ketoacyl group of Fas2. As previously discussed, BOT-O possesses two *FAS* genes containing most domains of *FAS1* and *FAS2*, and therefore also two ketoacyl domains (Figure 3.8). Therefore, this yeast was hypothesised to potentially have a higher tolerance towards cerulenin than BOT-A2, but this could not be confirmed by the experiment. BOT-A2 itself has a more conventional split between the different *FAS* subunits, which is consistent with that of other *R. toruloides* strains, and therefore was expected to behave similarly. Obviously, the results should be corroborated, and the experiment repeated to confirm the findings. The Fas2 protein of BOT-A2 was also subjected to a computational structural investigation by using models produced by Phyre2 [381] to investigate the cerulenin binding site, which was inconclusive. It is possible that a point or other mutation in *FAS2* caused cerulenin tolerance in BOT-A2, as was reported on in a cerulenin tolerant *S. cerevisiae* strain [374].

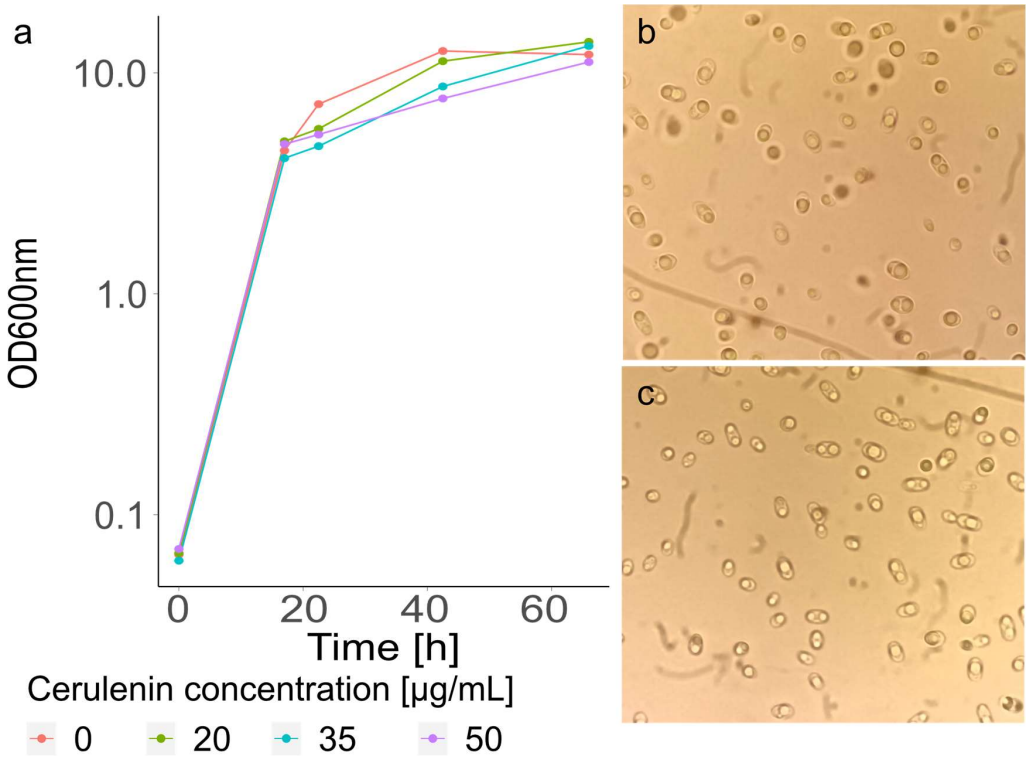

Figure 4.4. Cultivations of *R. toruloides* BOT-A2 in CN80 with glucose as the carbon source in shake flasks. Growth is displayed as OD_{600nm} (a). Cerulenin was added at 17 h. Cells were observed under a microscope at a 1000x magnification, with cells growing in medium without cerulenin in (b), and cells growing in 50 µg/mL in (c). Microscopy pictures were taken 74 h after the addition of cerulenin. Lipid bodies are clearly visible in both pictures.

Chapter 4: Summary

- There is a wide variety of interesting and valuable bioproducts that can help replace fossil use.
- Acetyl-CoA is a major precursor for native and non-native products in yeasts.
- Oleaginous yeasts and some filamentous fungi possess an ATP citrate lyase leading to higher flux towards cytosolic acetyl-CoA, making them interesting cell factory candidates.
- Oleaginous yeasts already efficiently convert cytosolic acetyl-CoA for lipid production, and potentially other products. If they are to become production hosts for other acetyl-CoA derived products, flux would need to be redirected from their native products to the desired product.
- Flux redirection could be achieved with genetic engineering, which requires established engineering methods and regulatable promoters, or by addition of inhibiting agents.
- Six novel promoters, regulated during nitrogen starvation, were identified for *R. toruloides* BOT-A2 in **Paper III**.
- The fatty acid synthase inhibitor cerulenin was successful at inhibiting lipid production in *P. hubeiensis* BOT-O, but not in *R. toruloides* BOT-A2.

Chapter 5: Renewable Feedstocks

The use of waste streams as feedstocks is gaining increasing interest in research and industry as these feedstocks are cheap, generally abundant, and renewable. Utilisation of these types of feedstocks is interesting for biorefineries, as well as for sustainable bioproduction, as previously mentioned. Bioproduct production from microorganisms is attractive alternative to conventional production. Microbial bioproduction generally requires less space than conventional production, and is overall more sustainable through the use of renewable waste materials [382].

5.1 The World of Waste Products

There are a wide variety of different waste streams that can be used as feedstocks for microbial bioproduction. Glycerol from biodiesel plants can be used as a carbon source by a variety of microorganisms [383, 384], including BOT-O and BOT-A2 as tested in **Paper I**. Other waste streams include volatile fatty acids from anaerobic waste digestion [385, 386], lipid-rich waste streams from vegetable oil productions [387, 388], or lignocellulose from agricultural, forestry or municipal solid waste [152, 389, 390]. Lignocellulose is perhaps one of the most promising sources of renewable feedstocks, which has been and still is investigated intensely. Hydrolysates from lignocellulose waste have been considered for example as a source of feedstock for yeast-based production of bioethanol [391], bioplastics [392], or lipids [393-395], also from *R. toruloides*.

As previously mentioned, lignocellulose primarily consists of cellulose, hemicellulose and lignin [25], with lignocellulosic composition varying depending on the plant itself, as well as growth conditions and growth stages [396]. Cellulose is a linear and homogenous polysaccharide consisting of repeated glucose units that are linked via β-1,4-glycosidic bonds [397]. Cellulose chains interact with each other via hydrogen and van der Waals bonds, which results in long, flat and parallel structures that form tightly packed microfibrils. Hemicellulose binds to cellulose, and other plant cell wall components, in order to provide strength and rigidity [398-401]. It is more complex and is composed of a variety of hexose and pentose sugars, including glucose, mannose, arabinose and xylose [397]. Composition differs depending on the plant. Hardwoods, as well as cereal crops and grasses contain more xylans, whereas softwoods have a higher glucomannan content [396]. In general, hemicellulose is easier to hydrolyse as it is not crystalline but rather amorphous in structure. The third major component of the plant cell wall, lignin, is a very complex polymer, with a high structural and compositional diversity [402]. Despite its variability, it mainly consists of 4-hydroxycinnamyl alcohols, coniferyl alcohol and sinapyl alcohol [403]. Lignin is also highly difficult to break down and poses challenges during pretreatment and hydrolysis of lignocellulosic biomass [404]. In fact, it has been suggested that it took about 60 million years for microorganisms to develop the ability to break down

lignin after the appearance of trees [45], although this is debated [405]. However, it was ancestors of today's present basidiomycete fungi that evolved to break down lignocellulose about 295 million years ago [45], possibly explaining the xylan-utilising ability of BOT-O.

Lignocellulosic biomass needs to be pretreated and hydrolysed in order to be utilisable by most microorganisms. During pretreatment, not just fermentable sugars are released, however, but also inhibitors that can significantly affect growth and product yields [26]. At a certain concentration, these can lead to the biomass not being fermentable, thus pretreatment has to be carefully balanced. Presently, acid-catalysed steam pretreatment is widely used, leading to a better recovery of hemicellulose-derived sugars [406-408]. The subsequent hydrolysis step is usually performed using a specific enzyme cocktail for enzymatic hydrolysis, or acids such as HCl and H_2SO_4 are used for dilute-acid hydrolysis [27]. Dilute-acid hydrolysis is fast but leads to the formation of a significant number of by-products. Resulting monosaccharide concentrations are also generally higher when enzymatic hydrolysis is used, as the use of carefully selected enzymes does not lead to sugar degradation reactions. Compositions of lignocellulosic biomass differ, as previously mentioned, and enzyme cocktails can be targeted towards the specific biomass used.

As mentioned above, pretreatment of lignocellulosic biomass leads to a release of sugars, but also the presence of inhibitors. Certain inhibitors are found in derivatives of certain structures (Figure 5.1). The breakdown of hemicellulose leads to the release of a variety of hexose and pentose sugars, but also to the release of inhibitors such acetic acid and furanic aldehydes such as furfural HMF, as well as formic acid, which is a byproduct of HMF and furfural breakdown [27, 409, 410]. Breaking down cellulose leads to similar results in terms of inhibitors, but sugar-wise only glucose is released. Lignin is hard to break down and generally leads to an increase in phenolic compounds in the hydrolysate. These, in contrast to the other inhibitors can also inhibit the hydrolysis step [27, 411]. Pretreatment of lignocellulosic biomass is constantly improving, and strategies to increase tolerance towards these inhibitors exist. These include introducing a short-term adaptation step to the cultivation [412-414] or the addition of nutrients such as vitamins or metals to the fermentation medium [415-421].

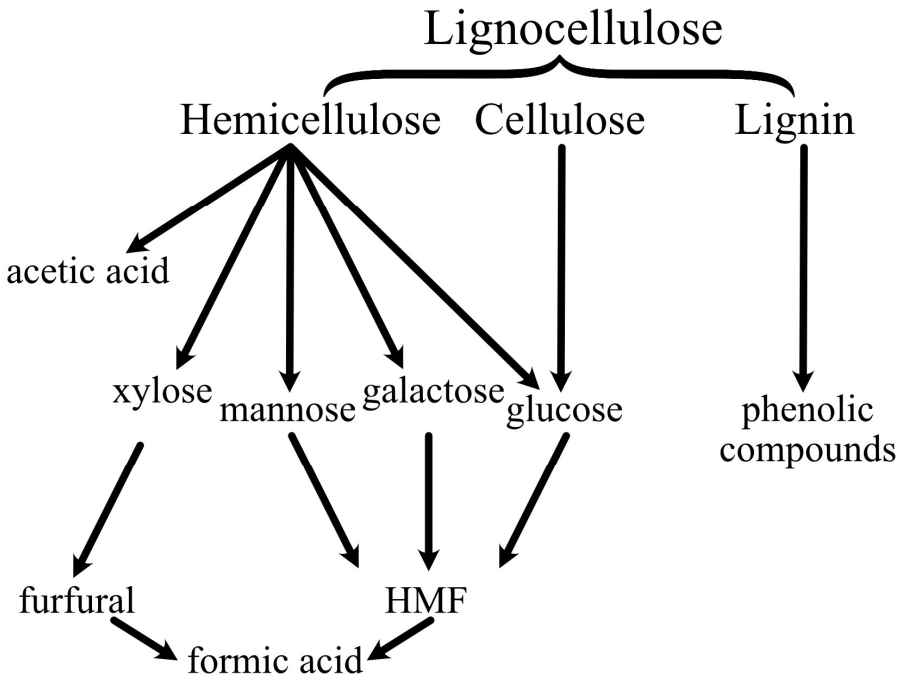

Figure 5.1. Lignocellulosic biomass and its breakdown during pretreatment. Monosaccharides and inhibitors are released. Adapted from [27].

5.2 Inhibitors and Inhibitor Tolerance

Lignocellulosic inhibitors are toxic to most microorganisms and inhibit growth and productivity in various ways (Figure 5.2). Weak acids, such as acetic acid and formic acid, have been shown to inhibit growth through uncoupling of oxidative phosphorylation. When the external pH is below that of the pKa value of the acid, the undissociated form of the acid will diffuse across the plasma membrane into the cytosol of the cell. There, it dissociates again due to the higher intracellular pH [422, 423], which leads to a decrease in cytosolic pH. This causes accumulation of anions in the cell since protons are pumped out to restore the regular intracellular pH, also leading to a depletion of ATP [424]. A negative effect on growth in terms of cell viability and formation of biomass can be observed [424], although specifically lower concentrations of weak acids have been shown to have a beneficial effect on productivity in *S. cerevisiae* [425]. Some yeasts, such as *R. toruloides*, have been described as being quite tolerant towards weak acids. *R. toruloides* has been reported to tolerate up to 80 mM of weak acids with minimal effects on growth [426, 427]. *S. cerevisiae* has been shown to be affected and reach lower biomass levels at lower concentrations of acetic acid [428], however tolerance is strongly strain related [429]. Mechanisms of inhibition differ for furanic aldehydes. Furfural in general has been associated with inhibition of enzymes, and specifically key glycolytic enzymes such as hexokinase, triosephosphate dehydrogenase, aldehyde dehydrogenase, pyruvate dehydrogenase and alcohol dehydrogenase. This in turn leads to a negative effect on growth

and other cellular functions [430-432]. Furthermore, the reactive aldehyde groups of furfural and also HMF cause DNA mutations, protein misfolding and membrane damage through an increase in ROS [433, 434]. Repair of the damage leads to an increase in spent ATP, NADH and NADPH, leading in turn to longer lag phases and growth inhibition [434]. In *Y. lipolytica*, growth rates and lag phases have also been reported to be affected, and the biomass yield was reduced due to the presence of furfural [435]. Both furanic aldehydes can be reduced to less toxic compounds such as furfuryl alcohol, furoic acid and 5-hydroxymethyl furfuryl alcohol [436-438], which has been suggested to be facilitated by alcohol dehydrogenases [433]. However, HMF has been reported to be reduced at a slower rate than furfural [439].

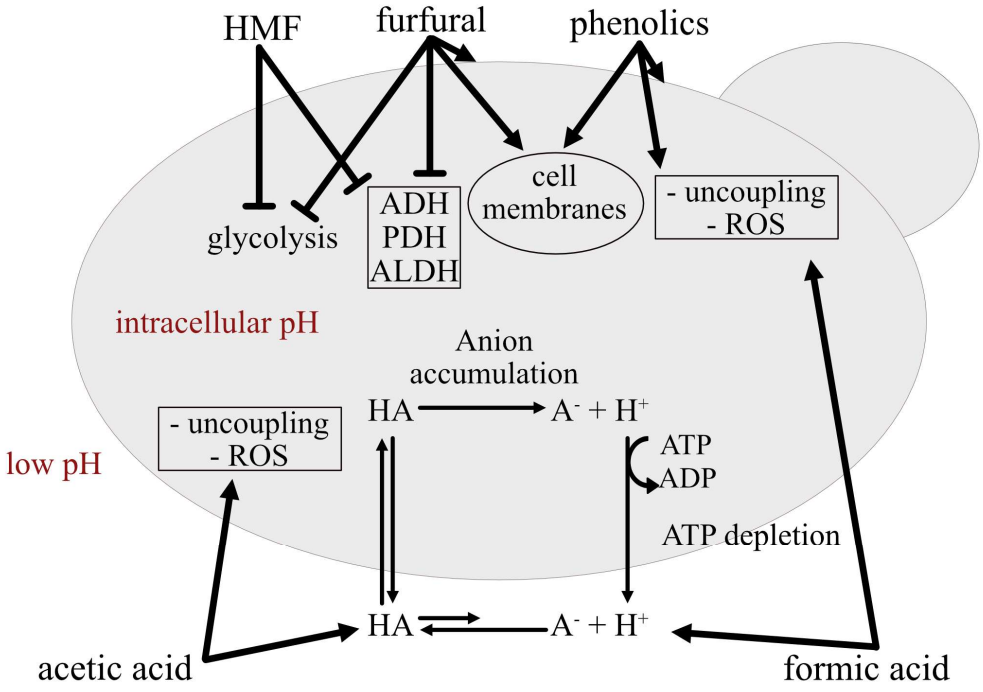

Figure 5.2. Overview over the mechanisms of inhibition of lignocellulosic inhibitors and their effect on *S. cerevisiae*. HMF inhibits glycolytic enzymes, as well as alcohol dehydrogenase (ADH), pyruvate dehydrogenase (PDH) and aldehyde dehydrogenase (ALDH). Furfural, in addition to inhibiting enzymes similarly to HMF, damages cell membranes. Phenolic compounds damage cell membranes as well, and inhibit through uncoupling and their reactive oxygen species. The two latter inhibition mechanisms can also be found for formic acid, in addition to anion accumulation and ATP depletion. These two are also crucial inhibition mechanisms for acetic acid. Adapted from [440].

In **Paper IV**, *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2 were tested for their tolerance towards the weak acids acetic acid and formic acid, as well as the furanic aldehydes furfural and HMF. Furthermore, growth on spruce wood hydrolysate was investigated, as well as the effect of the addition of vitamins. Not much prior information is available on lignocellulosic hydrolysate inhibitor tolerance in *P. hubeiensis*, but it was possible to see that *P. hubeiensis* BOT-O was able to tolerate higher furanic aldehyde concentrations than *R. toruloides* BOT-A2 (Figure 5.3).

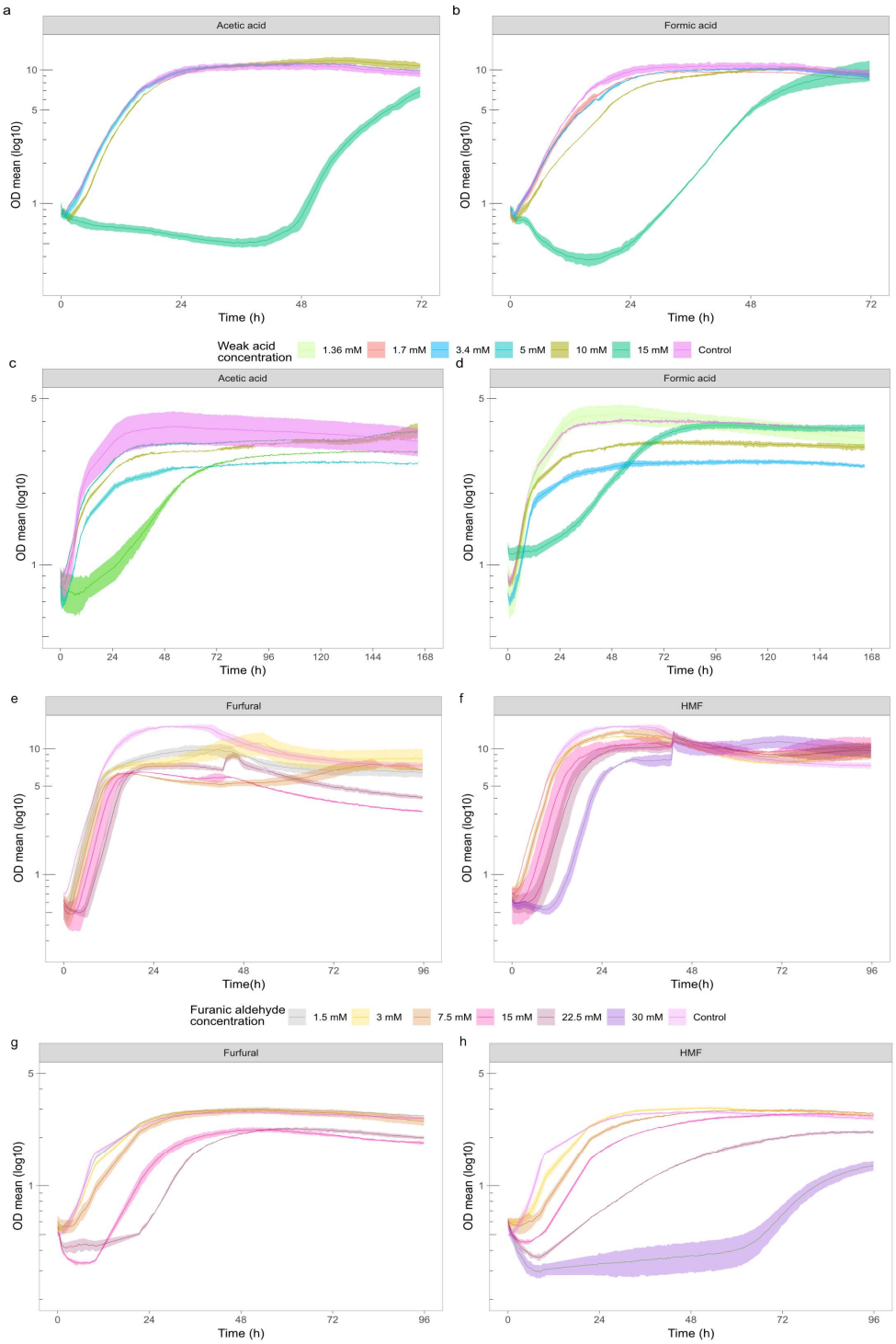

Figure 5.3. Inhibitor tolerance of *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2 as assessed in **Paper** IV. Multi-well plate cultivations were performed in media with either acetic acid, formic acid, furfural or HMF. Assessment of growth for BOT-O is shown in acetic acid in a, formic acid in b, furfural in e and HMF in f. Growth of BOT-A2 is shown in acetic acid in c, formic acid in d furfural in g and HMF in h. The main medium composition contained 25 g/L xylose, 25 g/L glucose, 5 g/L ammonium sulphate, 0.5 g/L magnesium sulphate, and 0.1 M potassium buffer. Working volume was 250 µL. Results are averages of triplicates with shadows depicting the standard deviation. Only selected concentrations are shown.

BOT-O did grow in all tested furfural conditions (up to 22.5 mM), and all tested HMF conditions (up to 30 mM), although with a longer lag phase at 30 mM. Growth of BOT-A2 however was already affected at 15 mM furfural or HMF, with significantly longer lag phases. BOT-A2 in turn was somewhat more tolerant towards the tested weak acids (Figure 5.3), but seemed to be less tolerant towards these inhibitors than what had been reported for other *R. toruloides* strains, showing negative effects on growth already at 20 mM acetic acid and 15 mM formic acid. This might be due to the use of both glucose and xylose in the tested medium, and BOT-A2 has been shown to grow slower on xylose than on glucose in **Paper III** despite the natural ability to metabolise both sugars. BOT-O in turn does grow equally well on either sugar or on a combination of them, but since no tolerance data from other strains of this species are available it is impossible to say if strains of *P. hubeiensis* are less tolerant towards weak acids in general or if the effect is strain- or condition-dependent.

Furthermore, when both yeasts were tested on spruce wood hydrolysate, BOT-O only grew on lower hydrolysate concentrations of up to 60% spruce hydrolysate, while BOT-A2 was able to tolerate higher concentrations of up to 70%. Short-term adaptations in form of propagations had a beneficial effect for both yeasts (Figure 6, **Paper IV**) and increased tolerance towards higher concentrations of hydrolysate drastically. Addition of vitamins only led to minor improvements and only increased tolerance in the presence of higher percentages of inhibitors at 20% for BOT-O and at 30% for BOT-A2. Overall, tolerance towards lignocellulosic inhibitors needs to be improved for both strains if lignocellulosic hydrolysates are to be used as feedstocks for the production of acetyl-CoA derived products in either *P. hubeiensis* BOT-O or *R. toruloides* BOT-A2. The work of identifying tolerances and upper limits is not made easier by the fact that the cells behave very different depending on the test medium. Cells growing in synthetic hydrolysate, a medium containing precise amounts of added inhibitors, carbon sources, etc., are more inhibited in terms of growth than cells growing in natural lignocellulosic hydrolysate. The presence of synergistic and antagonistic effects has been reported [441, 442]. Lignocellulosic hydrolysates also potentially contain compounds that have not been identified yet, further increasing differences between the synthetic and natural medium. It is also no secret that working with hydrolysates is difficult as composition changes depend on the origin of the biomass, the chosen pretreatment method, the age of the hydrolysate and the storage conditions. It is thus somewhat of an unreliable material to work with, but the necessity of using a sustainable and renewable feedstock warrants more research. Due to low product yields when lignocellulosic hydrolysates are used, investigation into new species and strains could help uncover unknown yeasts capable of utilising certain carbon sources or of tolerating higher amounts of inhibitors.

Chapter 5: Summary

- Lignocellulosic biomass from plant waste is a renewable and sustainable source of carbon.
- Through pretreatment of the biomass, fermentable sugars are released. However, so are inhibitors that affect growth and productivity of cell factories.
- BOT-A2 was more tolerant towards weak acids and lignocellulosic hydrolysate than BOT-O.
- BOT-O was more tolerant towards furanic aldehydes than BOT-A2.
- Further improvement of lignocellulosic hydrolysate utilisation is necessary to make these two strains competitive production hosts on lignocellulosic hydrolysates.

Chapter 6: Conclusions

The work presented in this thesis aimed to identify and characterise two novel oleaginous yeast strains that could be considered as cell factories for the production of acetyl-CoA derived bioproducts. Both yeasts are assumed to naturally have a high flux towards cytosolic acetyl-CoA during stress conditions such as nitrogen starvation, which is necessary for the production of a whole range of high-value products. In general, focusing on oleaginous yeasts for acetyl-CoA derived products is a viable strategy to identify natural high acetyl-CoA producers, making genetic engineering at least for this aspect unnecessary.

Isolation of new yeasts can be time consuming but can be rewarding as new species and strains with highly interesting traits could be uncovered. *P. hubeiensis* BOT-O, for example, is capable of growing equally well on glucose and on xylose, a trait that is not often seen in yeasts. In fact, this trait is so desired that a variety of yeasts have been genetically engineered to be capable of doing this, and yet they do not utilise the different sugars at rates as similar as those of BOT-O. In general, more easily available sequencing methods for both genome and RNA sequencing are a great help in the characterisation of new yeast strains, although difficulties remain when one does not work with the most conventional species such as *S. cerevisiae* as many tools and resources are targeted towards these species. Genome and RNA sequencing were able to reveal the central metabolic pathways in both *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2. BOT-A2, naturally capable of utilising xylose, however with significantly lower growth and sugar consumption rates than glucose, showed greater differences in transcription profiles when grown on the different sugars compared to BOT-O. In general, the sugar-independent traits of BOT-O make it an interesting specimen to study further to understand the underlying mechanisms of its efficient xylose metabolism, and their putative applications in bioproduction. While BOT-A2 did not excel in xylose utilisation, it had the advantage that it could be genetically engineered, and six novel promoters, differentially active during nitrogen starvation, were identified. These can potentially be used for any strain of the *R. toruloides* species, providing a valuable addition to the *R. toruloides* engineering toolbox.

For a more sustainable bioproduction, renewable materials such as lignocellulosic hydrolysates are important as feedstocks. In order to evaluate the tolerance of the two novel strains characterised in this work, both strains were grown on spruce wood hydrolysate, as well as single and a combination of lignocellulosic inhibitors. While BOT-O was naturally capable of using different xylans, it struggled more on hydrolysate than BOT-A2. Tolerance towards only furanic inhibitors was improved compared to BOT-A2 however, while weak acid tolerance was again better in BOT-A2 than in BOT-O, suggesting differences in inhibitor tolerance and tolerance mechanisms between the two investigated basidiomycetes.

Overall, characterisation of both *P. hubeiensis* BOT-O and *R. toruloides* BOT-A2 was performed using different approaches including physiology, genomics and transcriptomics, and morphology. Although genetic information is available, engineering of non-conventional yeasts is not straight forward and requires intensive method and tool development. Furthermore, non-homologous end-joining leads to random integration of genetic constructs in most non-conventional yeasts, including BOT-A2, and thus complicates genetic engineering even further. However, due to their various advantages and natural products such as storage lipids, biosurfactants and carotenoids, investigation of novel species, and here specifically BOT-O and BOT-A2, is worth the effort.

Chapter 7: Future Perspectives

Despite the detailed characterisation that was performed in this work in **Paper I**, **Paper II**, and **Paper III**, there is more work to do with respect to the development of these strains towards utilisation. *P. hubeiensis* BOT-O is, for example, still not engineerable, and will require more research to establish a functioning transformation method. However, future approaches for the development of such a method are clear, and lean on the similarities between BOT-O and its close relative *U. maydis*, for which engineering strategies exist [166], indicating that future engineering of BOT-O will be possible. Furthermore, hypothesis-driven genetic engineering could be a great aid in determining exactly how and why BOT-O is such an efficient xylose utiliser, as was revealed by extensive physiological characterisation in **Paper II** and is considered as one of BOT-O's greatest assets. Proteomics could also be used to improve knowledge and insight into this matter. Furthermore, a transcriptomic study investigating differential gene expression in a medium with mixed sugars containing glucose and xylose might be able to reveal more details.

For BOT-A2, while engineering is possible, currently only non-targeted integration has been achieved. This requires extensive screening of transformed strain candidates. Development of a targeted integration method would be beneficial, as well as determining potential beneficial integration sites. In *R. toruloides*, targeted integration has been achieved, however with low homologous recombination efficiency [443]. The used strategies could potentially be applied to BOT-A2, but also do require improvements for a higher efficiency. Only after targeted integration has been achieved, identification of potential integration sites is reasonable. In *S. cerevisiae*, with targeted integration, beneficial integration sites can for example be identified by integration of GFP as a reporter gene into different sites and screening for high fluorescence levels, for example [444]. For BOT-A2, the same could be done if targeted integration has been achieved. However, already genetically engineered strains with high fluorescence levels from **Paper III** for example could be sequenced to potentially determine beneficial integration sites. CRISPR methods for *R. toruloides* have been described [445-447] and could be tested in BOT-A2. Furthermore, the six promoters presented in **Paper III** should be tested further in other strains of the *R. toruloides* species to determine their generality and usefulness to extend the genetic toolbox for this species. Another interesting tool would be a genome-scale model for this strain, which could help with identifying new strategies for genetic and process engineering in this strain. Genome-scale models do exist for *R. toruloides* NP11 [448] and *R. toruloides* IFO0880 [449]. A model for BOT-A2 could be built based on these, while also using the BOT-A2 *de novo* genome sequence, as well as the obtained transcriptomic data. Similarly, a genome-scale model for BOT-O could be built using its *de novo* genome sequence and transcriptomic data, and based on models of other basidiomycete yeasts such as *U. maydis* [450] or *R. toruloides*. In general, genome-scale

models have been getting increasing attention since their first appearance [451], and will continue to do so in this more and more digital world.

Over the past few decades, vast improvements have been made when it comes to sequencing methods, and will likely continue to be made. Especially the rise of TGS with its long reads has somewhat revolutionised the field, and has been a great aid in achieving more complete genome assemblies including hard to capture repeat sequences [232], and more easily achievable *de novo* sequences, for example, which also in turn leads to better data that can be used in all kinds of projects. Today, long-read sequencing approaches can already be used to capture perfect, that is complete and error-free, bacterial genomes [452]. This leads to the assumption that, with improvements, the same might be possible for organisms with larger genomes such as yeasts in the future. Furthermore, improvements made also will affect and further improve RNA sequencing, leading to better and more complete transcriptomic data free from PCR-introduced bias.

When it comes to inhibitor tolerance and growth on lignocellulosic hydrolysates, different hydrolysates should be tested as compositions differ vastly depending on the raw material. Similarly, production capacities for some of the native bioproducts that BOT-A2 or BOT-O produce, such as storage lipids, carotenoids or MELs, should be investigated during inhibited growth. RNA sequencing during growth on lignocellulosic inhibitors could potentially reveal more about the variations in transcriptomic responses towards different groups of inhibitors in BOT-O and BOT-A2, as were observed in **Paper IV**. However, lignocellulosic hydrolysates are not the only available, renewable feedstocks of interest. Other waste streams that can be used as feedstocks exist, and are also gaining attention, among them food waste, and industrial waste streams from pulp and dairy industries [453-455]. Use of waste streams for bioproduction is an important factor for more sustainability. However, properties of waste streams are very dependent on their origins, and efficient use will require the presence of microorganisms with diverse profiles. Therefore, further investigation into new species and strains is important in order to find the optimal microorganisms for the specific tasks at hand.

While *S. cerevisiae* is still the primary production host for many processes, non-conventional yeasts are gaining importance. It is only reasonable that, with a change in desired feedstocks and products, production hosts change too. The new developments in biotechnology will certainly be of help for establishing non-conventional yeasts. New sequencing techniques are making the achievement of high-quality *de novo* sequences of new species and strains more attainable, laying a good basis for further work with those non-conventional yeasts. More high-throughput methods and more possibilities for screening are also leading the way for easier investigation of non-conventional yeasts, and more and more available engineering methods pave a road towards their implementation, even in harder to engineer strains.

Acknowledgements

Wow, writing this part now makes the end of this PhD feel a bit more real. It has been an interesting, challenging, rewarding, confusing, eye-opening experience, and overall I am very grateful I got to go through this somewhat exciting period in my career. I feel stronger, more determined and more self-assured after these four years of doctoral studies. I am thankful for the opportunities I've been given to learn and experience new things, the help and support I've received, and the fun moments I got to share with friends and colleagues. It has certainly been a wild ride across multiple divisions in the Department of LIFE, but I would not want to miss the experiences I've gained along the way.

First of all, I'd like to thank my supervisors Verena, Thomas, Joakim and Ed. Verena, thank you for jumping in when needed, for your always super valuable and to-the-point comments, your fast replies and constant willingness to help. Writing manuscripts and this thesis and making posters was much easier knowing that I could rely on your feedback. Thomas, I am grateful for the chance you gave me to work on this exciting project, which has allowed me to apply and really expand my knowledge and broaden my horizon. Joakim, thank you for your insightful comments and your valuable addition to our group meetings. Ed, thank you for your help with some of the bioinformatics and the introduction the genome scale models.

I also want to extend a thank you to Elin, first as a director of studies and later as my examiner, for your support throughout my entire PhD, and for taking the time to acquaint yourself with my research topic.

I am also grateful I got the opportunity to supervise a master student during my PhD studies. Davide, I could not have asked for a better master student. Thank you for being so easy to work with, a fast learner and great at working independently too. I know you'll do great, no matter what you decide to do in the future.

Furthermore, I'd like to thank our collaborators in Lund, Gunnar and Marie, as well as the whole TMB division whom I had the pleasure to visit for a few weeks. The atmosphere and positivity in the group was truly engaging.

A very, very, very big thank you has to go to Daniel, of course, you definitely deserve your very own paragraph in this acknowledgement section. I think I have told this story to quite a few people, but when I started my PhD, I was a little panicked thinking about the paper requirements (like probably a lot of fellow PhD students), and immediately after the first time we met (digitally) I messaged you. And I feel like since then I did not worry as much. Our collaboration throughout these past four years has been a pleasure, and while we had

fun, we also were extremely productive, managed to learn from each other and produce (at least in my opinion) amazing results. Thank you for encouraging me, my thoughts and ideas, and thank you for the great discussions too. My PhD definitely would have not been the same without you. You are going to be a great supervisor for your future students.

I started my work at Chalmers already during my master thesis work, and thus managed to experience three divisions in total: IndBio, FNS, and SysBio. I want to extend a general thank you to all my colleagues across all the divisions, past and present, whom I had the pleasure to work and socialise with. And a big extra thank you for all the research engineers I got to work with throughout the years who make sure that everything runs as smoothly as possible. I want to extend some thank you also to the administrative staff that help make sure the divisions and the department keep on going too. I also want to express my gratitude towards Calle, Lisbeth and Marlous, without whom I would have never come to Sweden. Doing my master thesis work in IndBio, surrounded by a fun environment including Dan, Jeroen, Jae, Elena, Vero, Andrea, Cathi, Jocke, Fabio, and of course also Ibai and Mauri, my fellow master students at the time, really made me consider doing a PhD, and staying in Sweden and at Chalmers to pursue this degree. I also do not want to forget some of the "newer" IndBio people (you really did not start that much later after I left): Amanda, escaping with you and Valle, or tipping kayaks and getting rescued by you is always fun.

After starting my PhD in FNS, my new colleagues were super welcoming and supportive, and I want to extend another thank you to all of you. I will definitely miss the Friday fikas. I do not want to mention too many names, but I definitely want to thank you, Agata, for the fun hours we spent teaching and chatting. Olle and Vanessa, boardgame nights with you are fun. And I of course also have to mention João, part of the small Friday gym gang (although you really did not show up to a lot of the Fridays recently), for the fun (and serious) conversations, heavy lifts and the overall fun times we spent together. If I can leave you with one piece of advice though, stop dressing like a 50-year-old.

Mauri, you are definitely one of the oldest friends I have here in Sweden (pun intended). Thank you for being you, and thanks for showing up to steak night with me and Dan when Jeroen does not anymore. Also, those t-shirts are still amazing, and I love being part of a group that has special t-shirts, thank you.

Ibai, it is time to thank you too. I still remember when you walked into that master student office we had on the second floor and how I at first thought you talked too much. Well, you still do sometimes, but I'm also guilty of that and I still enjoy talking with you. We have come so far and I can't wait for you to finish up soon too. Thanks for also being my gym partner and a member of our Friday gym gang. Thanks for helping me to push myself more and for being there so we can look stupid and suffer together.

Now it is time to thank my family, and my friends at home in Berlin and across Germany. Thank you for dealing with my slow and sometimes non-existent replies, and for accepting that I am more of an in-person person. Thank you for visiting me here, or for letting me

hang out with you whenever I manage to come over. A special thanks goes to Mini and Theresa who have been on this science journey with me for more than 10 years ago. And of course extra special thanks to my parents, to Mama and Papa, for raising me the way you did, and for supporting my decisions of going abroad. Thanks to my siblings Juschi and Basti, for keeping everything running smoothly while I'm not there, and thank you to my grannies, my aunts and uncles (and especially Tante Wilma for coming to visit me basically every year), and the rest of my family. I love you all and I do miss you.

Lastly, I want to thank Dan. I got to watch you succeed and finish your PhD with flying colours (although you might not think that, but I am proud of you for what you have achieved), and now it is my turn. I cannot thank you enough for all the support you have given me throughout the years. You encouraged me to stand up for myself, you listened to my random science thoughts, helped proof-read manuscripts, reports and the like, were there when urgent help was required, in- and outside the lab, and always manage to help me calm me down when necessary (although, to be honest, I am not nearly half as dramatic as you are). Without you, these whole past four years would have been immeasurably harder and less fun, and I am extremely grateful to have had you by my side throughout this journey. Here is to many more fun adventures together. I love you.

References

1. Cavalieri, D., et al., *Evidence for S. cerevisiae fermentation in ancient wine.* J Mol Evol, 2003. **57 Suppl 1**: p. S226-32.
2. McGovern, P.E., et al., *The beginnings pf winemaking and vinivulture in the ancient Near East and Egypt.* Expedition Magazine, 1997. **39**: p. 3-21.
3. McGovern, P.E., et al., *Fermented beverages of pre- and proto-historic China.* Proc Natl Acad Sci U S A, 2004. **101**(51): p. 17593-8.
4. Gest, H., *The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society.* Notes Rec R Soc Lond, 2004. **58**(2): p. 187-201.
5. Pasteur, L., *Mémoire sur la fermentation alcoolique*. 1857: BoD–Books on Demand.
6. Barnett, J.A., *Beginnings of microbiology and biochemistry: the contribution of yeast research.* Microbiology, 2003. **149**(3): p. 557-567.
7. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
8. Kaunitz, J.D., *The Discovery of PCR: ProCuRement of Divine Power.* Dig Dis Sci, 2015. **60**(8): p. 2230-1.
9. Legras, J.-L., et al., *Bread, beer and wine: Saccharomyces cerevisiae diversity reflects human history.* Molecular Ecology, 2007. **16**(10): p. 2091-2102.
10. Li, M. and I. Borodina, *Application of synthetic biology for production of chemicals in yeast Saccharomyces cerevisiae.* FEMS Yeast Res, 2015. **15**(1): p. 1-12.
11. Bellissimi, E., et al., *Effects of acetic acid on the kinetics of xylose fermentation by an engineered, xylose-isomerase-based Saccharomyces cerevisiae strain.* FEMS yeast research, 2009. **9**(3): p. 358-364.
12. Demeke, M.M., et al., *Combining inhibitor tolerance and D-xylose fermentation in industrial Saccharomyces cerevisiae for efficient lignocellulose-based bioethanol production.* Biotechnology for biofuels, 2013. **6**(1): p. 1-17.
13. Casey, E., et al., *Effect of acetic acid and pH on the cofermentation of glucose and xylose to ethanol by a genetically engineered strain of Saccharomyces cerevisiae.* FEMS yeast research, 2010. **10**(4): p. 385-393.
14. Rizzi, M., et al., *Xylose fermentation by yeasts. 4. Purification and kinetic studies of xylose reductase from Pichia stipitis.* Appl Microbiol Biotechnol, 1988. **29**: p. 148-154.
15. Kwak, S., et al., *Production of biofuels and chemicals from xylose using native and engineered yeast strains.* Biotechnology Advances, 2019. **37**(2): p. 271-283.
16. Schmidt, I., et al., *Biotechnological production of astaxanthin with Phaffia rhodozyma/Xanthophyllomyces dendrorhous.* Appl Microbiol Biotechnol, 2011. **89**(3): p. 555-71.
17. Konzock, O., et al., *Altering the fatty acid profile of Yarrowia lipolytica to mimic cocoa butter by genetic engineering of desaturases.* Microbial Cell Factories, 2022. **21**(1): p. 25.
18. Morita, T., et al., *Production of glycolipid biosurfactants by basidiomycetous yeasts*. Biotechnol Appl Biochem, 2009. **53**(Pt 1): p. 39-49.
19. Braunwald, T., et al., *Effect of different C/N ratios on carotenoid and lipid production by Rhodotorula glutinis.* Appl Microbiol Biotechnol, 2013. **97**(14): p. 6581-8.
20. Yuan, J.S., et al., *Plants to power: bioenergy to fuel the future.* Trends in Plant Science, 2008. **13**(8): p. 421-429.
21. Dias De Oliveira, M.E., B.E. Vaughan, and E.J. Rykiel, *Ethanol as Fuel: Energy, Carbon Dioxide Balances, and Ecological Footprint.* BioScience, 2005. **55**(7): p. 593-602.
22. Waclawovsky, A.J., et al., *Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content.* Plant Biotechnology Journal, 2010. **8**(3): p. 263-276.

23. Chylenski, P., et al., *Enzymatic degradation of sulfite-pulped softwoods and the role of LPMOs.* Biotechnology for Biofuels, 2017. **10**(1): p. 177.
24. Kim, S. and B.E. Dale, *Global potential bioethanol production from wasted crops and crop residues.* Biomass and Bioenergy, 2004. **26**(4): p. 361-375.
25. Fan, L.T., Y.-H. Lee, and M.M. Gharpuray. *The nature of lignocellulosics and their pretreatments for enzymatic hydrolysis*. 1982. Berlin, Heidelberg: Springer Berlin Heidelberg.
26. Palmqvist, E. and B. Hahn-Hägerdahl, *Fermentation of lignocellulosic hydrolysates I - inhibition and detoxification.* Bioresource Technology, 2000. **74**: p. 17-24.
27. Palmqvist, E. and B. Hahn-Hägerdahl, *Fermentation of lignocellulosic hydrolysates II - inhibitors and mechanisms of inhibition.* Bioresource Technology, 2000. **74**: p. 25-33.
28. Camarasa, C., et al., *Phenotypic landscape of Saccharomyces cerevisiae during wine fermentation: evidence for origin-dependent metabolic traits.* PLoS One, 2011. **6**(9): p. e25147.
29. Wang, Q.M., et al., *Surprisingly diverged populations of Saccharomyces cerevisiae in natural environments remote from human activity.* Mol Ecol, 2012. **21**(22): p. 5404-17.
30. Connell, L., et al., *Diversity of Soil Yeasts Isolated from South Victoria Land, Antarctica.* Microbial Ecology, 2008. **56**(3): p. 448-459.
31. Kutty, S.N. and R. Philip, *Marine yeasts—a review.* Yeast, 2008. **25**(7): p. 465-483.
32. Robinson, H.A., A. Pinharanda, and D. Bensasson, *Summer temperature can predict the distribution of wild yeast populations.* Ecology and Evolution, 2016. **6**(4): p. 1236-1250.
33. Mortimer, R.K., *Evolution and variation of the yeast (Saccharomyces) genome.* Genome research, 2000. **10**(4): p. 403-409.
34. Lachance, M.-A., *Yeast biodiversity: how many and how much?* Biodiversity and ecophysiology of yeasts, 2006: p. 1-9.
35. Piškur, J., et al., *How did Saccharomyces evolve to become a good brewer?* TRENDS in Genetics, 2006. **22**(4): p. 183-186.
36. Kurtzman, C.P., J.W. Fell, and T. Boekhout, *The Yeasts (Fifth Edition)*. The Yeasts (Fifth Edition), ed. C.P. Kurtzman, J.W. Fell, and T. Boekhout. 2011, London: Elsevier. xiii.
37. Research, V.M. *Yeast Market*. 2023 09.05.2023]; Available from: https://www.vantagemarketresearch.com/industry-report/yeast-market-1392.
38. González Peña, O.I., M.Á. López Zavala, and H. Cabral Ruelas, *Pharmaceuticals market, consumption trends and disease incidence are not driving the pharmaceutical research on water and wastewater.* International journal of environmental research and public health, 2021. **18**(5): p. 2532.
39. Research, P. *Biotechnology Market Size, share, Growth, Forecast 2023-2030*. 2023 [cited 2023 2023-07-25]; Available from: https://www.precedenceresearch.com/biotechnology-market.
40. Molina-Espeja, P., *Next Generation Winemakers: Genetic Engineering in Saccharomyces cerevisiae for Trendy Challenges.* Bioengineering, 2020. **7**(4): p. 128.
41. Mussatto, S.I., et al., *Technological trends, global market, and challenges of bio-ethanol production.* Biotechnol Adv, 2010. **28**(6): p. 817-30.
42. Nielsen, J., C.B. Tillegreen, and D. Petranovic, *Innovation trends in industrial biotechnology.* Trends in Biotechnology, 2022. **40**(10): p. 1160-1172.
43. Mordor Intelligence. *Yeast and yeast extract market - Growth, trends, COVID-19 impact, and forecasts (2023-2028)*. 2023 09.05.2023]; Available from: https://www.mordorintelligence.com/industry-reports/yeast-and-yeast-extract-market?gclid=CjwKCAjw3ueiBhBmEiwA4BhspBwHiWB-V2V99TCkrIQrB5dn1kCjWrI62OB0_afi36QCR4swQnlxVxoC6BoQAvD_BwE.
44. Bennett, R.J. and B.G. Turgeon, *Fungal Sex: The Ascomycota.* Microbiology Spectrum, 2016. **4**(5): p. 4.5.20.
45. Schmidt-Dannert, C., *Biocatalytic portfolio of Basidiomycota.* Curr Opin Chem Biol, 2016. **31**: p. 40-9.

46. National Center for Biotechnology Information (NCBI). *Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information*. 2023 [cited 2023 2023-06-14]; Available from: https://www.ncbi.nlm.nih.gov/.
47. Mattanovich, D., M. Sauer, and B. Gasser, *Yeast biotechnology: teaching the old dog new tricks.* Microbial Cell Factories, 2014. **13**(1): p. 34.
48. Porro, D., et al., *Replacement of a Metabolic Pathway for Large-Scale Production of Lactic Acid from Engineered Yeasts.* Applied and Environmental Microbiology, 1999. **65**(9): p. 4211-4215.
49. Della-Bianca, B.E., et al., *What do we know about the yeast strains from the Brazilian fuel ethanol industry?* Applied Microbiology and Biotechnology, 2013. **97**(3): p. 979-991.
50. Parapouli, M., et al., *Saccharomyces cerevisiae and its industrial applications.* AIMS Microbiol, 2020. **6**(1): p. 1-31.
51. Bankar, A., L. Jadhav, and V. Phalke, *Metagenomic Insights of Yarrowia lipolytica in Food Industry*, in *Metagenomic Systems Biology: Integrative Analysis of the Microbiome*, S. Singh, Editor. 2020, Springer Singapore: Singapore. p. 159-183.
52. Miller, K.K. and H.S. Alper, *Yarrowia lipolytica: more than an oleaginous workhorse.* Appl Microbiol Biotechnol, 2019. **103**(23-24): p. 9251-9262.
53. Karim, A., N. Gerliani, and M. Aïder, *Kluyveromyces marxianus: An emerging yeast cell factory for applications in food and biotechnology.* International Journal of Food Microbiology, 2020. **333**: p. 108818.
54. Varela, J.A., et al., *Applications of Kluyveromyces marxianus in Biotechnology*, in *Yeast Diversity in Human Welfare*, T. Satyanarayana and G. Kunze, Editors. 2017, Springer Singapore: Singapore. p. 439-453.
55. Amaya-Delgado, L., et al., *Performance evaluation of Pichia kluyveri, Kluyveromyces marxianus and Saccharomyces cerevisiae in industrial tequila fermentation.* World Journal of Microbiology and Biotechnology, 2013. **29**(5): p. 875-881.
56. Karbalaei, M., S.A. Rezaee, and H. Farsiani, *Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins.* J Cell Physiol, 2020. **235**(9): p. 5867-5881.
57. Manfrão-Netto, J.H.C., A.M.V. Gomes, and N.S. Parachin, *Advances in Using Hansenula polymorpha as Chassis for Recombinant Protein Production.* Frontiers in Bioengineering and Biotechnology, 2019. **7**.
58. Ilmén, M., et al., *Efficient production of L-lactic acid from xylose by Pichia stipitis.* Appl Environ Microbiol, 2007. **73**(1): p. 117-23.
59. Yong, K.J. and T.Y. Wu, *Second-generation bioenergy from oilseed crop residues: Recent technologies, techno-economic assessments and policies.* Energy Conversion and Management, 2022. **267**: p. 115869.
60. Johnson, E.A. and M.J. Lewis, *Astaxanthin Formation by the Yeast Phaffia rhodozyma.* Microbiology, 1979. **115**(1): p. 173-183.
61. Yan, W., et al., *Biotechnological applications of the non-conventional yeast Meyerozyma guilliermondii.* Biotechnol Adv, 2021. **46**: p. 107674.
62. Wijayawardene, N.N., K.D. Hyde, and D.-Q. Dai, *Outline of Ascomycota*, in *Encyclopedia of Mycology*, Ó. Zaragoza and A. Casadevall, Editors. 2021, Elsevier: Oxford. p. 246-254.
63. Ruchala, J., et al., *Construction of advanced producers of first- and second-generation ethanol in Saccharomyces cerevisiae and selected species of non-conventional yeasts (Scheffersomyces stipitis, Ogataea polymorpha).* Journal of Industrial Microbiology and Biotechnology, 2020. **47**(1): p. 109-132.
64. Tikka, C., et al., *Isolation and characterization of ethanol tolerant yeast strains.* Bioinformation, 2013. **9**(8): p. 421-5.
65. Goffeau, A., et al., *Life with 6000 genes.* Science, 1996. **274**(5287): p. 546, 563-7.
66. Duan, S.-F., et al., *The origin and adaptive evolution of domesticated populations of yeast from Far East Asia.* Nature Communications, 2018. **9**(1): p. 2690.

67. Wang, Q.-M., et al., *Surprisingly diverged populations of Saccharomyces cerevisiae in natural environments remote from human activity.* Molecular Ecology, 2012. **21**(22): p. 5404-5417.
68. Mortimer, R. and M. Polsinelli, *On the origins of wine yeast.* Res Microbiol, 1999. **150**(3): p. 199-204.
69. Wolfe, K.H. and D.C. Shields, *Molecular evidence for an ancient duplication of the entire yeast genome.* Nature, 1997. **387**(6634): p. 708-713.
70. De Deken, R.H., *The Crabtree Effect: A Regulatory System in Yeast.* Microbiology, 1966. **44**(2): p. 149-156.
71. Malina, C., et al., *Adaptations in metabolism and protein translation give rise to the Crabtree effect in yeast.* Proceedings of the National Academy of Sciences, 2021. **118**(51): p. e2112836118.
72. Bakalinsky, A.T. and R. Snow, *The chromosomal constitution of wine strains of Saccharomyces cerevisiae.* Yeast, 1990. **6**(5): p. 367-82.
73. Gibson, B.R., et al., *Yeast responses to stresses associated with industrial brewery handling.* FEMS Microbiol Rev, 2007. **31**(5): p. 535-69.
74. Pizarro, F., F.A. Vargas, and E. Agosin, *A systems biology perspective of wine fermentations.* Yeast, 2007. **24**(11): p. 977-991.
75. Pretorius, I.S., *Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking.* Yeast, 2000. **16**(8): p. 675-729.
76. Qiao, K., et al., *Engineering lipid overproduction in the oleaginous yeast Yarrowia lipolytica.* Metabolic engineering, 2015. **29**: p. 56-65.
77. Zinjarde, S.S., *Food-related applications of Yarrowia lipolytica.* Food Chemistry, 2014. **152**: p. 1-10.
78. Li, H. and H.S. Alper, *Enabling xylose utilization in Yarrowia lipolytica for lipid production.* Biotechnol J, 2016. **11**(9): p. 1230-40.
79. 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.
80. Bonekamp, F.J. and J. Oosterom, *On the safety of Kluyveromyces lactis - a review.* Applied Microbiology and Biotechnology, 1994. **41**(1): p. 1-3.
81. Singh, D., et al., *Ethanol production at elevated temperatures and alcohol concentrations: Part II – Use of Kluyveromyces marxianus IMB3.* World Journal of Microbiology and Biotechnology, 1998. **14**(6): p. 823-834.
82. Cheon, Y., et al., *A biosynthetic pathway for hexanoic acid production in Kluyveromyces marxianus.* J Biotechnol, 2014. **182-183**: p. 30-6.
83. McTaggart, T.L., et al., *Synthesis of polyketides from low cost substrates by the thermotolerant yeast Kluyveromyces marxianus.* Biotechnol Bioeng, 2019. **116**(7): p. 1721-1730.
84. Lin, Y.J., et al., *Metabolic engineering a yeast to produce astaxanthin.* Bioresour Technol, 2017. **245**(Pt A): p. 899-905.
85. Kötter, P. and M. Ciriacy, *Xylose fermentation by Saccharomyces cerevisiae.* Applied Microbiology and Biotechnology, 1993. **38**(6): p. 776-783.
86. Skoog, K. and B. Hahn-Hägerdal, *Effect of Oxygenation on Xylose Fermentation by Pichia stipitis.* Appl Environ Microbiol, 1990. **56**(11): p. 3389-94.
87. Verho, R., et al., *Engineering redox cofactor regeneration for improved pentose fermentation in Saccharomyces cerevisiae.* Appl Environ Microbiol, 2003. **69**(10): p. 5892-7.
88. Kötter, P., et al., *Isolation and characterization of the Pichia stipitis xylitol dehydrogenase gene, XYL2, and construction of a xylose-utilizing Saccharomyces cerevisiae transformant.* Curr Genet, 1990. **18**(6): p. 493-500.
89. Tantirungkij, M., et al., *Construction of xylose-assimilating Saccharomyces cerevisiae.* Journal of Fermentation and Bioengineering, 1993. **75**(2): p. 83-88.
90. du Preez, J.C., M. Bosch, and B.A. Prior, *Temperature profiles of growth and ethanol tolerance of the xylose-fermenting yeasts Candida shehatae and Pichia stipitis.* Applied Microbiology and Biotechnology, 1987. **25**(6): p. 521-525.

91. Breuer, U. and H. Harms, *Debaryomyces hansenii - an extremophilic yeast with biotechnological potential.* Yeast, 2006. **23**(6): p. 415-37.
92. Ramírez-Orozco, M., N.Y. Hernández-Saavedra, and J.-L. Ochoa, *Debaryomyces hansenii growth in nonsterile seawater ClO_2-peptone-containing medium.* Canadian journal of microbiology, 2001. **47**(7): p. 676-679.
93. Buzzini, P. and A. Martini, *Large-scale screening of selected Candida maltosa, Debaryomyces hansenii and Pichia anomala killer toxin activity against pathogenic yeasts.* Sabouraudia, 2001. **39**(6): p. 479-482.
94. Ryabova, O.B., O.M. Chmil, and A.A. Sibirny, *Xylose and cellobiose fermentation to ethanol by the thermotolerant methylotrophic yeast Hansenula polymorpha.* FEMS Yeast Research, 2003. **4**(2): p. 157-164.
95. Vyas, A., et al., *Fission Yeast Schizosaccharomyces pombe: A Unicellular "Micromammal" Model Organism.* Curr Protoc, 2021. **1**(6): p. e151.
96. Daniel, H.-M., M.-A. Lachance, and C.P. Kurtzman, *On the reclassification of species assigned to Candida and other anamorphic ascomycetous yeast genera based on phylogenetic circumscription.* Antonie van Leeuwenhoek, 2014. **106**(1): p. 67-84.
97. Kurtzman, C.P. and C.J. Robnett, *Relationships among genera of the Saccharomycotina (Ascomycota) from multigene phylogenetic analysis of type species.* FEMS Yeast Research, 2013. **13**(1): p. 23-33.
98. Kidd, S.E., A. Abdolrasouli, and F. Hagen, *Fungal Nomenclature: Managing Change is the Name of the Game.* Open Forum Infectious Diseases, 2023. **10**(1).
99. Webster, J. and R. Weber, *Basidiomycete yeasts*, in *Introduction to Fungi*, J. Webster and R. Weber, Editors. 2007, Cambridge University Press: Cambridge. p. 658-672.
100. Maoka, T., *Carotenoids as natural functional pigments.* J Nat Med, 2020. **74**(1): p. 1-16.
101. Miller, N.J., et al., *Antioxidant activities of carotenes and xanthophylls.* FEBS Lett, 1996. **384**(3): p. 240-2.
102. Moore, M.M., M.W. Breedveld, and A.P. Autor, *The role of carotenoids in preventing oxidative damage in the pigmented yeast, Rhodotorula mucilaginosa.* Arch Biochem Biophys, 1989. **270**(2): p. 419-31.
103. McCready, S.J., F. Osman, and A. Yasui, *Repair of UV damage in the fission yeast Schizosaccharomyces pombe.* Mutat Res, 2000. **451**(1-2): p. 197-210.
104. Moliné, M., et al., *Photoprotective role of carotenoids in yeasts: Response to UV-B of pigmented and naturally-occurring albino strains.* J Photochem Photobiol B, 2009. **95**(3): p. 156-61.
105. Vachali, P., P. Bhosale, and P.S. Bernstein, *Microbial carotenoids.* Methods Mol Biol, 2012. **898**: p. 41-59.
106. Goodwin, T., *The biochemistry of the carotenoids: volume I plants*. 2012: Springer Science & Business Media.
107. Inácio, J., et al., *Estimation and diversity of phylloplane mycobiota on selected plants in a Mediterranean-type ecosystem in Portugal.* Microb Ecol, 2002. **44**(4): p. 344-53.
108. Buzzini, P., et al., *Basidiomycetous Yeasts for Production of Carotenoids*, in *Encyclopedia of Industrial Biotechnology*. 2010. p. 1-13.
109. Martínez-Soto, D. and J. Ruiz-Herrera, *Transcriptomic analysis of the dimorphic transition of Ustilago maydis induced in vitro by a change in pH.* Fungal Genetics and Biology, 2013. **58-59**: p. 116-125.
110. Horst, R.J., et al., *The Ustilago maydis Nit2 Homolog Regulates Nitrogen Utilization and Is Required for Efficient Induction of Filamentous Growth.* Eukaryotic Cell, 2012. **11**(3): p. 368-380.
111. Rodríguez-Kessler, M., et al., *Isolation of UmRrm75, a gene involved in dimorphism and virulence of Ustilago maydis.* Microbiological Research, 2012. **167**(5): p. 270-282.
112. Vollmeister, E., et al., *Fungal development of the plant pathogen Ustilago maydis.* FEMS Microbiology Reviews, 2012. **36**(1): p. 59-77.
113. Morrison, E.N., M.E. Donaldson, and B.J. Saville, *Identification and analysis of genes expressed in the Ustilago maydis dikaryon: uncovering a novel class of pathogenesis genes.* Canadian Journal of Plant Pathology, 2012. **34**(3): p. 417-435.

114. Feldbrügge, M., et al., *Regulation of mating and pathogenic development in Ustilago maydis.* Current Opinion in Microbiology, 2004. **7**(6): p. 666-672.
115. Klosterman, S.J., et al., *Genetics of Morphogenesis and Pathogenic Development of Ustilago maydis*, in *Advances in Genetics*. 2007, Academic Press. p. 1-47.
116. Steinberg, G. and J. Perez-Martin, *Ustilago maydis, a new fungal model system for cell biology.* Trends in Cell Biology, 2008. **18**(2): p. 61-67.
117. Zou, K., et al., *Early infection response of fungal biotroph Ustilago maydis in maize.* Frontiers in Plant Science, 2022. **13**.
118. Juárez-Montiel, M., et al., *Huitlacoche (corn smut), caused by the phytopathogenic fungus Ustilago maydis, as a functional food.* Revista Iberoamericana de Micología, 2011. **28**(2): p. 69-73.
119. Santos, A., et al., *Ustilago maydis killer toxin as a new tool for the biocontrol of the wine spoilage yeast Brettanomyces bruxellensis.* International Journal of Food Microbiology, 2011. **145**(1): p. 147-154.
120. Klement, T., et al., *Biomass pretreatment affects Ustilago maydis in producing itaconic acid.* Microbial Cell Factories, 2012. **11**(1): p. 43.
121. Teichmann, B., et al., *A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from Ustilago maydis.* Molecular Microbiology, 2007. **66**(2): p. 525-533.
122. Olicón-Hernández, D.R., et al., *Response of Ustilago maydis against the Stress Caused by Three Polycationic Chitin Derivatives.* Molecules, 2017. **22**(12): p. 1745.
123. Geiser, E., et al., *Identification of an endo-1,4-beta-xylanase of Ustilago maydis.* BMC Biotechnology, 2013. **13**(1): p. 59.
124. Holliday, R., *Early studies on recombination and DNA repair in Ustilago maydis.* DNA Repair, 2004. **3**(6): p. 671-682.
125. Morita, T., et al., *Characterization of the genus Pseudozyma by the formation of glycolipid biosurfactants, mannosylerythritol lipids.* FEMS Yeast Res, 2007. **7**(2): p. 286-92.
126. Mawani, J., J. Jadhav, and A. Pratap, *Fermentative Production of Mannosylerythritol Lipids using Sweetwater as Waste Substrate by Pseudozyma antarctica (MTCC 2706).* Tenside Surfactants Detergents, 2021. **58**(4): p. 246-258.
127. Shinozaki, Y., et al., *Biodegradable plastic-degrading enzyme from Pseudozyma antarctica: cloning, sequencing, and characterization.* Appl Microbiol Biotechnol, 2013. **97**(7): p. 2951-9.
128. Takakuwa, N., et al., *Efficient conversion of crude glycerol into triacylglycerol by the yeast Pseudozyma sp. TYC-2187 for biodiesel production.* J Oleo Sci, 2013. **62**(8): p. 605-12.
129. Ratledge, C. and J.P. Wynn, *The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms.* Advances in Applied Microbiology, 2002. **51**.
130. Thorpe, R.F. and C. Ratledge, *Fatty Acid Distribution in Triglycerides of Yeasts Grown on Glucose or n-Alkanes.* Microbiology, 1972. **72**(1): p. 151-163.
131. Garay, L.A., et al., *Eighteen new oleaginous yeast species.* Journal of Industrial Microbiology and Biotechnology, 2016. **43**(7): p. 887-900.
132. Papanikolaou, S. and G. Aggelis, *Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production.* European Journal of Lipid Science and Technology, 2011. **113**(8): p. 1031-1051.
133. Qvirist, L., et al., *Screening of xylose utilizing and high lipid producing yeast strains as a potential candidate for industrial application.* BMC Microbiol, 2022. **22**(1): p. 173.
134. Dujon, B., et al., *Genome evolution in yeasts.* Nature, 2004. **430**(6995): p. 35-44.
135. Mamaev, D. and R. Zvyagilskaya, *Yarrowia lipolytica: A multitalented yeast species of ecological significance.* FEMS Yeast Research, 2021. **21**(2): p. foab008.
136. Madzak, C., *Yarrowia lipolytica Strains and Their Biotechnological Applications: How Natural Biodiversity and Metabolic Engineering Could Contribute to Cell Factories Improvement.* J Fungi (Basel), 2021. **7**(7).

137. 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.
138. André, A., et al., *Biotechnological conversions of bio-diesel-derived crude glycerol by Yarrowia lipolytica strains.* Engineering in Life Sciences, 2009. **9**(6): p. 468-478.
139. Wang, K., et al., *Engineering the Lipid and Fatty Acid Metabolism in Yarrowia lipolytica for Sustainable Production of High Oleic Oils.* ACS Synth Biol, 2022. **11**(4): p. 1542-1554.
140. Groenewald, M., et al., *Yarrowia lipolytica: Safety assessment of an oleaginous yeast with a great industrial potential.* Critical Reviews in Microbiology, 2014. **40**(3): p. 187-206.
141. Butinar, L., et al., *Yeast diversity in hypersaline habitats.* FEMS Microbiology Letters, 2005. **244**(2): p. 229-234.
142. Chi, Z., et al., *Selection of Yarrowia lipolytica strains with high protein content from yeasts isolated from different marine environments.* Journal of Ocean University of China, 2007. **6**(4): p. 360-364.
143. Sutherland, J., C. Cornelison, and S. Crow Jr, *CANDIDA | Yarrowia lipolytica (Candida lipolytica).* 2014.
144. Peter, J., et al., *Genome evolution across 1,011 Saccharomyces cerevisiae isolates.* Nature, 2018. **556**(7701): p. 339-344.
145. Barth, G. and C. Gaillardin, *Physiology and genetics of the dimorphic fungus Yarrowia lipolytica.* FEMS Microbiology Reviews, 1997. **19**(4): p. 219-237.
146. Papanikolaou, S., et al., *Biosynthesis of lipids and organic acids by Yarrowia lipolytica strains cultivated on glucose.* European Journal of Lipid Science and Technology, 2009. **111**(12): p. 1221-1232.
147. Pérez-Campo, F.M. and A. Domínguez, *Factors affecting the morphogenetic switch in Yarrowia lipolytica.* Current microbiology, 2001. **43**: p. 429-433.
148. Madzak, C., *Yarrowia lipolytica: recent achievements in heterologous protein expression and pathway engineering.* Applied Microbiology and Biotechnology, 2015. **99**(11): p. 4559-4577.
149. Spagnuolo, M., et al., *Alternative Substrate Metabolism in Yarrowia lipolytica.* Frontiers in Microbiology, 2018. **9**.
150. Ryu, S., J. Hipp, and C.T. Trinh, *Activating and Elucidating Metabolism of Complex Sugars in Yarrowia lipolytica.* Applied and Environmental Microbiology, 2016. **82**(4): p. 1334-1345.
151. Ledesma-Amaro, R., et al., *Metabolic engineering of Yarrowia lipolytica to produce chemicals and fuels from xylose.* Metab Eng, 2016. **38**: p. 115-124.
152. Quarterman, J., et al., *A survey of yeast from the Yarrowia clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate.* Applied Microbiology and Biotechnology, 2017. **101**(8): p. 3319-3334.
153. Abeln, F. and C.J. Chuck, *The history, state of the art and future prospects for oleaginous yeast research.* Microbial Cell Factories, 2021. **20**(1): p. 221.
154. Mierke, F., et al., *Functional genome annotation and transcriptome analysis of Pseudozyma hubeiensis BOT-O, an oleaginous yeast that utilizes glucose and xylose at equal rates.* Fungal Genet Biol, 2023. **166**: p. 103783.
155. Wang, Q.M., J.H. Jia, and F.Y. Bai, *Pseudozyma hubeiensis sp. nov. and Pseudozyma shanxiensis sp. nov., novel ustilaginomycetous anamorphic yeast species from plant leaves.* Int J Syst Evol Microbiol, 2006. **56**(Pt 1): p. 289-93.
156. Botto, E., et al., *Pseudozyma sp. isolation from Eucalyptus leaves and its hydrolytic activity over xylan.* Biocatalysis and Agricultural Biotechnology, 2019. **21**: p. 101282.
157. Morita, T., et al., *Isolation of Pseudozyma churashimaensis sp. nov., a novel ustilaginomycetous yeast species as a producer of glycolipid biosurfactants, mannosylerythritol lipids.* Journal of Bioscience and Bioengineering, 2011. **112**(2): p. 137-144.

158. Sugita, T., et al., *The First Isolation of Ustilaginomycetous Anamorphic Yeasts, Pseudozyma Species, from Patients' Blood and a Description of Two New Species: P. parantarctica and P. thailandica.* Microbiology and Immunology, 2003. **47**(3): p. 183-190.
159. Chang, M.-H., et al., *The isolation and characterization of Pseudozyma sp. JCC 207, a novel producer of squalene.* Applied Microbiology and Biotechnology, 2008. **78**(6): p. 963-972.
160. Tanimura, A., et al., *Lipid production through simultaneous utilization of glucose, xylose, and L-arabinose by Pseudozyma hubeiensis: a comparative screening study.* AMB Express, 2016. **6**(1): p. 58.
161. Konishi, M., et al., *Efficient production of mannosylerythritol lipids with high hydrophilicity by Pseudozyma hubeiensis KM-59.* Appl Microbiol Biotechnol, 2008. **78**(1): p. 37-46.
162. Morita, T., et al., *Accumulation of cellobiose lipids under nitrogen-limiting conditions by two ustilaginomycetous yeasts, Pseudozyma aphidis and Pseudozyma hubeiensis.* FEMS Yeast Res, 2013. **13**(1): p. 44-9.
163. Konishi, M., et al., *Efficient and simple electro-transformation of intact cells for the basidiomycetous fungus Pseudozyma hubeiensis.* Biotechnol Lett, 2015. **37**(8): p. 1679-85.
164. Omae, N., et al., *Disruption of protease A and B orthologous genes in the basidiomycetous yeast Pseudozyma antarctica GB-4(0) yields a stable extracellular biodegradable plastic-degrading enzyme.* PLoS One, 2021. **16**(3): p. e0247462.
165. Yarimizu, T., et al., *Targeted gene replacement at the URA3 locus of the basidiomycetous yeast Pseudozyma antarctica and its transformation using lithium acetate treatment.* Yeast, 2017. **34**(12): p. 483-494.
166. Schuster, M., et al., *Genome editing in Ustilago maydis using the CRISPR–Cas system.* Fungal Genetics and Biology, 2016. **89**: p. 3-9.
167. Wang, Q.M., et al., *Phylogenetic classification of yeasts and related taxa within Pucciniomycotina.* Stud Mycol, 2015. **81**: p. 149-89.
168. Hawksworth, D.L., et al., *The Amsterdam declaration on fungal nomenclature.* IMA fungus, 2011. **2**: p. 105-111.
169. Kraisintu, P., W. Yongmanitchai, and S. Limtong, *Selection and optimization for lipid production of a newly isolated oleaginous yeast, Rhodosporidium toruloides DMKU3-TK16.* Agriculture and Natural Resources, 2010. **44**(3): p. 436-445.
170. Kim, J., et al., *Lipid extract derived from newly isolated Rhodotorula toruloides LAB-07 for cosmetic applications.* Computational and Structural Biotechnology Journal, 2023. **21**: p. 2009-2017.
171. Machado, W., et al., *Production of carotenoids by Rhodotorula toruloides isolated from Brazilian tropical savannah.* International food research journal, 2019. **26**(4).
172. Gerber, U., et al., *Metabolism-dependent bioaccumulation of uranium by Rhodosporidium toruloides isolated from the flooding water of a former uranium mine.* PLOS ONE, 2018. **13**(8): p. e0201903.
173. Coelho, M.A., et al., *Identification of mating type genes in the bipolar basidiomycetous yeast Rhodosporidium toruloides: first insight into the MAT locus structure of the Sporidiobolales.* Eukaryot Cell, 2008. **7**(6): p. 1053-61.
174. Akada, R., et al., *Genomic organization of multiple genes coding for rhodotorucine A, a lipopeptide mating pheromone of the basidiomycetous yeast Rhodosporidium toruloides.* Archives of Microbiology, 1989. **152**(5): p. 484-487.
175. Casselton, L.A. and M.P. Challen, *The Mating Type Genes of the Basidiomycetes*, in *Growth, Differentiation and Sexuality*, U. Kües and R. Fischer, Editors. 2006, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 357-374.
176. May, G., et al., *The signature of balancing selection: fungal mating compatibility gene evolution.* Proceedings of the National Academy of Sciences, 1999. **96**(16): p. 9172-9177.

177. James, T.Y., et al., *Evolution of the gene encoding mitochondrial intermediate peptidase and its cosegregation with the A mating-type locus of mushroom fungi.* Fungal Genetics and Biology, 2004. **41**(3): p. 381-390.
178. Christensen, J., *Studies on the genetics of Ustilago zeae*. 1931.
179. Rowell, J., *Genetics of Ustilago zeae in relation to basic problems of its pathogenicity.* Phytopathology, 1954. **44**: p. 356-362.
180. Puhalla, J.E., *Genetic studies of the b incompatability locus of Ustilago maydis.* Genetics Research, 1970. **16**(2): p. 229-232.
181. Starmer, W.T. and M.-A. Lachance, *Chapter 6 - Yeast Ecology*, in *The Yeasts (Fifth Edition)*, C.P. Kurtzman, J.W. Fell, and T. Boekhout, Editors. 2011, Elsevier: London. p. 65-83.
182. Gai, C.S., et al., *Diversity of endophytic yeasts from sweet orange and their localization by scanning electron microscopy.* J Basic Microbiol, 2009. **49**(5): p. 441-51.
183. Randhawa, H.S., A.Y. Mussa, and Z.U. Khan, *Decaying wood in tree trunk hollows as a natural substrate for Cryptococcus neoformans and other yeast-like fungi of clinical interest.* Mycopathologia, 2001. **151**(2): p. 63-9.
184. Into, P., et al., *Yeast Diversity Associated with the Phylloplane of Corn Plants Cultivated in Thailand.* Microorganisms, 2020. **8**(1).
185. Sampaio, A., J.P. Sampaio, and C. Leão, *Dynamics of yeast populations recovered from decaying leaves in a nonpolluted stream: a 2-year study on the effects of leaf litter type and decomposition time.* FEMS Yeast Res, 2007. **7**(4): p. 595-603.
186. Sláviková, E., R. Vadkertiová, and D. Vránová, *Yeasts colonizing the leaf surfaces.* J Basic Microbiol, 2007. **47**(4): p. 344-50.
187. Sláviková, E., R. Vadkertiová, and D. Vránová, *Yeasts colonizing the leaves of fruit trees.* Annals of Microbiology, 2009. **59**(3): p. 419-424.
188. Zhou, Z.X., et al., *Microbial community on healthy and diseased leaves of an invasive plant Eupatorium adenophorum in Southwest China.* J Microbiol, 2010. **48**(2): p. 139-45.
189. Sibanda, T., et al., *Potential biotechnological capabilities of cultivable mycobiota from carwash effluents.* MicrobiologyOpen, 2017. **6**(5): p. e00498.
190. Baas-Becking, L.G.M., *Geobiologie; of inleiding tot de milieukunde*. 1934: WP Van Stockum & Zoon NV.
191. O'Malley, M.A., *'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography.* Stud Hist Philos Biol Biomed Sci, 2008. **39**(3): p. 314-25.
192. Lachance, M.-A., *Here and There or Everywhere?* BioScience, 2004. **54**(10): p. 884-884.
193. Yurkov, A., *Temporal and Geographic Patterns in Yeast Distribution*, in *Yeasts in Natural Ecosystems: Ecology*, P. Buzzini, M.-A. Lachance, and A. Yurkov, Editors. 2017, Springer International Publishing: Cham. p. 101-130.
194. De Chiara, M., et al., *Domestication reprogrammed the budding yeast life cycle.* Nature Ecology & Evolution, 2022. **6**(4): p. 448-460.
195. Ksiezopolska, E., et al., *Narrow mutational signatures drive acquisition of multidrug resistance in the fungal pathogen Candida glabrata.* Current Biology, 2021. **31**(23): p. 5314-5326. e10.
196. Goddard, M.R. and D. Greig, *Saccharomyces cerevisiae: a nomadic yeast with no niche?* FEMS Yeast Research, 2015. **15**(3).
197. Cavalieri, D., B. Valentini, and I. Stefanini, *Going wild: ecology and genomics are crucial to understand yeast evolution.* Current Opinion in Genetics & Development, 2022. **75**: p. 101922.
198. Lachance, M.-A., *Approaches to yeast identification*, in *Yeast Biotechnology*, D.R. Berry, I. Russell, and G.G. Stewart, Editors. 1987, Springer Netherlands: Dordrecht. p. 33-51.
199. Aon, J.C., R.C. Tecson, and V. Loladze, *Saccharomyces cerevisiae morphological changes and cytokinesis arrest elicited by hypoxia during scale-up for production of therapeutic recombinant proteins.* Microbial Cell Factories, 2018. **17**(1): p. 195.

200. Marechal, P.-A. and P. Gervais, *Passive response of Saccharomyces cerevisiae to osmotic shifts: Cell volume variations depending on the physiological state.* Biochemical and Biophysical Research Communications, 1996. **227**(2): p. 519-523.
201. Knop, M., *Yeast cell morphology and sexual reproduction – A short overview and some considerations.* Comptes Rendus Biologies, 2011. **334**(8): p. 599-606.
202. Gancedo, J.M., *Control of pseudohyphae formation in Saccharomyces cerevisiae.* FEMS Microbiol Rev, 2001. **25**(1): p. 107-23.
203. Zhang, Y., et al., *Comparative Fatty Acid Compositional Profiles of Rhodotorula toruloides Haploid and Diploid Strains under Various Storage Conditions.* Fermentation, 2022. **8**(9): p. 467.
204. Nerurkar, V., et al., *Identifying Candida and other yeast-like fungi: utility of an identification algorithm in resource limited setting.* J Clin Diagn Res, 2014. **8**(12): p. Dc01-4.
205. Kurtzman, C.P. and C.J. Robnett, *Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences.* Antonie Van Leeuwenhoek, 1998. **73**(4): p. 331-71.
206. Chen, Y.C., et al., *Identification of Medically Important Yeasts Using PCR-Based Detection of DNA Sequence Polymorphisms in the Internal Transcribed Spacer 2 Region of the rRNA Genes.* Journal of Clinical Microbiology, 2000. **38**(6): p. 2302-2310.
207. Evans, C.T. and C. Ratledge, *Effect of Nitrogen Source on Lipid Accumulation in Oleaginous Yeasts.* Microbiology, 1984. **130**(7): p. 1693-1704.
208. Albers, E., et al., *Influence of the nitrogen source on Saccharomyces cerevisiae anaerobic growth and product formation.* Appl Environ Microbiol, 1996. **62**(9): p. 3187-95.
209. Wang, M., et al., *Can sustainable ammonia synthesis pathways compete with fossil-fuel based Haber–Bosch processes?* Energy & Environmental Science, 2021. **14**(5): p. 2535-2548.
210. Antonetti, E., et al., *Waste-to-Chemicals for a Circular Economy: The Case of Urea Production (Waste-to-Urea).* ChemSusChem, 2017. **10**(5): p. 912-920.
211. Brabender, M., et al., *Urea and urine are a viable and cost-effective nitrogen source for Yarrowia lipolytica biomass and lipid accumulation.* Appl Microbiol Biotechnol, 2018. **102**(5): p. 2313-2322.
212. Zhu, Z., et al., *A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides.* Nat Commun, 2012. **3**: p. 1112.
213. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* J. Mol. Biol., 1975. **94**: p. 441-448.
214. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the national academy of sciences, 1977. **74**(12): p. 5463-5467.
215. Libkind, D., et al., *Towards yeast taxogenomics: lessons from novel species descriptions based on complete genome sequences.* FEMS Yeast Research, 2020. **20**(6).
216. Maxam, A.M. and W. Gilbert, *A new method for sequencing DNA.* Proc. Natl. Acad. Sci., 1977. **74**(2): p. 560-564.
217. Messing, J., R. Crea, and P.H. Seeburg, *A system for shotgun DNA sequencing.* Nucleic Acids Res, 1981. **9**(2): p. 309-321.
218. Anderson, S., *Shotgun DNA sequencing using cloned DNase I-generated fragments.* Nucleic Acids Research, 1981. **9**(13).
219. International Human Genome Sequencing Consortium, *Initial sequencing and analysis of the human genome.* Articles, 2001.
220. Venter, J.C., et al., *The sequence of the human genome.* Science 2001. **291**: p. 1304-1351.
221. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**(7057): p. 376-80.
222. Schatz, M.C., A.L. Delcher, and S.L. Salzberg, *Assembly of large genomes using second-generation sequencing.* Genome Res, 2010. **20**(9): p. 1165-73.

223. Illumina Inc. *Sequencing by Synthesis (SBS) Technology*. 2023 [cited 2023 05.06.2023]; Available from: https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html.
224. GenScript. *Advancing genomics, medicine and health together - by semiconductor DNA synthesis technology*. 2022 15.06.2023]; Available from: https://www.genscript.com/advancing-genomics-medicine-and-health-together-by-semiconductor-dna-synthesis-technology-summary.html.
225. Illumina Inc. *TruSeq DNA PCR-Free Library Preparation Kit data sheet*. 2017 [cited 2023 14-07-2023]; Available from: https://www.illumina.com/documents/products/datasheets/datasheet_truseq_dna_pcr_free_sample_prep.pdf.
226. Nakazato, T., T. Ohta, and H. Bono, *Experimental Design-Based Functional Mining and Characterization of High-Throughput Sequencing Data in the Sequence Read Archive.* PLOS ONE, 2013. **8**(10): p. e77910.
227. Illumina Inc. *An introduction to Next-Generation Sequencing Technology*. 2017 [cited 2023 14-07-2023]; Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.
228. Treangen, T.J. and S.L. Salzberg, *Repetitive DNA and next-generation sequencing: computational challenges and solutions.* Nat Rev Genet, 2011. **13**(1): p. 36-46.
229. Sholes, S.L., et al., *Chromosome-specific telomere lengths and the minimal functional telomere revealed by nanopore sequencing.* Genome Res, 2022. **32**(4): p. 616-628.
230. De Roeck, A., et al., *NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION.* Genome Biology, 2019. **20**(1): p. 239.
231. Zimin, A.V. and S.L. Salzberg, *The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies.* PLoS Comput Biol, 2020. **16**(6): p. e1007981.
232. Marx, V., *Method of the year: long-read sequencing.* Nature Methods, 2023. **20**(1): p. 6-11.
233. Wang, Y., et al., *Nanopore sequencing technology, bioinformatics and applications.* Nature Biotechnology, 2021. **39**(11): p. 1348-1365.
234. Giordano, F., et al., *De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms.* Sci Rep, 2017. **7**(1): p. 3935.
235. Zhang, Y.-z., et al., *Nanopore basecalling from a perspective of instance segmentation.* BMC Bioinformatics, 2020. **21**(3): p. 136.
236. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
237. De Coster, W., et al., *NanoPack: visualizing and processing long-read sequencing data.* Bioinformatics, 2018. **34**(15): p. 2666-2669.
238. Li, H., *Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences.* Bioinformatics, 2016. **32**(14): p. 2103-10.
239. Li, H., *Minimap2: pairwise alignment for nucleotide sequences.* Bioinformatics, 2018. **34**(18): p. 3094-3100.
240. Tyson, J.R., et al., *MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome.* Genome research, 2018. **28**(2): p. 266-274.
241. Fournier, T., et al., *High-Quality de Novo Genome Assembly of the Dekkera bruxellensis Yeast Using Nanopore MinION Sequencing.* G3 (Bethesda), 2017. **7**(10): p. 3243-3250.
242. Liu, H., et al., *SMARTdenovo: a de novo assembler using long noisy reads.* Gigabyte, 2021. **2021**.
243. Jia, N., et al., *Large-scale comparative analyses of tick genomes elucidate their genetic diversity and vector capacities.* Cell, 2020. **182**(5): p. 1328-1340. e13.
244. Lim, C.S., et al., *Analysis of fungal genomes reveals commonalities of intron gain or loss and functions in intron-poor species.* Molecular Biology and Evolution, 2021. **38**(10): p. 4166-4186.

245. Jung, H., et al., *Twelve quick steps for genome assembly and annotation in the classroom.* PLOS Computational Biology, 2020. **16**(11): p. e1008325.
246. Gribskov, M., J. Devereux, and R.R. Burgess, *The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression.* Nucleic Acids Res, 1984. **12**(1 Pt 2): p. 539-49.
247. Staden, R. and A.D. McLachlan, *Codon preference and its use in identifying protein coding regions in long DNA sequences.* Nucleic Acids Res, 1982. **10**(1): p. 141-56.
248. Yandell, M. and D. Ence, *A beginner's guide to eukaryotic genome annotation.* Nature Reviews Genetics, 2012. **13**(5): p. 329-342.
249. Campbell, M.S., et al., *Genome Annotation and Curation Using MAKER and MAKER-P.* Current Protocols in Bioinformatics, 2014. **2014**: p. 4.11.1-4.11.39.
250. Emms, D.M. and S. Kelly, *OrthoFinder: phylogenetic orthology inference for comparative genomics.* Genome Biology, 2019. **20**(1): p. 238.
251. Wang, X., Q. Liu, and B. Zhang, *Leveraging the complementary nature of RNA-Seq and shotgun proteomics data.* Proteomics, 2014. **14**(23-24): p. 2676-87.
252. McDermaid, A., et al., *Interpretation of differential gene expression results of RNA-seq data: review and integration.* Brief Bioinform, 2019. **20**(6): p. 2044-2054.
253. Nagalakshmi, U., et al., *The transcriptional landscape of the yeast genome defined by RNA sequencing.* Science, 2008. **320**(5881): p. 1344-9.
254. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nature reviews genetics, 2009. **10**(1): p. 57-63.
255. Oxford Nanopore Technologies, *The value of full-length transcripts without bias*. 2019.
256. Soneson, C. and M. Delorenzi, *A comparison of methods for differential expression analysis of RNA-seq data.* BMC Bioinformatics, 2013. **14**(1): p. 91.
257. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2009. **26**(1): p. 139-140.
258. Love, M., S. Anders, and W. Huber, *Differential analysis of count data–the DESeq2 package.* Genome Biol, 2014. **15**(550): p. 10-1186.
259. Schurch, N.J., et al., *How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?* Rna, 2016. **22**(6): p. 839-51.
260. Gallego Romero, I., et al., *RNA-seq: impact of RNA degradation on transcript quantification.* BMC Biology, 2014. **12**(1): p. 42.
261. Schroeder, A., et al., *The RIN: an RNA integrity number for assigning integrity values to RNA measurements.* BMC Molecular Biology, 2006. **7**(1): p. 3.
262. Conrad, M., et al., *Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae.* FEMS Microbiol Rev, 2014. **38**(2): p. 254-99.
263. Tesniere, C., et al., *Relief from nitrogen starvation triggers transient destabilization of glycolytic mRNAs in Saccharomyces cerevisiae cells.* Mol Biol Cell, 2018. **29**(4): p. 490-498.
264. Aliyu, H., et al., *Global Transcriptome Profile of the Oleaginous Yeast Saitozyma podzolica DSM 27192 Cultivated in Glucose and Xylose.* Journal of Fungi, 2021. **7**(9): p. 758.
265. Paul, D., et al., *Genome Sequence of the Oleaginous Yeast Rhodotorula glutinis ATCC 204091.* Genome Announc, 2014. **2**(1).
266. Bommareddy, R.R., W. Sabra, and A.P. Zeng, *Glucose-mediated regulation of glycerol uptake in Rhodosporidium toruloides: Insights through transcriptomic analysis on dual substrate fermentation.* Eng Life Sci, 2017. **17**(3): p. 282-291.
267. Lahtvee, P.J., et al., *Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast.* Cell Syst, 2017. **4**(5): p. 495-504.e5.
268. den Ridder, M., et al., *A systematic evaluation of yeast sample preparation protocols for spectral identifications, proteome coverage and post-isolation modifications.* Journal of Proteomics, 2022. **261**: p. 104576.

269. Martín-Hernández, G.C., et al., *Chromosome-level genome assembly and transcriptome-based annotation of the oleaginous yeast Rhodotorula toruloides CBS 14.* Genomics, 2021. **113**(6): p. 4022-4027.
270. Shaigani, P., et al., *Mastering targeted genome engineering of GC-rich oleaginous yeast for tailored plant oil alternatives for the food and chemical sector.* Microbial Cell Factories, 2023. **22**(1): p. 25.
271. Bracharz, F., et al., *Opportunities and challenges in the development of Cutaneotrichosporon oleaginosus ATCC 20509 as a new cell factory for custom tailored microbial oils.* Microbial Cell Factories, 2017. **16**(1): p. 178.
272. Lyu, L., et al., *Engineering the Oleaginous Yeast Rhodosporidium toruloides for Improved Resistance Against Inhibitors in Biomass Hydrolysates.* Frontiers in Bioengineering and Biotechnology, 2021. **9**.
273. Zhang, S., et al., *Metabolic engineering of the oleaginous yeast Rhodosporidium toruloides IFO0880 for lipid overproduction during high-density fermentation.* Appl Microbiol Biotechnol, 2016. **100**(21): p. 9393-9405.
274. Zhang, S., et al., *Engineering Rhodosporidium toruloides for increased lipid production.* Biotechnology and Bioengineering, 2016. **113**(5): p. 1056-1066.
275. Zhu, Z., et al., *Expanding the product portfolio of fungal type I fatty acid synthases.* Nat Chem Biol, 2017. **13**(4): p. 360-362.
276. Finn, R.D., J. Clements, and S.R. Eddy, *HMMER web server: interactive sequence similarity searching.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W29-37.
277. Wernig, F., et al., *Fusing alpha and beta subunits of the fungal fatty acid synthase leads to improved production of fatty acids.* Sci Rep, 2020. **10**(1): p. 9780.
278. Geijer, C., R. Ledesma-Amaro, and E. Tomás-Pejó, *Unraveling the potential of non-conventional yeasts in biotechnology.* FEMS Yeast Research, 2022. **22**(1).
279. Park, Y.-K., et al., *Bioproducts generation from carboxylate platforms by the non-conventional yeast Yarrowia lipolytica.* FEMS Yeast Research, 2021. **21**(6).
280. Martinez-Jimenez, F.D., et al., *Exploiting the Non-conventional Yeast Spathaspora passalidarum as a Platform for Hemicellulosic Hydrolysate Conversion into Bioproducts: a Mini Review.* BioEnergy Research, 2021. **14**(2): p. 689-708.
281. Cardoso, L.A., et al., *Biotechnological production of carotenoids and their applications in food and pharmaceutical products.* Carotenoids, 2017. **8**: p. 125-141.
282. Albers, S.C., A.M. Berklund, and G.D. Graff, *The rise and fall of innovation in biofuels.* Nature Biotechnology, 2016. **34**(8): p. 814-821.
283. Zabed, H., et al., *Bioethanol production from renewable sources: Current perspectives and technological progress.* Renewable and Sustainable Energy Reviews, 2017. **71**: p. 475-501.
284. Balat, M., *Production of bioethanol from lignocellulosic materials via the biochemical pathway: A review.* Energy Conversion and Management, 2011. **52**(2): p. 858-875.
285. Hsieh, W.-D., et al., *Engine performance and pollutant emission of an SI engine using ethanol–gasoline blended fuels.* Atmospheric Environment, 2002. **36**(3): p. 403-410.
286. Wu, C.W., et al. *The influence of air-fuel ratio on engine performance and pollutant emission of an SI engine using ethanol-gasoline-blended fuels*. in *Atmospheric Environment*. 2004.
287. Liska, A.J., et al., *Improvements in Life Cycle Energy Efficiency and Greenhouse Gas Emissions of Corn-Ethanol.* Journal of Industrial Ecology, 2009. **13**(1): p. 58-74.
288. Searchinger, T., et al., *Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change.* Science, 2008. **319**(5867): p. 1238-40.
289. Crutzen, P.J., et al., *N_2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels.* Paul J. Crutzen: A pioneer on atmospheric chemistry and climate change in the anthropocene, 2016: p. 227-238.
290. Macedo, I.C., J.E. Seabra, and J.E. Silva, *Green house gases emissions in the production and use of ethanol from sugarcane in Brazil: the 2005/2006 averages and a prediction for 2020.* Biomass and bioenergy, 2008. **32**(7): p. 582-595.

291. Bryngemark, E., *Second generation biofuels and the competition for forest raw materials: A partial equilibrium analysis of Sweden.* Forest Policy and Economics, 2019. **109**: p. 102022.
292. Demirbas, A., *Biofuels securing the planet's future energy needs.* Energy Conversion and Management, 2009. **50**(9): p. 2239-2249.
293. Goldemberg, J., *Ethanol for a sustainable energy future.* science, 2007. **315**(5813): p. 808-810.
294. Jansen, M.L.A., et al., *Saccharomyces cerevisiae strains for second-generation ethanol production: from academic exploration to industrial implementation.* FEMS Yeast Research, 2017. **17**(5).
295. Patel, A., et al., *Sustainable biodiesel production from oleaginous yeasts utilizing hydrolysates of various non-edible lignocellulosic biomasses.* Renewable and Sustainable Energy Reviews, 2016. **62**: p. 836-855.
296. Kosior, E. and J. Mitchell, *Chapter 6 - Current industry position on plastic production and recycling*, in *Plastic Waste and Recycling*, T.M. Letcher, Editor. 2020, Academic Press. p. 133-162.
297. Lemoigne, M., *Produits de deshydration et de polymerisation de l'acide 3-oxybutyrique.* Bull Soc Chim Biol (Paris), 1926. **8**: p. 770.
298. Varsha, Y. and R. Savitha, *Overview on polyhydroxyalkanoates: a promising biopol.* J. Microb. Biochem. Technol, 2011. **3**(05): p. 99-105.
299. Tang, R., et al., *Metabolic engineering of Cupriavidus necator H16 for improved chemoautotrophic growth and PHB production under oxygen-limiting conditions.* Metabolic engineering, 2020. **61**: p. 11-23.
300. Breuer, U., et al., *Yeasts as producers of polyhydroxyalkanoates - Genetic engineering of Saccharomyces cerevisiae.* Macromol Biosci, 2002. **2**: p. 380-386.
301. de Las Heras, A.M., et al., *Anaerobic poly-3-D-hydroxybutyrate production from xylose in recombinant Saccharomyces cerevisiae using a NADH-dependent acetoacetyl-CoA reductase.* Microb Cell Fact, 2016. **15**(1): p. 197.
302. Abd-El-haleem, D., et al., *Biosynthesis of biodegradable polyhydroxyalkanotes biopolymers in genetically modified yeasts.* International Journal of Environmental Science & Technology, 2007. **4**: p. 513-520.
303. Aristya, G.R., et al., *Polyhydroxybutyrate (PHB) production from crude glycerol by genetic engineering of Rhodotorula glutinis.* Bioresource Technology Reports, 2022. **18**: p. 101048.
304. Caporusso, A., A. Capece, and I. De Bari, *Oleaginous Yeasts as Cell Factories for the Sustainable Production of Microbial Lipids by the Valorization of Agri-Food Wastes.* Fermentation, 2021. **7**(2): p. 50.
305. Verna, R., *The history and science of chocolate.* Malays J Pathol, 2013. **35**(2): p. 111-21.
306. Schroth, G., et al., *Vulnerability to climate change of cocoa in West Africa: Patterns, opportunities and limits to adaptation.* Sci Total Environ, 2016. **556**: p. 231-41.
307. Clough, Y., H. Faust, and T. Tscharntke, *Cacao boom and bust: sustainability of agroforests and opportunities for biodiversity conservation.* Conservation Letters, 2009. **2**(5): p. 197-205.
308. Bergenholm, D., et al., *Modulation of saturation and chain length of fatty acids in Saccharomyces cerevisiae for production of cocoa butter-like lipids.* Biotechnol Bioeng, 2018. **115**(4): p. 932-942.
309. Wei, Y., et al., *Expression of cocoa genes in Saccharomyces cerevisiae improves cocoa butter production.* Microbial Cell Factories, 2018. **17**(1): p. 11.
310. Wei, Y., et al., *Increasing cocoa butter-like lipid production of Saccharomyces cerevisiae by expression of selected cocoa genes.* AMB Express, 2017. **7**(1): p. 34.
311. Ferreira, R., et al., *Metabolic engineering of Saccharomyces cerevisiae for overproduction of triacylglycerols.* Metabolic Engineering Communications, 2018. **6**: p. 22-27.
312. Ward, O.P. and A. Singh, *Omega-3/6 fatty acids: Alternative sources of production.* Process Biochemistry, 2005. **40**(12): p. 3627-3652.

313. Wei, Y., V. Siewers, and J. Nielsen, *Cocoa butter-like lipid production ability of non-oleaginous and oleaginous yeasts under nitrogen-limited culture conditions.* Applied microbiology and biotechnology, 2017. **101**: p. 3577-3585.
314. Papanikolaou, S., et al., *Accumulation of a cocoa-butter-like lipid by Yarrowia lipolytica cultivated on agro-industrial residues.* Current Microbiology, 2003. **46**: p. 0124-0130.
315. Fitzherbert, E.B., et al., *How will oil palm expansion affect biodiversity?* Trends in ecology & evolution, 2008. **23**(10): p. 538-545.
316. Lovejoy, T.E. and C. Nobre, *Amazon Tipping Point.* Science Advances, 2018. **4**(2): p. eaat2340.
317. Ellwanger, J.H., et al., *Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health.* Anais da Academia Brasileira de Ciências, 2020. **92**.
318. Spier, F., J.G. Buffon, and C.A. Burkert, *Bioconversion of raw glycerol generated from the synthesis of biodiesel by different oleaginous yeasts: lipid content and fatty acid profile of biomass.* Indian Journal of Microbiology, 2015. **55**: p. 415-422.
319. Anschau, A., et al., *Effect of feeding strategies on lipid production by Lipomyces starkeyi.* Bioresource technology, 2014. **157**: p. 214-222.
320. Juanssilfero, A.B., et al., *Lipid production by Lipomyces starkeyi using sap squeezed from felled old oil palm trunks.* Journal of bioscience and bioengineering, 2019. **127**(6): p. 726-731.
321. Liu, Y., et al., *Enhanced lipid production with undetoxified corncob hydrolysate by Rhodotorula glutinis using a high cell density culture strategy.* Bioresource technology, 2015. **180**: p. 32-39.
322. Misra, S., A. Ghosh, and J. Dutta, *Production and composition of microbial fat from Rhodotorula glutinis.* Journal of the Science of Food and Agriculture, 1984. **35**(1): p. 59-65.
323. Kunthiphun, S., et al., *Diversity and characterization of cultivable oleaginous yeasts isolated from mangrove forests.* World Journal of Microbiology and Biotechnology, 2018. **34**: p. 1-11.
324. Saini, R., et al., *Effect of creating a fed-batch like condition using carbon to nitrogen ratios on lipid accumulation in Rhodosporidium toruloides-1588.* Bioresource Technology, 2021. **337**: p. 125354.
325. Parsons, S., et al., *Techno-economic analysis (TEA) of microbial oil production from waste resources as part of a biorefinery concept: assessment at multiple scales under uncertainty.* Journal of Chemical Technology & Biotechnology, 2019. **94**(3): p. 701-711.
326. Sandmann, G., *Antioxidant protection from UV-and light-stress related to carotenoid structures.* Antioxidants, 2019. **8**(7): p. 219.
327. Paul, D., P.K. Kumari, and N. Siddiqui, *Yeast Carotenoids: Cost-Effective Fermentation Strategies for Health Care Applications.* Fermentation, 2023. **9**(2): p. 147.
328. Sreeharsha, R.V. and S.V. Mohan, *Obscure yet Promising Oleaginous Yeasts for Fuel and Chemical Production.* Trends in Biotechnology, 2020. **38**(8): p. 873-887.
329. Ravn, J.L., et al., *CAZyme prediction in ascomycetous yeast genomes guides discovery of novel xylanolytic species with diverse capacities for hemicellulose hydrolysis.* Biotechnology for Biofuels, 2021. **14**(1): p. 150.
330. De Pourcq, K., K. De Schutter, and N. Callewaert, *Engineering of glycosylation in yeast and other fungi: current state and perspectives.* Applied Microbiology and Biotechnology, 2010. **87**(5): p. 1617-1631.
331. Nielsen, J., *Production of biopharmaceutical proteins by yeast: advances through metabolic engineering.* Bioengineered, 2013. **4**(4): p. 207-11.
332. Madhavan, A., et al., *Customized yeast cell factories for biopharmaceuticals: from cell engineering to process scale up.* Microbial Cell Factories, 2021. **20**(1): p. 124.
333. Kulagina, N., et al., *Yeasts as Biopharmaceutical Production Platforms.* Frontiers in Fungal Biology, 2021. **2**.
334. Chen, G.Q. and X.R. Jiang, *Engineering microorganisms for improving polyhydroxyalkanoate biosynthesis.* Curr Opin Biotechnol, 2018. **53**: p. 20-25.

335. Kildegaard, K.R., et al., *Engineering and systems-level analysis of Saccharomyces cerevisiae for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway.* Microb Cell Fact, 2016. **15**: p. 53.
336. Martin, V.J., et al., *Engineering a mevalonate pathway in Escherichia coli for production of terpenoids.* Nat Biotechnol, 2003. **21**(7): p. 796-802.
337. Rodriguez, S., et al., *ATP citrate lyase mediated cytosolic acetyl-CoA biosynthesis increases mevalonate production in Saccharomyces cerevisiae.* Microbial Cell Factories, 2016. **15**(1): p. 48.
338. Tang, X., H. Feng, and W.N. Chen, *Metabolic engineering for enhanced fatty acids synthesis in Saccharomyces cerevisiae.* Metabolic Engineering, 2013. **16**: p. 95-102.
339. Hossain, A.H., et al., *Metabolic engineering with ATP-citrate lyase and nitrogen source supplementation improves itaconic acid production in Aspergillus niger.* Biotechnology for Biofuels, 2019. **12**(1): p. 233.
340. Chen, Y., V. Siewers, and J. Nielsen, *Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in Saccharomyces cerevisiae.* PLoS One, 2012. **7**(8): p. e42475.
341. Shi, L. and B.P. Tu, *Acetyl-CoA and the regulation of metabolism: mechanisms and consequences.* Curr Opin Cell Biol, 2015. **33**: p. 125-31.
342. Srere, P.A., *The citrate cleavage enzyme: I. Distribution and purification.* Journal of Biological Chemistry, 1959. **234**(10): p. 2544-2547.
343. Chen, Y., et al., *Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism.* Metabolic Engineering, 2013. **15**: p. 48-54.
344. Ku, J.T., A.Y. Chen, and E.I. Lan, *Metabolic Engineering Design Strategies for Increasing Acetyl-CoA Flux.* Metabolites, 2020. **10**(4).
345. Nielsen, J., *Synthetic Biology for Engineering Acetyl Coenzyme A Metabolism in Yeast.* mBio, 2014. **5**(6): p. 10.1128/mbio.02153-14.
346. Brink, D.P., et al., *Evaluation of the potential of the novel Rhodotorula toruloides isolate BOT-A2 for lipid production and genetic engineering applications* Manuscript in progress, 2023.
347. Hwang, H.-H., M. Yu, and E.-M. Lai, *Agrobacterium-mediated plant transformation: biology and applications.* The Arabidopsis Book, 2017. **15**.
348. Smith, E.F. and C.O. Townsend, *A plant-tumor of bacterial origin.* Science, 1907. **25**(643): p. 671-673.
349. Bundock, P., et al., *Trans-kingdom T-DNA transfer from Agrobacterium tumefaciens to Saccharomyces cerevisiae.* The EMBO journal, 1995. **14**(13): p. 3206-3214.
350. De Groot, M.J., et al., *Agrobacterium tumefaciens-mediated transformation of filamentous fungi.* Nature biotechnology, 1998. **16**(9): p. 839-842.
351. Kunik, T., et al., *Genetic transformation of HeLa cells by Agrobacterium.* Proceedings of the National Academy of Sciences, 2001. **98**(4): p. 1871-1876.
352. Lacroix, B., et al., *A case of promiscuity: Agrobacterium's endless hunt for new partners.* Trends in Genetics, 2006. **22**(1): p. 29-37.
353. Hooykaas, P.J.J., et al., *Agrobacterium-Mediated Transformation of Yeast and Fungi*, in *Agrobacterium Biology: From Basic Science to Biotechnology*, S.B. Gelvin, Editor. 2018, Springer International Publishing: Cham. p. 349-374.
354. Kiyokawa, K., et al., *Construction of versatile yeast plasmid vectors transferable by Agrobacterium-mediated transformation and their application to bread-making yeast strains.* Journal of Bioscience and Bioengineering, 2023.
355. Liu, Y., et al., *Characterization of glyceraldehyde-3-phosphate dehydrogenase gene RtGPD1 and development of genetic transformation method by dominant selection in oleaginous yeast Rhodosporidium toruloides.* Appl Microbiol Biotechnol, 2013. **97**(2): p. 719-29.
356. Wang, Y., et al., *Cloning and evaluation of different constitutive promoters in the oleaginous yeast Rhodosporidium toruloides.* Yeast, 2016. **33**(3): p. 99-106.
357. Liu, Y., et al., *Developing a set of strong intronic promoters for robust metabolic engineering in oleaginous Rhodotorula (Rhodosporidium) yeast species.* Microbial Cell Factories, 2016. **15**(1): p. 200.

358. Nora, L.C., et al., *A toolset of constitutive promoters for metabolic engineering of Rhodosporidium toruloides.* Microb Cell Fact, 2019. **18**(1): p. 117.
359. Guo, X., et al., *Mining and application of constitutive promoters from Rhodosporidium toruloides.* AMB Express, 2023. **13**(1): p. 17.
360. Johns, A.M., J. Love, and S.J. Aves, *Four Inducible Promoters for Controlled Gene Expression in the Oleaginous Yeast Rhodotorula toruloides.* Front Microbiol, 2016. **7**: p. 1666.
361. Iiizumi, S., et al., *Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting.* Nucleic Acids Res, 2008. **36**(19): p. 6333-42.
362. Kirby, J., et al., *Further engineering of R. toruloides for the production of terpenes from lignocellulosic biomass.* Biotechnology for Biofuels, 2021. **14**(1): p. 101.
363. Parekh, R.N., M.R. Shaw, and K.D. Wittrup, *An integrating vector for tunable, high copy, stable integration into the dispersed Ty delta sites of Saccharomyces cerevisiae.* Biotechnol Prog, 1996. **12**(1): p. 16-21.
364. Tatchell, K. and L.C. Robinson, *Use of green fluorescent protein in living yeast cells.* Methods Enzymol, 2002. **351**: p. 661-83.
365. He, L., et al., *In vivo study of gene expression with an enhanced dual-color fluorescent transcriptional timer.* eLife, 2019. **8**: p. e46181.
366. Matsumae, A., Y. Kamio, and T. Hata, *Studies on cerulenin. I. Studies on cerulenin producing strain.* J Antibiot (Tokyo), 1963. **16**: p. 236-8.
367. Hata, T., et al., *Study of new antifungal antibiotic.* Jpn. J. Bacteriol, 1960. **15**: p. 1075-1077.
368. Nomura, S., et al., *The action mechanism of cerulenin. I. Effect of cerulenin on sterol and fatty acid biosynthesis in yeast.* J Biochem, 1972. **71**(5): p. 783-96.
369. Vance, D., et al., *Inhibition of fatty acid synthetases by the antibiotic cerulenin.* Biochemical and Biophysical Research Communications, 1972. **48**(3): p. 649-656.
370. Johansson, P., et al., *Inhibition of the fungal fatty acid synthase type I multienzyme complex.* Proc Natl Acad Sci U S A, 2008. **105**(35): p. 12803-8.
371. Tehlivets, O., K. Scheuringer, and S.D. Kohlwein, *Fatty acid synthesis and elongation in yeast.* Biochim Biophys Acta, 2007. **1771**(3): p. 255-70.
372. Liu, D., et al., *Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor–actuator.* ACS synthetic biology, 2015. **4**(2): p. 132-140.
373. Ichikawa, E., et al., *Breeding of a Sake Yeast with Improved Ethyl Caproate Productivity.* Agricultural and Biological Chemistry, 1991. **55**(8): p. 2153-2154.
374. Nagai, Y., et al., *A Saccharomyces cerevisiae strain encoding a novel FAS2 mutation produces high levels of caprylic acid.* Mycoscience, 2016. **57**(3): p. 228-231.
375. Yamada, R., T. Kashihara, and H. Ogino, *Improvement of lipid production by the oleaginous yeast Rhodosporidium toruloides through UV mutagenesis.* World J Microbiol Biotechnol, 2017. **33**(5): p. 99.
376. Yamada, R., et al., *Evaluation of lipid production from xylose and glucose/xylose mixed sugar in various oleaginous yeasts and improvement of lipid production by UV mutagenesis.* Biochemical Engineering Journal, 2017. **128**: p. 76-82.
377. Tapia V, E., et al., *Optimization of lipid production by the oleaginous yeast Lipomyces starkeyi by random mutagenesis coupled to cerulenin screening.* AMB Express, 2012. **2**(1): p. 64.
378. Katre, G., et al., *Mutants of Yarrowia lipolytica NCIM 3589 grown on waste cooking oil as a biofactory for biodiesel production.* Microb Cell Fact, 2017. **16**(1): p. 176.
379. Zhang, M., et al., *Metabolic engineering of Rhodotorula toruloides for resveratrol production.* Microbial Cell Factories, 2022. **21**(1): p. 270.
380. Liu, Y., et al., *Identification of novel genes in the carotenogenic and oleaginous yeast Rhodotorula toruloides through genome-wide insertional mutagenesis.* BMC Microbiol, 2018. **18**(1): p. 14.
381. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature Protocols, 2015. **10**(6): p. 845-858.

382. Vasconcelos, B., et al., *Oleaginous yeasts for sustainable lipid production-from biodiesel to surf boards, a wide range of "green" applications.* Appl Microbiol Biotechnol, 2019. **103**(9): p. 3651-3667.
383. Johnson, D.T. and K.A. Taconi, *The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production.* Environmental Progress, 2007. **26**(4): p. 338-348.
384. Papanikolaou, S. and G. Aggelis, *Lipid production by Yarrowia lipolytica growing on industrial glycerol in a single-stage continuous culture.* Bioresource technology, 2002. **82**(1): p. 43-49.
385. Chang, H.N., et al., *Biomass-derived volatile fatty acid platform for fuels and chemicals.* Biotechnology and Bioprocess Engineering, 2010. **15**: p. 1-10.
386. Fontanille, P., et al., *Bioconversion of volatile fatty acids into lipids by the oleaginous yeast Yarrowia lipolytica.* Bioresource technology, 2012. **114**: p. 443-449.
387. Gonçalves, C., et al., *Biological treatment of olive mill wastewater by non-conventional yeasts.* Bioresource technology, 2009. **100**(15): p. 3759-3763.
388. Papanikolaou, S., et al., *Citric acid production by Yarrowia lipolytica cultivated on olive-mill wastewater-based media.* Bioresource Technology, 2008. **99**(7): p. 2419-2428.
389. Mathew, A.K., et al., *Lignocellulosic biorefinery wastes, or resources?*, in *Waste biorefinery*. 2018, Elsevier. p. 267-297.
390. Balat, M. and H. Balat, *Recent trends in global production and utilization of bio-ethanol fuel.* Applied energy, 2009. **86**(11): p. 2273-2282.
391. Mohd Azhar, S.H., et al., *Yeasts in sustainable bioethanol production: A review.* Biochemistry and Biophysics Reports, 2017. **10**: p. 52-61.
392. Brodin, M., et al., *Lignocellulosics as sustainable resources for production of bioplastics – A review.* Journal of Cleaner Production, 2017. **162**: p. 646-664.
393. Wu, S., et al., *Microbial lipid production by Rhodosporidium toruloides under sulfate-limited conditions.* Bioresour Technol, 2011. **102**(2): p. 1803-7.
394. Zhao, X., et al., *Lipid production by Rhodosporidium toruloides Y4 using different substrate feeding strategies.* Journal of Industrial Microbiology & Biotechnology, 2010. **38**(5): p. 627-632.
395. Slininger, P.J., et al., *Comparative lipid production by oleaginous yeasts in hydrolyzates of lignocellulosic biomass and process strategy for high titers.* Biotechnol Bioeng, 2016. **113**(8): p. 1676-90.
396. Gírio, F.M., et al., *Hemicelluloses for fuel ethanol: a review.* Bioresource technology, 2010. **101**(13): p. 4775-4800.
397. Bajpai, P., *Structure of lignocellulosic biomass.* Pretreatment of lignocellulosic biomass for biofuel production, 2016: p. 7-12.
398. Morris, S., S. Hanna, and M. Miles, *The self-assembly of plant cell wall components by single-molecule force spectroscopy and Monte Carlo modelling.* Nanotechnology, 2004. **15**(9): p. 1296.
399. Balakshin, M., et al., *Quantification of lignin–carbohydrate linkages with high-resolution NMR spectroscopy.* Planta, 2011. **233**: p. 1097-1110.
400. Tan, L., et al., *An Arabidopsis cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein.* The Plant Cell, 2013. **25**(1): p. 270-287.
401. Scheller, H.V. and P. Ulvskov, *Hemicelluloses.* Annual review of plant biology, 2010. **61**: p. 263-289.
402. Ralph, J., et al., *Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids.* Phytochemistry Reviews, 2004. **3**(1): p. 29-60.
403. Vanholme, R., et al., *Lignin biosynthesis and structure.* Plant physiology, 2010. **153**(3): p. 895-905.
404. Vermaas, J.V., et al., *Mechanism of lignin inhibition of enzymatic biomass deconstruction.* Biotechnology for biofuels, 2015. **8**(1): p. 1-16.
405. Nelsen, M.P., et al., *Delayed fungal evolution did not cause the Paleozoic peak in coal production.* Proceedings of the National Academy of Sciences, 2016. **113**(9): p. 2442-2447.

406. Clark, T. and K. Mackie, *Steam explosion of the softwood Pinus radiata with sulphur dioxide addition. I. Process optimisation.* Journal of Wood Chemistry and Technology, 1987. **7**(3): p. 373-403.
407. Tengborg, C., et al., *Comparison of SO_2 and H_2SO_4 impregnation of softwood prior to steam pretreatment on ethanol production.* Applied Biochemistry and Biotechnology, 1998. **70**: p. 3-15.
408. Linde, M., et al., *Steam pretreatment of dilute H_2SO_4-impregnated wheat straw and SSF with low yeast and enzyme loadings for bioethanol production.* Biomass and bioenergy, 2008. **32**(4): p. 326-332.
409. Dunlop, A.P., *Furfural Formation and Behavior.* Industrial & Engineering Chemistry, 1948. **40**(2): p. 204-209.
410. Ulbricht, R.J., S.J. Northup, and J.A. Thomas, *A review of 5-hydroxymethylfurfural (HMF) in parenteral solutions.* Fundam Appl Toxicol, 1984. **4**(5): p. 843-53.
411. Jönsson, L.J., B. Alriksson, and N.-O. Nilvebrant, *Bioconversion of lignocellulose: inhibitors and detoxification.* Biotechnology for biofuels, 2013. **6**(1): p. 1-10.
412. Tomás-Pejó, E., et al., *Adaptation of the xylose fermenting yeast Saccharomyces cerevisiae F12 for improving ethanol production in different fed-batch SSF processes.* Journal of Industrial Microbiology and Biotechnology, 2010. **37**(11): p. 1211-1220.
413. Tomás-Pejó, E. and L. Olsson, *Influence of the propagation strategy for obtaining robust Saccharomyces cerevisiae cells that efficiently co-ferment xylose and glucose in lignocellulosic hydrolysates.* Microbial biotechnology, 2015. **8**(6): p. 999-1005.
414. Nielsen, F., et al., *Short-term adaptation during propagation improves the performance of xylose-fermenting Saccharomyces cerevisiae in simultaneous saccharification and co-fermentation.* Biotechnology for Biofuels, 2015. **8**(1): p. 219.
415. Dombek, K.M. and L.O. Ingram, *Magnesium limitation and its role in apparent toxicity of ethanol during yeast fermentation.* Appl Environ Microbiol, 1986. **52**(5): p. 975-81.
416. Winter, J.F., M.-O. Loret, and J.-L. Uribelarrea, *Inhibition and growth factor deficiencies in alcoholic fermentation by Saccharomyces cerevisiae.* Current Microbiology, 1989. **18**(4): p. 247-252.
417. Alfenore, S., et al., *Improving ethanol production and viability of Saccharomyces cerevisiae by a vitamin feeding strategy during fed-batch process.* Appl Microbiol Biotechnol, 2002. **60**(1-2): p. 67-72.
418. Zhao, X.Q., et al., *Impact of zinc supplementation on the improvement of ethanol tolerance and yield of self-flocculating yeast in continuous ethanol fermentation.* J Biotechnol, 2009. **139**(1): p. 55-60.
419. Ko, J.K., Y. Um, and S.M. Lee, *Effect of manganese ions on ethanol fermentation by xylose isomerase expressing Saccharomyces cerevisiae under acetic acid stress.* Bioresour Technol, 2016. **222**: p. 422-430.
420. van Dijk, M., et al., *Nutrient-supplemented propagation of Saccharomyces cerevisiae improves its lignocellulose fermentation ability.* AMB Express, 2020. **10**(1): p. 157.
421. Stambuk, B.U., et al., *Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis.* Genome research, 2009. **19**(12): p. 2271-2278.
422. Russell, J.B., *Another explanation for the toxicity of fermentation acids at low pH: anion accumulation versus uncoupling.* Journal of Applied Bacteriology, 1992. **73**(5): p. 363-370.
423. Mira, N.P., M.C. Teixeira, and I. Sá-Correia, *Adaptive Response and Tolerance to Weak Acids in Saccharomyces cerevisiae: A Genome-Wide View.* OMICS: A Journal of Integrative Biology, 2010. **14**(5): p. 525-540.
424. Verduyn, C., et al., *Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation.* Yeast, 1992. **8**(7): p. 501-17.
425. Pampulha, M.E. and M.C. Loureiro-Dias, *Combined effect of acetic acid, pH and ethanol on intracellular pH of fermenting yeast.* Applied Microbiology and Biotechnology, 1989. **31**: p. 547-550.

426. Hu, C., et al., *Effects of biomass hydrolysis by-products on oleaginous yeast Rhodosporidium toruloides.* Bioresour Technology, 2009. **100**.
427. Zhao, X., et al., *Effects of some inhibitors on the growth and lipid accumulation of oleaginous yeast Rhodosporidium toruloides and preparation of biodiesel by enzymatic transesterification of the lipid.* Bioprocess Biosyst Eng, 2012. **35**(6): p. 993-1004.
428. Narendranath, N.V., K.C. Thomas, and W.M. Ingledew, *Effects of acetic acid and lactic acid on the growth of Saccharomyces cerevisiae in a minimal medium.* Journal of Industrial Microbiology and Biotechnology, 2001. **26**(3): p. 171-177.
429. Swinnen, S., et al., *The fraction of cells that resume growth after acetic acid addition is a strain-dependent parameter of acetic acid tolerance in Saccharomyces cerevisiae.* FEMS Yeast Research, 2014. **14**(4): p. 642-653.
430. Delgenes, J.P., R. Moletta, and J.M. Navarro, *Effects of lignocellulose degradation products on ethanol fermentations of glucose and xylose by Saccharomyces cerevisiae, Zymomonas mobilis, Pichia stipitis, and Candida shehatae.* Enzyme and Microbial Technology, 1996. **19**(3): p. 220-225.
431. Modig, T., G. Lidén, and M.J. Taherzadeh, *Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase.* Biochemical Journal, 2002. **363**(3): p. 769-776.
432. Banerjee, N., R. Bhatnagar, and L. Viswanathan, *Inhibition of glycolysis by furfural in Saccharomyces cerevisiae.* European journal of applied microbiology and biotechnology, 1981. **11**(4): p. 226-228.
433. Almeida, J.R., et al., *Metabolic effects of furaldehydes and impacts on biotechnological processes.* Appl Microbiol Biotechnol, 2009. **82**(4): p. 625-38.
434. Kim, D. and J.S. Hahn, *Roles of the Yap1 transcription factor and antioxidants in Saccharomyces cerevisiae's tolerance to furfural and 5-hydroxymethylfurfural, which function as thiol-reactive electrophiles generating oxidative stress.* Appl Environ Microbiol, 2013. **79**(16): p. 5069-77.
435. Konzock, O., S. Zaghen, and J. Norbeck, *Tolerance of Yarrowia lipolytica to inhibitors commonly found in lignocellulosic hydrolysates.* BMC Microbiol, 2021. **21**(1): p. 77.
436. Palmqvist, E., J.S. Almeida, and B. Hahn-Hägerdal, *Influence of furfural on anaerobic glycolytic kinetics of Saccharomyces cerevisiae in batch culture.* Biotechnol Bioeng, 1999. **62**(4): p. 447-54.
437. Taherzadeh, M.J., et al., *Physiological effects of 5-hydroxymethylfurfural on Saccharomyces cerevisiae.* Appl Microbiol Biotechnol, 2000. **53**(6): p. 701-8.
438. Taherzadeh, M.J., et al., *Conversion of furfural in aerobic and anaerobic batch fermentation of glucose by Saccharomyces cerevisiae.* J Biosci Bioeng, 1999. **87**(2): p. 169-74.
439. Larsson, S., et al., *The generation of fermentation inhibitors during dilute acid hydrolysis of softwood.* Enzyme and Microbial Technology, 1999. **24**(3): p. 151-159.
440. Almeida, J.R., et al., *Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae.* Journal of Chemical Technology & Biotechnology, 2007. **82**(4): p. 340-349.
441. Oliva, J.M., et al., *Effects of acetic acid, furfural and catechol combinations on ethanol fermentation of Kluyveromyces marxianus.* Process Biochemistry, 2006. **41**(5): p. 1223-1228.
442. Ding, M.-Z., et al., *Metabolomic Study of Interactive Effects of Phenol, Furfural, and Acetic Acid on Saccharomyces cerevisiae.* OMICS: A Journal of Integrative Biology, 2011. **15**(10): p. 647-653.
443. Wen, Z., et al., *Rhodosporidium toruloides - A potential red yeast chassis for lipids and beyond.* FEMS Yeast Res, 2020. **20**(5).
444. Kong, S., et al., *Expanding the neutral sites for integrated gene expression in Saccharomyces cerevisiae.* FEMS Microbiology Letters, 2022. **369**(1).
445. Otoupal, P.B., et al., *Multiplexed CRISPR-Cas9-based genome editing of Rhodosporidium toruloides.* mSphere, 2019. **4:e00099-19**.

446. Jiao, X., et al., *Developing a CRISPR/Cas9 System for Genome Editing in the Basidiomycetous Yeast Rhodosporidium toruloides.* Biotechnol J, 2019. **14**(7): p. e1900036.
447. Schultz, J.C., M. Cao, and H. Zhao, *Development of a CRISPR/Cas9 system for high efficiency multiplexed gene deletion in Rhodosporidium toruloides.* Biotechnol Bioeng, 2019. **116**(8): p. 2103-2109.
448. Tiukova, I.A., et al., *Genome-scale model of Rhodotorula toruloides metabolism.* Biotechnology and Bioengineering, 2019. **116**: p. 3396– 3408.
449. Dinh, H.V., et al., *A comprehensive genome-scale model for Rhodosporidium toruloides IFO0880 accounting for functional genomics and phenotypic data.* Metab Eng Commun, 2019. **9**: p. e00101.
450. Liebal, U.W., et al., *Ustilago maydis Metabolic Characterization and Growth Quantification with a Genome-Scale Metabolic Model.* J Fungi (Basel), 2022. **8**(5).
451. Zhang, C. and Q. Hua, *Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine.* Frontiers in Physiology, 2016. **6**.
452. Wick, R.R., L.M. Judd, and K.E. Holt, *Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing.* PLOS Computational Biology, 2023. **19**(3): p. e1010905.
453. Shen, J., et al., *Development of a novel, robust and cost-efficient process for valorizing dairy waste exemplified by ethanol production.* Microbial Cell Factories, 2019. **18**(1): p. 51.
454. Suwannarat, J. and R.J. Ritchie, *Anaerobic digestion of food waste using yeast.* Waste Management, 2015. **42**: p. 61-66.
455. Brandt, B.A., et al., *Adaptation of Saccharomyces cerevisiae in a concentrated spent sulphite liquor waste stream for increased inhibitor resistance.* Applied Microbiology and Biotechnology, 2022. **106**(1): p. 455-468.