S L U

Insights into genomics and gene expression of oleaginous yeasts of the genus *Rhodotorula*

GISELLE C. MARTÍN-HERNÁNDEZ

Insights into genomics and gene expression of oleaginous yeasts of the genus *Rhodotorula*

Giselle C. Martín-Hernández

Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Sciences
Uppsala

SLU

SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS

Uppsala 2023

Acta Universitatis Agriculturae Sueciae

2023:46

Insights into genomics and gene expression of oleaginous yeasts of the genus *Rhodotorula*

Abstract

The *Rhodotorula glutinis* cluster includes oleaginous yeast species with high biotechnological potential as cell factories for lipid, carotenoid, and oleochemical production. They can grow on various low-cost residues, such as lignocellulose and crude glycerol. The accumulated microbial lipids have a fatty acid composition similar to vegetable oils, representing a sustainable alternative in food, feed, and biofuel industries. Strain manipulation could further increase its economic attractiveness. A basic prerequisite for this is to understand the genome organization of *Rhodotorula* spp. and the lipid and carotenoid metabolism on residual products in detail. This doctoral thesis made a substantial contribution to this:

Hybrid genome assemblies from *R. toruloides*, *R. glutinis*, and *R. babjevae* were built *de novo* using a combination of short- and long-read sequencing. By doing this, high-quality genomes could be achieved, showing high completeness and contiguity (near-chromosome level). Among other things, this enabled the prediction of ploidy and the discovery of extrachromosomal structures. Phylogenetic analyses revealed high intraspecific divergence in *R. babjevae*. Using RNA-seq data generated from different growth conditions, gene annotation for *R. toruloides* could be significantly increased, and comprehensive gene expression data could be obtained. These allowed conclusions to be drawn about the activation and mechanism of lipid accumulation in *R. toruloides* grown on residual products such as crude glycerol and hemicellulosic hydrolysate.

Keywords: *Rhodotorula*, Microbial lipids, Genomics, *de novo* hybrid assembly, Annotation, Genome divergence, Ploidy, Transcriptomics, Lignocellulose, Glycerol

Insikter i genomik och genuttryck av lipdackumulerande jäst av släktet Rhodotorula

Sammanfattning

Rhodotorula glutinis-klustret inkluderar lipidackumulerande jästarter med hög bioteknologisk potential som cellfabriker för lipid-, karotenoid- och oleokemisk produktion. De kan växa på olika restprodukter, såsom lignocellulosa och råglycerol. De ackumulerade lipiderna har en fettsyrasammansättning som liknar vegetabiliska oljor, vilket representerar ett mer hållbart alternativ inom livsmedels-, foder- och biobränsleindustrin. Stammanipulation kan ytterligare öka dess ekonomiska attraktionskraft. En grundläggande förutsättning för detta är att förstå genomorganisationen hos *Rhodotorula* spp. och lipid- och karotenoidmetabolismen på restprodukter i detalj. Denna doktorsavhandling ger ett väsentligt bidrag till detta:

Hybridgenomsamlingar från *R. toruloides*, *R. glutinis* och *R. babjevae* byggdes *de novo* med en kombination av kort- och långläst sekvensering. Genom att göra detta kan kompletta och sammanhängande genom med hög kvalitet uppnås med en upplösning nära kromosomnivå. Det möjliggjorde bland annat förutsägelsen av ploidin och upptäckten av extrakromosomala strukturer. Fylogenetiska analyser avslöjade en hög intraspecifik divergens i *R. babjevae*. Genom att använda RNA-seq-data från olika tillväxtförhållanden kan antalet annoterade gener i denna art öka avsevärt och omfattande genuttrycksdata kan erhållas. Studien gjorde det möjligt att dra nya slutsatser om aktiveringen och mekanismen för lipidackumulering i *R. toruloides* odlad på restprodukter såsom råglycerol och hemicellulosahydrolysat.

Keywords: *Rhodotorula*, Mikrobiella lipider, Genomik, *de novo* hybridsammansättning, Annotation, Genomdivergens, Ploidy, Transcriptomics, Lignocellulosa, Glycerol

Dedication

To my family

"Ay. No hay que llorar

Que la vida es un carnaval

Y es más bello vivir cantando."

　　　Celia Cruz

Contents

List of publications

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

I. Martín-Hernández GC, Müller B, Chmielarz M, Brandt C, Hölzer M, Viehweger A, Passoth V (2021). Chromosome-level genome assembly and transcriptome-based annotation of the oleaginous yeast *Rhodotorula toruloides* CBS 14. Genomics, 113 (6), 4022-4027.

II. Martín-Hernádez GC, Müller B, Brandt C, Hölzer M, Viehweger A, Passoth V (2022). Chromosome-level genome assembly and annotation of *Rhodotorula babjevae* strains reveals high intraspecific divergence. Journal of Fungi, 8 (4), 323.

III. Martín-Hernández GC, Chmielarz M, Müller B, Hölzer M, Brandt C, Viehweger A, Passoth V (2023). Enhanced glycerol assimilation and lipid production in *Rhodotorula toruloides* CBS14 upon addition of hemicellulose primarily correlates with early transcription of energy-metabolism related genes. Biotechnology for Biofuels and Bioproducts, 16 (1), 42.

Papers I-III are reproduced with the permission of the publishers.

The contribution of Giselle C. Martín-Hernández to the papers included in this thesis was as follows:

I. Performed all laboratory work. Took part of formal analysis, investigation, visualization and writing—original draft preparation.

II. Took part in planning the project. Performed all laboratory work. Took part of formal analysis, investigation, visualization and writing—original draft preparation.

III. Took part in planning the project. Performed main parts of laboratory work. Took part of formal analysis, investigation, visualization and writing—original draft preparation.

The following paper was published during the timeframe of the doctoral education, but is not part of this thesis.

IV. Passoth V, Brandenburg J, Chmielarz M, Martín-Hernández GC, Nagaraj Y, Müller B, Blomqvist J (2023). Oleaginous yeasts for biochemicals, biofuels and food from lignocellulose-hydrolysate and crude glycerol. Yeast, 1-13, https://doi.org/10.1002/yea.3838

List of tables

List of figures

Abbreviations

ACL	ATP-citrate lyase
AMP	Adenosine monophosphate
ANI	Average Nucleotide Identity
ARS	Autonomously replicating sequence
ATMT	*Agrobacterium*-mediated transformation
BUSCO	Benchmarking of Universal Single-Copy Orthologs
CG	Crude glycerol
CGHH	Crude glycerol and hemicellulosic hydrolysate mixture
DDH	DNA–DNA homology
FA	Fatty acid
FAS	Fatty acid synthase
GDH	Glycerol dehydrogenase
GEM	Genome-scale model
HH	Hemicellulosic hydrolysate
IDH	Isocitrate dehydrogenase
IMP	Inosine monophosphate
ITS	Internal transcribed spacer
D1/D2 LSU	D1/D2 region of the large ribosomal subunit
ME	Malic enzyme

NGS	Next-generation sequencing
NHEJ	Nonhomologous end joining
ONT	Oxford Nanopore technologies
ORF	Open reading frames
PacBio	Pacific Biosciences
PCS	Protein-coding sequences
PFGE	Pulsed-field gel electrophoresis
PPP	Pentose-phosphate pathway
SNPs	Single nucleotide polymorphisms
TAG	Triacylglycerides
TCA	Tricarboxylic acid cycle
VO	Vegetable oils
WGD	Whole-genome duplication
YP	Medium containing yeast extract and peptone
YPG	Medium containing yeast extract, peptone and glucose

1. Background and aims

Vegetable oils (VO) are important food commodities with additional applications as lipidic feedstocks for biodiesel production. Biodiesel represents an opportunity for ensuring energy security, reducing the carbon footprint from the energy sector, aiding climate change mitigation, and improving air quality compared to fossil fuels (Fawzy et al., 2020; Halkos & Gkampoura, 2020). However, the dominant VO used as feedstock for biodiesel production are rapeseed-, palm-, soybean- and sunflower oil in the EU (Flach B., 2022). Besides rainforest clearance for creating farmland, some other issues that arise with using edible oils as a biodiesel feedstock are food supply impairment and unstable availability due to market competition (Bastos Lima, 2021; Hoang & Kanemoto, 2021; Schmidt, 2015). Extensive cultivation of VO also has a negative climate impact, including higher greenhouse gas emissions, freshwater consumption, and loss of biodiversity (Alcock et al., 2022; Bastos Lima, 2021; Koh & Wilcove, 2008).

Microbial lipids can be considered a sustainable alternative to oil-bearing crops as a lipid biofuel raw material (Carmona-Cabello et al., 2021). Besides biodiesel, microbial lipids have applications as feedstock for the production of oleochemicals and feed and food additives. For instance, microbial lipids are potential candidates for VO or fish oil replacement in fish feed formulations (Blomqvist et al., 2018; Brunel et al., 2022). Microbial biomass can also be used as an alternative protein source to fishmeal (Vidakovic et al., 2020). Exploiting such resources can alleviate food market competition while increasing sustainability and reducing the carbon footprint of aquaculture. Ultimately, the study of alternative food and feed ingredients from sustainable sources is essential for ensuring food and nutrition security (El Bilali et al., 2019).

Oleaginous microorganisms are those able to accumulate lipids in a proportion higher than 20% of their biomass weight (Ratledge & Wynn, 2002). They are widely spread in nature, including prokaryote and eukaryote species among bacteria, algae and fungi. Besides their ability for lipid accumulation, many oleaginous microorganisms have rapid growth rates and the capacity to grow on low-cost substrates (Liao et al., 2016; Passoth et al., 2023; Patel A, 2020). Cultivation can be done all year round without requiring farmland or competing with the food market. Furthermore, the choice of strain and culture conditions can vary microbial lipid content and thus aid in addressing specific demands from industry (Alvarez et al., 1997; Brandenburg et al., 2021; Menegazzo & Fonseca, 2019). Oleaginous microorganisms with lipid composition similar to those derived from VO are considered promising raw materials for producing biodiesel and food and feed additives (Brunel et al., 2022; Zhu et al., 2008).

A prominent example of such oleaginous microorganisms is *Rhodotorula* yeasts. Besides lipids, they are naturally able to accumulate carotenoids, glycolipids and exopolysaccharides of industrial interest (Garay et al., 2017; Zhao & Li, 2022; Zhao et al., 2021). *Rhodotorula* spp. have also the advantage compared to other oleaginous microorganisms of growing on low-cost residual products (Chmielarz et al., 2021; M. Liu et al., 2016). The residues after lipid production could close the loop towards a circular bioeconomy through their valorisation into agricultural fertilizers or as feedstock for biogas production to obtain energy (Karlsson et al., 2016, 2017). The application of a circular economy concept for *Rhodotorula* lipid production could increase its economic attractiveness and reduce natural resource consumption and the environmental impact of processes (Figure 1) (Grdic et al., 2020).
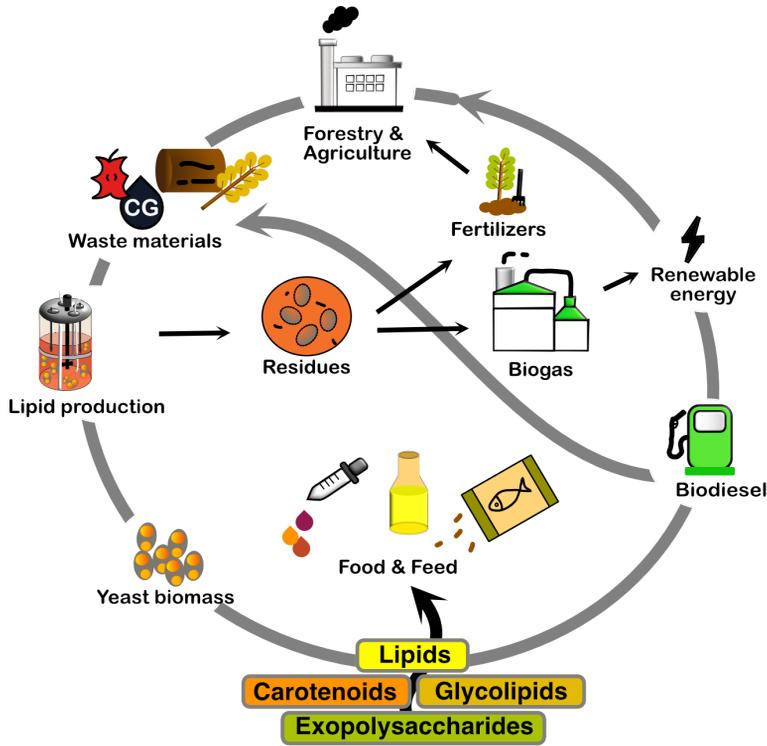
Figure 1. Biorefinery approach of *Rhodotorula* lipids from low-cost residual substrates.

Genetic engineering approaches represent an additional tool for improving the performance of the bioprocess and establishing *Rhodotorula* yeasts as versatile cell factories (Abeln & Chuck, 2021). For instance, strains could be developed towards having higher lipid accumulation rates, which is desirable for enhancing the economic feasibility of the bioprocess (Karlsson et al., 2016). Despite the advances in expanding the *R. toruloides* genetic engineering toolkit (Bonturi et al., 2022; Zhao et al., 2021), there is room for improvement and expansion to less studied but industrially attractive yeast species such as *R. babjevae* and *R. glutinis*. The availability of accurate genome sequences, metabolic networks, and knowledge of genome organization and ploidy could aid in tailoring engineering approaches to promising strains (Fidan & Zhan, 2015; Schindler, 2020; Zhang et al., 2014). For producing and improving such resources, which are still lacking for most *Rhodotorula* spp., genomics and transcriptomics studies on *Rhodotorula* yeasts can create valuable data. They are also needed to deepen

our understanding of the physiology of lipid production in *Rhodotorula* spp. and the biology and evolution of non-conventional yeast species. Thus, the focus of the thesis work was on performing genomics and gene expression studies in *R. toruloides*, *R. babjevae* and *R. glutinis* to understand their genome organization, evolution and physiology of lipid production in different carbon sources.

1.1 Oleaginous yeasts

Oleaginous yeasts and some filamentous fungi are the most promising hosts for the production of single-cell oil due to their ability to grow with higher tolerance to contaminants and at fast growth rates on low-cost carbon sources (Passoth et al., 2023; Venkata Subhash & Venkata Mohan, 2011; Zininga et al., 2019). For instance, they can be cultivated on organic residues from agriculture, forestry or the food industry as substrate (Abeln & Chuck, 2021; Passoth & Sandgren, 2019; Patel A, 2020; Sineli et al., 2022). Robust strains according to their industrial applications, are those naturally able to co-metabolise sugars derived from lignocellulosic hydrolysates and grow on aromatic compounds derived from lignin with high inhibitor tolerance. Some examples are *Lipomyces starkeyi*, *Trichosporon cutaneum*, *R. toruloides* and *R. babjevae* (Abeln & Chuck, 2021; Brandenburg et al., 2021; Chmielarz et al., 2021; Juanssilfero et al., 2018; J. Wang et al., 2016). *Rhodotorula* species and *Yarrowia lipolytica* are also attractive due to their capacity to efficiently convert crude glycerol (CG), a biodiesel-derived residue, into added-value metabolites (Abeln & Chuck, 2021; Chmielarz et al., 2021; Filippousi et al., 2019; Rakicka et al., 2015). In addition, yeast lipids have a favourable fatty acid (FA) composition on account of their similarity with VO (Blomqvist et al., 2018; Gong et al., 2016; Patel et al., 2016).

Other oleaginous microorganisms are found among bacteria and algae. Bacterial oleaginous species frequently have low yields of neutral lipids, polyhydroxyalkanoic acids its main class (Alvarez & Steinbüchel, 2002; Alvarez et al., 1997; Jones et al., 2019). Bacterial lipids can have promising industrial applications as nutraceuticals rather than as a VO replacement (Alvarez & Steinbüchel, 2002; Kaneda, 1991). On the other hand, biorefineries based on oleaginous microalgae have the asset of combining the production of biofuels and high-value nutrients with wastewater treatment (Shayesteh et al., 2022; Vadiveloo et al., 2021; Zhu et al., 2013).

However, large-scale production of microalgal biomass is still challenging due to high costs, harvesting methods, a high risk of biological contamination and demanding cultivation conditions, such as nutrient supply and geographical variations (Menegazzo & Fonseca, 2019; Shayesteh et al., 2022; Zhu et al., 2013). Hence, the suitability of oleaginous bacteria and microalgae lipids as a replacement for VO is rather low compared to yeasts.

1.2 Potential biotechnological applications of *Rhodotorula* yeasts

Rhodotorula yeasts can grow on various low-cost residual products such as lignocellulosic hydrolysates from the pulp- and paper industry and CG (Brandenburg et al., 2016; Chmielarz et al., 2021; M. Liu et al., 2016). In complex media, *Rhodotorula* spp. can synthesize a series of products of industrial interest such as sesquiterpenes, triacylglycerides (TAG), FA ethyl esters, fatty alcohols, D-arabitol, biotechnologically important enzymes, carotenoids, glycolipids and exopolysaccharides (Ayadi et al., 2018; Boviatsi et al., 2020; Geiselman et al., 2020; Guerfali et al., 2022; Guerfali et al., 2019; Liu et al., 2020; M. Liu et al., 2016; Mirzaei Seveiri et al., 2020; Rodriguez et al., 2019). Hence, *Rhodotorula* yeasts could be used as versatile microbial cell factories (Zhao et al., 2021).

While free FAs are present in a substantial proportion, TAG is the most common lipid class by the end of the cultivation (i.e., when carbon sources are consumed) (Nagaraj et al., 2022; Shapaval et al., 2019). FA composition varies between yeast strains (Shapaval et al., 2019). In *R. toruloides* CBS 14, the FA profile mainly includes oleic (C18:1), linoleic (C18:2), palmitic (C16:0), alpha-linolenic (C18:3), stearic (C18:0) and heptadecenoic (C17:1) acids (Nagaraj et al., 2022). Hence, the similarity of FA composition to VO supports their suitability for biodiesel and food and feed applications (Nagaraj et al., 2022; Sineli et al., 2022). *R. toruloides* has been shown to have a higher content of polyunsaturated FAs than *L. starkeyi* strains (Brandenburg et al., 2021).

R. toruloides can produce carotenoids such as torulene, torularhodin, β-carotene and γ-carotene in high numbers (Nagaraj et al., 2022; Pinheiro et al., 2020). Carotenoids have high antioxidant activity with promising applications in the pharmaceutical, food and feed industries (Guerfali et al., 2022; Rapoport et al., 2021). Other therapeutic properties include immune

modulation and prevention of cardiovascular disease and cancer (Rapoport et al., 2021). Carotenoids can also act as a preserving agent in fish feed formulations. *R. toruloides* biomass has been assessed as part of Arctic char feed pellets resulting in no differences in the fat content of muscle and liver tissues compared to fish under normal diets (Brunel et al., 2022). Moreover, carotenoids from *Rhodotorula* yeasts have an industrial interest as natural colourants with the environmental prospect of replacing chemically synthesized carotenoids (Rapoport et al., 2021).

R. babjevae can synthesize and secrete glycolipid biosurfactants such as sophorolipids and polyol lipids (Garay et al., 2017; Sen et al., 2017). These glycolipids have shown promising properties for various industrial sectors, including cosmetics, biodegradation of hydrocarbon pollutants as well as food and pharmaceutics due to their antiviral, antibacterial, antifungal, and anti-carcinogenic activities (Guerfali et al., 2022; Guerfali et al., 2019; Sen et al., 2017; Sen et al., 2020). Microbial biosurfactants have advantages over chemically produced surfactants because they are renewable, biodegradable and sustainably produced with low ecotoxicity (Marchant & Banat, 2012).

Furthermore, *R. glutinis* and *R. babjevae* have been shown to synthesize and secrete exopolysaccharides (Mirzaei Seveiri et al., 2020; Zhao & Li, 2022) that have application prospects in the pharmaceutical and food industries due to their physicochemical and health-promoting properties, such as emulsifying, antioxidant, anti-tumour and immune-modulating (Mirzaei Seveiri et al., 2020; Zhao & Li, 2022).

1.3   *Rhodotorula glutinis* cluster

There are 207 *Rhodotorula* species according to the latest records of Mycobank (www.mycobank.org, accessed on March 28th 2023), including *R. toruloides*, *R. glutinis*, *R. babjevae*, *R. mucilaginosa* and *R. graminis*. They are basidiomycetes oleaginous yeast species from class Microbotryomycetes, order Sporidiobolales and family Sporidiobolaceae. Due to the accumulation of carotenoids, these yeast species are pink-coloured and thus referred to as red yeasts (Figure 2). They have been isolated from various environmental sources ranging from seawater and soil to plants and the atmosphere. Interestingly, *R. glutinis* was first identified as causing infections in humans and thus contemplated as pathogenic (Morrow & Fraser, 2009; Tuon & Costa, 2008).
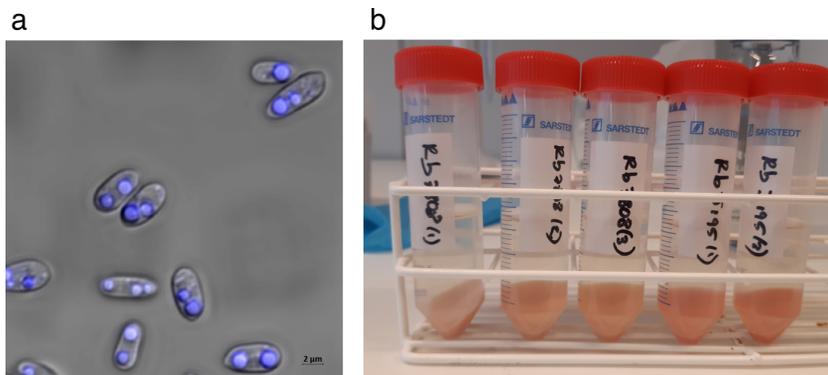
Figure 2. *Rhodotorula* cells. (a) Confocal microscopy of *R. toruloides* strain CBS 14. The lipid bodies (blue) were dyed using Nile red (paper **I**). (b) Cell pellet from *R. babjevae* after 24 h cultivation on YPD (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose).

Rhodotorula spp. were previously considered to be anamorphic, having only asexual haploid yeast form and no transition to filament forms (Morrow & Fraser, 2009). In contrast with *Rhodotorula*, strains of *Rhodosporidium* were considered to be those in a teleomorph state, with bipolar mating behaviour (Coelho et al., 2008). *R. toruloides* have been described as both, *Rhodotorula* and *Rhodosporidium*. In 2011 the scientific community agreed on following the principle of "one fungus = one name" for genus nomenclature, delineating *Rhodotorula* as the genus name for the group (Hawksworth et al., 2011; Wang et al., 2015).

 R. toruloides type strain is CBS 6016, IFO8766 in the IFO collection. It is a self-fertile hybrid, which resulted after the conjugation of CBS 14 (IFO0559, ATCC 10788, mating type A1) and CBS 349 (IFO0880, ATCC 10657, mating type A2) by Banno in 1967 (Banno, 1967; Sampaio, 2011a; Zhao et al., 2021). CBS 7808 is the type strain from *R. babjevae*, with mating type A1. It was isolated in 1970 by I.P. Babjevae in Moscow region, Russia (Sampaio, 2011a). Lastly, the *R. glutinis* type strain is CBS 20 (IFO1125).

1.4   Efforts and tools for improving lipid accumulation

1.4.1   Genetic manipulation

Besides the natural robustness of *Rhodotorula* yeasts, their biotechnological importance has boosted the development of engineering approaches aiming

at bioproduction optimization (Zhao et al., 2021). Some studies have focused on optimizing the cultivation conditions (Bonturi et al., 2017; Diamantopoulou et al., 2020; Guerfali et al., 2022; Lopes et al., 2020; Mussagy et al., 2021; Pinheiro et al., 2020; Zhang et al., 2020; Zhao & Li, 2022) and others in improving the yeast strains through genetic engineering or strain adaptation (Bonturi et al., 2017; Lin et al., 2014; Pinheiro et al., 2020; Zhang, Skerker, et al., 2016). Some relevant goals for genetic modification are to vary microbial lipid composition, increase lipid productivity and improve the tolerance to inhibitors present in culture media (Zhao et al., 2021).

There are far fewer genetic engineering tools and background knowledge for the non-conventional *Rhodotorula* yeasts than in the model yeast species *Y. lipolytica* and *Saccharomyces cerevisiae* (Zhao et al., 2021). Besides, the autonomously replicating sequence (ARS) elements have yet to be identified in *Rhodotorula* spp. (Bonturi et al., 2022). From *Rhodotorula* spp., *R. toruloides* is the most widely studied and, consequently, the one with higher availability of resources (Abeln & Chuck, 2021).

In 1985, a spheroplast transformation method was developed for exogenous DNA integration into *R. toruloides* (Tully & Gilbert, 1985). A more practical manipulation method was established in 2014, consisting of *Agrobacterium*-mediated transformation (ATMT) (Lin et al., 2014). Other approaches for DNA random integration centred on the nonhomologous end joining (NHEJ) pathway include lithium acetate/PEG-mediated chemical transformation and integration of linear sequences through electroporation (Coradetti et al., 2018; Liu et al., 2017; Tsai et al., 2017). Targeted gene deletion/integration recently emerged for the species through CRISPR/Cas9 systems and ATMT in KU70-deficient mutants (Koh et al., 2014; Otoupal et al., 2019; Schultz et al., 2019; Zhang, Ito, et al., 2016). In addition, (Liu et al., 2019) achieved downregulation of targeted gene expression through an RNA interference approach. A golden gate assembly platform was also established in the species (Bonturi et al., 2022).

Some examples of genetic elements developed for engineering *R. toruloides* are constitutive- (*GPD1*, *PGI1*, *PGK1*, *FBA1* and *TPI1*) (Liu et al., 2013; Y. Wang et al., 2016), intronic- (*LDP1*, *ACC1* and *FAS1*) (Y. Liu et al., 2016) and inducible promoters (*NAR1*, *ICL1*, *CTR3*, and *MET16*) (Johns et al., 2016), and selectable markers (Lin et al., 2014). Co-expression of multiple genes using a single promoter has also been established, further

expanding the genetic engineering toolkit (Jiao et al., 2018). Furthermore, some examples of genes whose overexpression has increased lipid production are *ME*, *SCD*, *ACC1* and *DGA1* (Díaz et al., 2018; Zhang, Ito, et al., 2016; Zhang, Skerker, et al., 2016).

Despite these advances, there is still a need to continue expanding the genetic engineering toolkit and developing manipulation methods that achieve higher transformation efficiencies in *Rhodotorula* yeasts, especially in less studied species such as *R. babjevae* and *R. glutinis*. The availability of genome sequences with accurate annotation and knowledge of their genome organization and gene expression regulation can be a valuable aid (Fidan & Zhan, 2015). Generating and combining genomics and transcriptomics data can be useful for building such robust genome assemblies.

1.4.2 Genomics and transcriptomics of *Rhodotorula* spp.

Genomics unravels the genetic information within genomes by applying whole-genome sequencing technologies. It is a powerful tool for expanding the availability of reference genomes and characterizing biodiversity (Formenti et al., 2022). Genome assemblies of high quality (high contiguity, accuracy and annotation) represent good reference genomes in a species. They can provide a comprehensive framework for characterizing genome structure and organization (Figure 3) as well as studying genomic variation and evolution between multiple conspecific individuals (Formenti et al., 2022). *R. toruloides* gathers most of the genomics studies in the *R. glutinis* cluster. There are some genome sequences available from species such as *R. glutinis* and *R. graminis* (Firrincieli et al., 2015; Li et al., 2020), while others like *R. babjevae* lack such data. In addition, most available genome sequences were built using short-read sequencing technologies, often lacking contiguity and completeness and incomplete annotation.
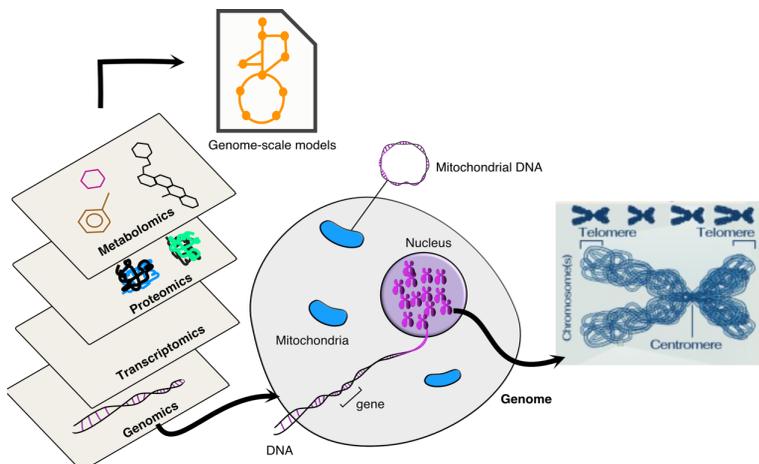
Figure 3. Reference genomes providing insights into genome organization in a species. Figure modified from Formenti et al. (2022).

Transcriptomics is the study of all RNA molecules in a cell or organism. For instance, RNA-Seq is a sequencing technology that quantifies transcript levels giving information on gene activity under different environmental conditions (Jagtap et al., 2021). The retrieved transcriptome provides insights into gene expression and regulation and can guide and improve gene annotation (Zhu et al., 2012). Most transcriptomic studies in the *R. glutinis* cluster have focused on *R. toruloides*. Thus, genomics and transcriptomics studies are needed in the *R. glutinis* cluster to achieve high-quality genome assemblies with accurate gene annotation.

1.4.3 Understanding mechanisms of lipid accumulation

In general, a combination of omics studies (i.e., genomics, transcriptomics, proteomics, and metabolomics) can be used to understand the mechanisms of lipid accumulation and substrate utilization in *Rhodotorula* yeasts. They provide a holistic view of cells´ growth and adaptation to changing environments (Figure 3) (Pinu et al., 2019). A variety of -omics studies have been performed on species from the *R. glutinis* cluster (Bo et al., 2022; Coradetti et al., 2018; Gong et al., 2019; Jagtap et al., 2021; Li et al., 2020; Nora et al., 2023; Tiukova, Brandenburg, et al., 2019; Zhao & Li, 2022; Zhu et al., 2012). Most of these studies have focused on *R. toruloides* and, to a smaller extent, on *R. glutinis*.

In addition, comprehensive metabolic networks have been constructed in *R. toruloides* integrating multi-omics studies and systems biology methods (Bommareddy et al., 2015; Castañeda et al., 2018; Dinh et al., 2019; Kim et al., 2021; Tiukova, Prigent, et al., 2019; Wang et al., 2018). For instance, the genome-scale model (GEM) generated by Tiukova et al. (2019) includes 852 genes, 2,731 reactions, and 2,277 metabolites in *R. toruloides* NP11, with growth predictions on glucose, xylose and glycerol (Tiukova, Prigent, et al., 2019). Figure 4 summarizes the lipid metabolism in the species based on the available metabolic models. GEMs have created the bases towards understanding the molecular mechanisms behind lipid accumulation in red oleaginous yeast, besides providing potential targets for genetic engineering (Castañeda et al., 2018; Patra et al., 2021). However, as research interest in red oleaginous yeast is continuously growing, there is a need for the existing GEMs to be improved and extended to the other species in the *R. glutinis* cluster through new data and knowledge.
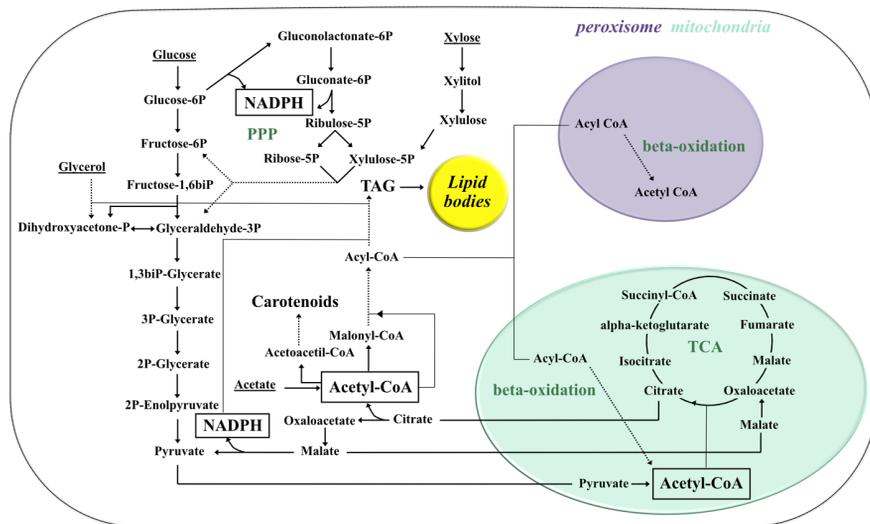
Figure 4. Simplified map of lipid metabolism in *Rhodotorula toruloides* and growth prediction on glucose, xylose and glycerol. Dash lines represent multiple reactions.

1.5 Aims

The available knowledge in the genome organization of strains from the *R. glutinis* cluster is still limited, especially within the closely related species *R.*

babjevae and *R. glutinis*. That hinders their molecular manipulation and analysis. Furthermore, there is a gap in understanding the molecular physiology of the above-mentioned yeasts when converting residual products such as CG into lipids, carotenoids and biosurfactants, which have high industrial attractiveness.

Aims:

1. Optimization and establishment of methods for genomic DNA isolation, Nanopore sequencing and genome assembly of strains of the *Rhodotorula glutinis* cluster (paper **I**, paper **II**).
2. Genome analysis of the yeast species *Rhodotorula babjevae* and comparison to other *Rhodotorula* species (paper **II**).
3. Gaining insights into metabolic pathways of *R. toruloides* activated on different carbon sources (paper **III**).

2. Genome assemblies in *Rhodotorula* spp*.*

Whole-genome sequencing and assembly methods have generated an increasing availability of reference genomes (Figure 3), revolutionizing the study of yeast biology and evolution. Competing sequencing technologies have been established with enhanced accessibility and high quality and robustness of the generated genome assemblies (Ding et al., 2023; Hu et al., 2021; Lang et al., 2020; Willenbücher et al., 2022). Year by year, these sequencing technologies are updated with improved chemistries in terms of higher throughput, sequence quality, read length, and applications (Hu et al., 2021; Pervez et al., 2022). Thus, DNA sequencing methods and bioinformatics approaches have experienced a dramatic transformation from only 50 years back.

2.1   DNA sequencing methods

First-generation sequencing technologies emerged in the 1970s, including the widely used Sanger sequencing (Sanger et al., 1977). This reliable and reproducible method was published nearly 25 years after the DNA structure was suggested by Watson and Crick (Watson & Crick, 1953). Sanger sequencing has been applied for sequencing genomes from bacteria and phages to humans (Levy et al., 2007; Sanger et al., 1977). The budding yeast was the first eukaryotic organism to be completely sequenced (Goffeau et al., 1996). The human genome project was also executed using first-generation sequencing, yielding the human "draft" sequence data in an international effort for understanding its structure and function (Collins et al., 1998; Gibbs, 2020). However, first-generation technologies have become obsolete for large genomes due to their high cost and labour-intensive, time-consuming, and low-efficiency procedure (Hu et al., 2021).

Next-generation sequencing (NGS) technologies were introduced between 2004 and 2006. They resulted in a huge increase in data output compared with Sanger-sequencing (Bentley et al., 2008; Hu et al., 2021). With the innovative nanotechnology principles from NGS, single DNA molecules are massively sequenced in parallel. Other advantages include less DNA input, faster processing time, and lower costs (Pervez et al., 2022). NGS technologies can be grouped as second or third-generation.

Second-generation platforms (e.g., Illumina) commonly start with a step of DNA fragmentation and thus are considered "short-read" technologies. The collected sequencing data presents challenges such as *de novo* assembly of low-complexity genome regions, identification of transcript isoforms, and errors related to DNA amplification (Hu et al., 2021). Despite that, Illumina sequencing technology has become a quick and non-expensive tool of choice for a wide range of applications, such as the identification of structural variations, whole-genome resequencing, and haplotype phasing (Bentley et al., 2008; Duan et al., 2022; Kajiya-Kanegae et al., 2021; Pervez et al., 2022; Xu et al., 2019).

Third-generation sequencing technologies (e.g., Oxford Nanopore technologies, ONT and Pacific Biosciences, PacBio) are referred to as "long-read" methods because of their ability to generate individual reads up to 100 kb or even higher (Brancaccio et al., 2021; Lang et al., 2020). These technologies can overcome the described issues encountered in short-read data while having the constraint of needing a high molecular weight DNA (Brandt et al., 2020). Another drawback of these technologies compared to second-generation methods is high error rates (Hu et al., 2021). However, new advances in sequencing chemistry and computational tools are continuously arising to reach higher read accuracies. In addition, a combination of short- and long-read technologies in a hybrid sequencing approach has also been used to overcome high error rates (Hu et al., 2021; Luo et al., 2023; paper **I**; paper **II**). The two cutting-edge sequencing platforms, ONT and PacBio, are broadly used, enabling the production of new and better-quality reference genomes from numerous species (Brancaccio et al., 2021; Brandt et al., 2020; Kingan et al., 2019; Zhang et al., 2022). Furthermore, ONT instruments (Figure 5) are portable, significantly smaller, and low-cost compared to PacBio, which is desirable for applications in low-income settings or field sequencing (Hu et al., 2021).
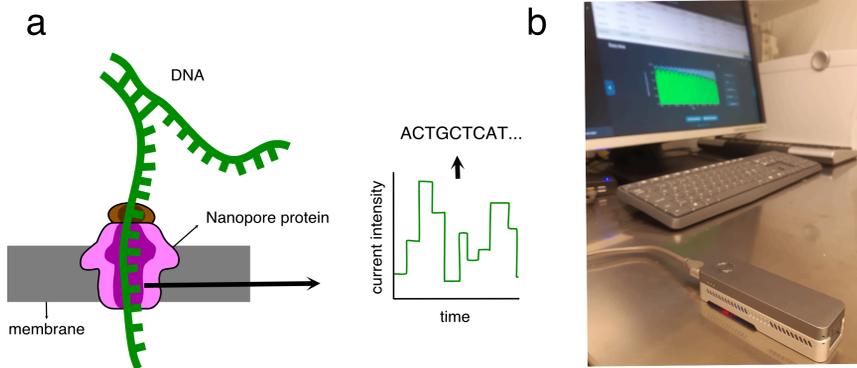
Figure 5. Oxford Nanopore Technology. (a) Sequencing overview. (b) MinION.

Each of the available sequencing technologies has advantages and disadvantages. The choice between short- and long-reads, sequencing devices, or a combination of them depends on the research question and resources. Hybrid genome assembly approaches, which combine short- and long-read sequencing data have shown improved accuracy and contiguity in yeast *de novo* genome assemblies (Olsen et al., 2015; Tiukova, Pettersson, et al., 2019; paper **I**; paper **II**).

2.2   Genome assemblies in the *Rhodotorula glutinis* cluster

There is an increasing availability of draft genome assemblies in the *R. glutinis* cluster (Table 1). *R. toruloides* have the highest representation in sequencing studies, where draft genomes have been built for different strains (Hu & Ji, 2016; Kumar et al., 2012; Morin et al., 2014; Paul et al., 2014; Sambles et al., 2017; Tran et al., 2019; Zhang, Skerker, et al., 2016; Zhu et al., 2012; paper **I**). However, most of them have been determined using only short-read sequencing technologies, except for paper **I**. The reviewed *R. toruloides* assemblies contain an average GC content of 61.9%. Their average genome size is 20.5 Mbp, excluding the strain CGMCC 2.1609, which genome is much larger (Table 1). In the study by Sambles et al. (2017), the genome from CGMCC 2.1609 was sequenced using short-read technologies and a low read depth (13x). A typical yeast genomics study will produce a genome assembly with about 50x coverage or higher for high-accuracy reads (Douglass et al., 2019). Even though the genome organization

of this strain can be different and consequently affect genome size, further studies are needed, including an updated genome assembly.

Table 1. Genome assembly statistics from strains belonging to *Rhodotorula toruloides*, *R. glutinis* and *R. babjevae*

Reference	Strain	Size (Mbp)	GC (%)	No. of contigs	No. of scaffolds	No. of PCS
(Kumar et al., 2012)	MTCC 457	20.1	62.0	689	644	5,993 c
(Zhu et al., 2012)	NP11	20.2	NA	17,814	34	8,171
(Morin et al., 2014)	CECT1137	20.5	61.9	NA	62	8,206
(Zhang, Skerker, et al., 2016)	IFO0559	20.3	NA	NA	246	8,100 c
	IFO0880	20.4	NA	NA	219	7,920 c
(Hu & Ji, 2016)	ATCC 10788	20.8	62.0	NA	61	7,730
	ATCC 10657	21.5	61.8	NA	137	7,800
(Sambles et al., 2017)	CGMCC 2.1609	33.4	61.9	868	365	9,820
(Tran et al., 2019)	VN1	20.0	61.8	424	NA	8,021
(Paper I)	CBS14	20.5	61.8	23	3	9,464
(Paul et al., 2014)	ATCC 204091	20.5	61.9	186	29	3,359 b
(Li et al., 2020)	ZHK	21.8	67.8	NA	30	6,774
(Bo et al., 2022)	X-20	23.0	68.2	20 a	NA	6,892 b
(Paper II)	CBS 20	22.9	67.5	32	0	7,607
	CBS 7808	21.9	68.2	24	3	7,591
	DBVPG 8058	21.5	68.2	33	1	7,481

Strain names that are represented in red, blue or green fonts belong to *R. toruloides*, *R. glutinis* and *R. babjevae*, respectively; NA, not available; PCS, protein-coding sequences; a, b and c refer to the number of contigs or scaffolds with size larger than 500 kb, putative genes and predicted proteins, respectively.

R. glutinis ATCC 204091 was thought to be the first published genome sequence for the species (Paul et al., 2014). However, the strain was later classified as *R. toruloides* due to its high sequence identity with strains such as IFO0880 (Liu et al., 2021; Paul et al., 2014; Zhang, Skerker, et al., 2016). Recently, whole-genome sequencing from *R. glutinis* strain ZHK was studied as part of a multi-omics approach (Li et al., 2020). Other studies that have published genomic analyses from *R. glutinis*, and similarly used a combination of short- and long-read sequencing data are (Bo et al., 2022) and paper **II**. The average GC content and size in the three *R. glutinis* strains are 67.8% and 22.6 Mbp, respectively. This high GC proportion, even higher than in *R. toruloides,* could represent a challenge for many sequencing technologies as well as for developing genetic engineering approaches

(Bonturi et al., 2022). That is because high GC levels can hinder gene synthesis, polymerase efficiency and accuracy in PCR amplification, and due to the necessity for codon-optimization of heterologous genes (Lin et al., 2014).

Two *de novo* genome assemblies from *R. babjevae* strains were first built in our group through a hybrid approach of short- and long-reads (paper **II**). Their average GC content is 68.2%, a higher GC value similar to *R. glutinis*. Their genome size is about 21.7 Mbp, likewise most other *Rhodotorula* spp. (Table 1).

In addition, several genome assemblies from other *Rhodotorula* species have been generated, though using mostly short-read technologies (Buedenbender et al., 2020; Daudu et al., 2020; Deligios et al., 2015; Fakankun et al., 2021; Su et al., 2019; Tkavc et al., 2018). One exception is the high-quality draft genome assembly from *R. mucilaginosa* JGTA-S1, which was built using a hybrid approach (Sen et al., 2019). It has a total size of 20.1 Mbp, 46 scaffolds, and 5,922 annotated protein-coding genes (Sen et al., 2019). The genome sequence from *R. graminis* WP1 was obtained using a Sanger whole genome shotgun approach (Firrincieli et al., 2015). The 21.0 Mbp genome is organized in 323 contigs and 28 scaffolds with an average GC content of 67.8%.

2.2.1 Hybrid genome assemblies and annotation

An example of a hybrid *de novo* genome assembly and annotation approach is illustrated in Figure 6 and represents the bioinformatic steps followed (paper **I**, paper **II**). Long ONT reads enabled the assembly of genomic DNA in a significantly lower number of contigs and scaffolds compared to other studies (Table 1). Mostly, these represent sequences resolved near the chromosome level (paper **I**, paper **II**). Thus, together with the high coverage achieved by the hybrid approach, *Rhodotorula* genome assemblies with higher contiguity, and completeness were generated. *R. mucilaginosa* JGTA-S1 hybrid assembly also showed improved contiguity compared to that for strains C2.5t1 (1,034 scaffolds) and RIT389 (250 contigs), which used only Illumina short reads (Deligios et al., 2015; Gan et al., 2017; Sen et al., 2019).
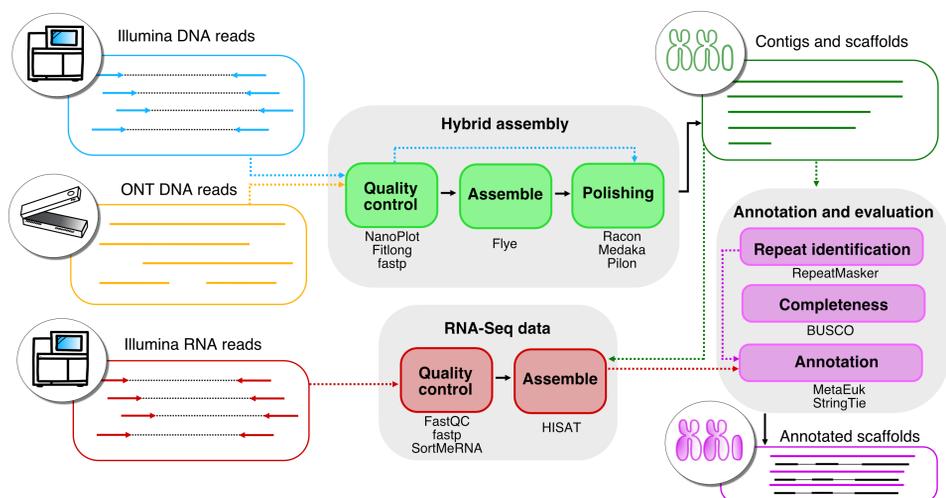
Figure 6. Simplified representation of the hybrid genome assembly workflow. Dash lines represent input data.

Moreover, *R. toruloides* genome annotation has been performed (Table 1). The highest number of annotated transcripts was achieved in *R. toruloides* strain CBS 14 (paper **I**). That could be related to the availability of RNA-Seq data from different growth conditions and the high contiguity of the *de novo* hybrid assembly. Furthermore, the annotation of protein-coding genes in *R. glutinis* and *R. babjevae* also had the highest described numbers in paper **II** for both species.

The number of complete BUSCO (Benchmarking of Universal Single-Copy Orthologs) genes is another indication of the completeness and quality of a genome assembly. Using the fungi_odb9 database, our approach identified 96.9%, 81.2%, 95.5%, and 96.9% complete BUSCO for the strains CBS 14, CBS 20, CBS 7808, and DBVPG 8058, respectively (paper **I**, paper **II**). When the number of complete BUSCO is closer to 100%, it implies a higher quality of the assembled genome. Some BUSCO genes might be undetectable because of assembly problems that would need to be addressed with additional sequencing and manual analysis. However, the studied strains may also not contain some of the missing BUSCO genes or had no significant matches with the selected BUSCO "fungi" profile, independently of the assembly completeness level (paper **II**). Some other hybrid approaches which reached high numbers of complete BUSCO in high-quality genome assemblies are those from *R. glutinis* strains ZHK and X-20 with values of 94.4% (Li et al., 2020) and 97.0% (Bo et al., 2022), respectively.

2.3 Mitochondrial genomes

Several mitochondrial sequences from *Rhodotorula* spp. have been recovered and deposited in public databases (Gan et al., 2017; Goordial et al., 2016; Li et al., 2020; Sen et al., 2019; Zhao et al., 2013; Zhou et al., 2020; paper **I**; paper **II**). They were identified to have a circular topology, similar to other Basidiomycota mitochondrial genomes (mitogenomes) (Gan et al., 2017; Zhao et al., 2013; Zhou et al., 2020; paper **I**; paper **II**). The length of available sequences ranges from 28.4 to 125.9 kb, with a 40.4 kb median. This relatively broad range could be a consequence of assembly artefacts because most assembly programs cannot clearly identify the ends of circular DNA sequences from their linear representation, and tandem repeats could be added or removed, affecting the predicted length (paper **I**). The presence of introns, spacer DNA, undetermined ORF (open reading frames), and integrated mitochondrial plasmids could also introduce variations in the mitochondrial genome size (Hao, 2022; Zhao et al., 2013).

Despite the differences in mitogenome length, the encoded genes are considered very stable in yeast (Hao, 2022). Some of these studies in *Rhodotorula* spp. included mitogenome annotation (Gan et al., 2017; Zhao et al., 2013; Zhou et al., 2020). In summary, the annotation approaches identified 15-20 core mitochondrial ORF, 2-3 rRNA, and 22-28 tRNAs, besides some ORF with unknown functions.

Furthermore, yeast mitogenomes have generally low GC contents, with numbers that can go as low as 7.6% in *Saccharomycodes ludwigii* (Hao, 2022). The average GC content from *Rhodotorula* mitogenomes is 40.1%, hence being relatively less A/T biased. Zhao et al. (2013) suggested that the *Rhodotorula* mitogenomes could have evolved with less mutation bias towards A/T than other known Basidiomycota yeasts.

The comparison of *Rhodotorula* genomes and mitogenomes, especially those with high-quality and annotated protein-coding sequences, has the potential to aid in inferring evolutionary changes in the genus. Schematic representations of *R. toruloides* CBS 14 genome and mitogenome assemblies are illustrated in Figure 7. High-quality genome assemblies are also necessary for studying genome structure and synteny of complete chromosomes across species (paper **II**).

Figure 7. Schematic representations of *Rhodotorula toruloides* CBS 14 genome. (a) Nuclear genome. The concentric circles show from outside to inside: contig names, sizes, and in 10 kb windows, gene densities in three replicas of CBS 14 grown on a mixture of hemicellulosic hydrolysate and crude glycerol, and three replicas of cells grown on crude glycerol only (grey circles). Figure modified from paper **III.** (b) Mitogenome scheme in which core mitochondrial genes are organized according to their position in the assembled contig. Quality scores are indicated per gene. Figure modified from paper **I.**

38

3. Genome organization and evolution in *Rhodotorula* spp*.*

In Eukaryotes, the genome is an organized physical structure in which DNA sequences are hierarchically folded into different layers, from chromatin to chromosome units to the whole genome (Figure 3). Such a structure is rather dynamic, affecting or regulating nuclear processes such as transcription, replication, and repair (Tokuda et al., 2012; Zimmer & Fabre, 2011). Ploidy levels indicate the number of complete sets of chromosomes that are present in a cell. They vary widely from strain to strain or even between individual cells in a population (Gilchrist & Stelkens, 2019; Harari et al., 2018; Legrand et al., 2019). Understanding the genome organization and ploidy of industrially interesting strains from the *R. glutinis* cluster is highly important in aiding the development of tailored genetic engineering approaches and synthetic yeast genomes (Schindler, 2020; Zhang et al., 2014).

3.1 Karyotypical variability

The number and structure of chromosomes per haploid genome represent a cell´s karyotype. For example, the 12 Mbp haploid genome from *Saccharomyces* spp. harbours about 16 chromosomes, while a reconstructed ancestor had about eight chromosomes before experiencing a whole-genome duplication (WGD) event (Wolfe & Shields, 1997) during its evolution (Borovkova et al., 2020; Gordon et al., 2009). High plasticity in chromosome number and sizes have been reported within yeast species even in the same geographical clade (Ahmad et al., 2014; Bravo Ruiz et al., 2019; Brown et al., 2011; Wendland & Walther, 2014). This karyotypical variability can result from chromosomal rearrangements, which provide genetic diversity

and thus drive evolution with different fitness implications (Ahmad et al., 2014).

The karyotype from *Rhodotorula* spp. was first studied by de Jonge et al. (1986) through pulsed-field gel electrophoresis (PFGE). They identified ten chromosomal bands in *R. toruloides* CBS 14 and *R. mucilaginosa* CBS 17. Yet, Ahmad et al. (2014) identified through PFGE only nine chromosomal bands in different *R. mucilaginosa* isolates, one of which had a different arrangement of band sizes. These electrophoretic karyotypes, however, cannot be directly translated into exact numbers of chromosomes. Some chromosomal DNA molecules might escape detection due to co-migrating with others or being unusually folded or very large and thus do not enter the agarose gel (de Jonge et al., 1986).

Moreover, 16 chromosomes were identified in *R. toruloides* NP11 through PFGE and genomic studies (Zhu et al., 2012). The estimated lengths of chromosomal DNA ranged from 0.65-1.9 Mbp. We predicted 18 chromosomes in *R. toruloides* CBS 14 after comparing contig and scaffold lengths from its high-quality genome assembly (paper **I**) and DNA bands from de Jonge et al. (1986). From the predicted putative chromosomes, 16 ranged from 0.62-1.97 Mbp, relatively similar to the results from Zhu et al. (2012), and two had smaller sizes (paper **I**). Additionally, *R. babjevae* harbours a maximum of 21 chromosomes according to nucleotide alignment patterns between homologous contig and scaffolds from two strains (paper **II**). Putative chromosomes´ lengths range from 0.4-2.4 Mbp.

Based on the total number of assembled scaffolds with telomeric repeats at either end of the DNA sequence, Tkavc et al. (2018) concluded that the genome from *R. taiwanensis* MD1149 is organized in at least 13 chromosomes. Telomeres comprise nucleoprotein structures at both ends of linear chromosomes, whose function is to maintain genome stability (De Lange, 2005). When both telomere DNA sequences are identified in an assembled contig or scaffold, it represents a complete chromosome. Telomeric repeats were detected at both termini of two contigs in CBS 14 (0.67 and 0.90 Mbp lengths), two in *R. babjevae* DBVPG 8058 (0.30 and 0.53 Mbp), and six in *R. glutinis* CBS 20 (0.77, 0.91, 1.00, 1.09, 1.35, and 1.67 Mbp), potentially representing complete chromosomes in each of the species (paper **I**, paper **II**).

3.2 Extrachromosomal circular DNA

In addition to chromosomes, extrachromosomal endogenous DNA has been detected in yeast genomes (Møller et al., 2015; Strope et al., 2015; paper **I**, paper **II**). These elements can be either linear or circular. Extrachromosomal circular DNA has been found as self-propagating DNA molecules ranging from 1-38 kb in *S. cerevisiae* (Gresham et al., 2010; Møller et al., 2015). They arise from genomic regions that contain many repetitive sequences or short and no repetitive sequences, representing a form of copy number variation with asymmetric segregation (Hull et al., 2019; Møller et al., 2015). Hence, they can be part of chromosomal structure evolution in yeast populations. One example of extrachromosomal circular DNA in *S. cerevisiae* is rDNA circles (Sinclair & Guarente, 1997). Furthermore, DNA mitochondrial plasmids have been reported in Basidiomycota filamentous fungi (Cahan & Kennell, 2005; Griffiths, 1995; Wang et al., 2008). Mitochondrial DNA plasmids are most frequently linear, with lengths ranging from 2.5–5 kb, and encoding enzymes that are involved in its replication (Cahan & Kennell, 2005; Wang et al., 2008).

Four circular contigs were predicted in *R. toruloides* CBS 14 hybrid genome assembly, including the mitochondrial genome (paper **I**). One of them, contig 63, had an 11 kb length, similar GC content to the chromosomes, and a read depth about three times higher, which could be indicative of relaxed replication regulation (Figure 8) (paper **I**). However, DNA replication sequences similar to ARS from *S. cerevisiae* (Dhar et al., 2012) were not found in contig 63 (paper **I**). Yet, AT-rich elements within the ARS consensus sequence can be degenerate and even absent in closely related species to *S. cerevisiae* (Dhar et al., 2012; Iwakiri et al., 2005). Contig 63 was annotated with the genes *UTP22* (encoding for RNA- associated protein 22), *H2A* (Histone H2A), and *H2B* (Histone H2B) (Figure 8a) (paper **I**). The fourth gene annotated in contig 63 was identified through Blast search as a truncated copy of *UTP22*. The circular structure of contig 63 was confirmed through PCR amplification and Sanger sequencing of overlapping fragments (Figure 8) (paper **I**). That was the first time such extrachromosomal circular DNA had been detected in basidiomycetes.
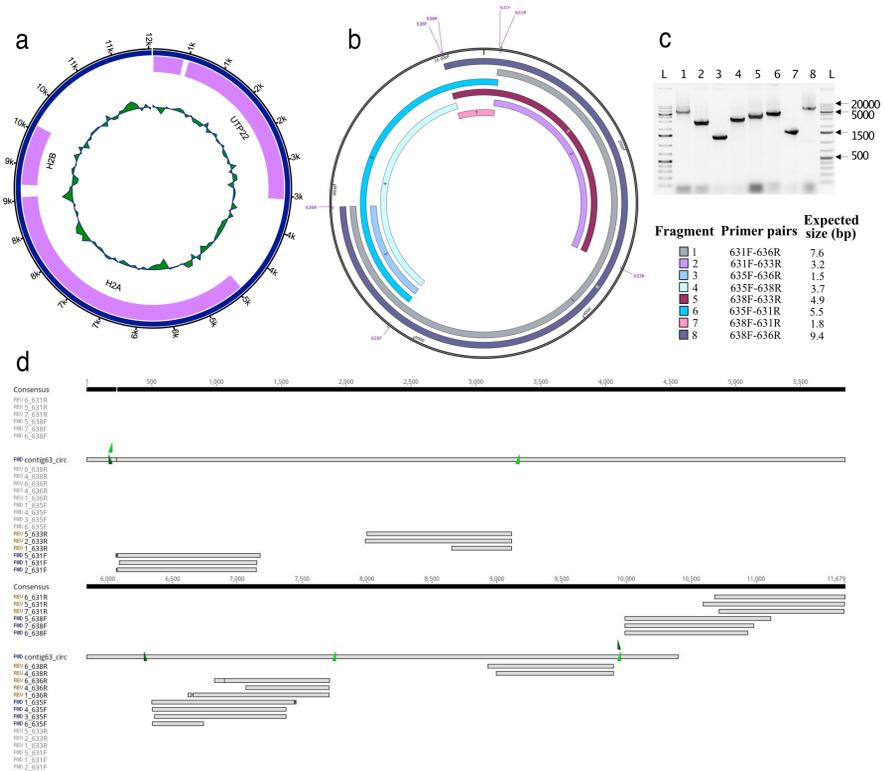
Figure 8. Extrachromosomal circular DNA from *Rhodotorula toruloides* CBS 14 assembled genome. (a) Contig 63 sequence overview. The concentric circles represent from outside to inside: size, annotated genes, identified gene names, and deviation from average GC content (61.8%) in non-overlapping 100 bp windows. (b) Different theoretical amplicons depending on the primer combinations. (c) Agarose gel electrophoresis using 1% agarose. Lane L, 1 Kb plus DNA ladder. The expected amplicon sizes from each primer pair are stated (table) (d) Assembly of amplicon sequences obtained after Sanger sequencing. The primer sites are represented in green: dark green for forward primers and light green for reverse primers. Primer names are indicated in the sequence name. The black bar represents the consensus sequence. Amplicon sequences are represented as grey bars when the average pair-wise identity is 100% and as a horizontal line when there are gaps or mismatches. Figure modified from paper **I**.

The *R. babjevae* genomes of strain CBS 7808 and DBVPG 8058 contain two strain-specific contigs each (paper **II**). Three represent small (6.0-8.7 kb) linear DNA sequences with higher read depth, one of which is even encoding Retrovirus-related Pol poly-protein from transposon 17.6 (paper **II**). The fourth strain-specific contig (30 kb) is a circular DNA molecule with lower read depth than chromosomes with shared homology between both strains.

GC content in these four contigs is variable but lower than the average GC ratio from chromosomes (paper **II**).

The relatively short circular genetic elements may be used to identify ARS- sequences in the *R. glutinis* cluster. The identification of such sequences could give a dramatic improvement to transformation efficiency in these yeast species, and thus further studies in the identified extrachromosomal endogenous DNA sequences are required.

3.3 Annotated transcripts and RNA splicing

Accurate gene and transcript annotations are necessary for understanding genome organization and gene functions within different cellular processes. The number of annotated protein-coding genes in *Rhodotorula* spp. ranges from 5,625 to 9,464 (7,518 average), strongly impacted by the quality of annotation approaches and reference genomes (Table 1) (Bo et al., 2022; Deligios et al., 2015; Fakankun et al., 2021; Firrincieli et al., 2015; Gan et al., 2017; Goordial et al., 2016; Hu & Ji, 2016; Li et al., 2020; Morin et al., 2014; Sen et al., 2019; Tkavc et al., 2018; Tran et al., 2019; Zhu et al., 2012; paper **I**; paper **II**). The average gene size among available annotated genomes is 2.1 kb (Table 2).

Table 2. Transcriptome information from strains belonging to *Rhodotorula toruloides*, *R. glutinis* and *R. babjevae*

Reference	Strain	Split genes	Av. exons per gene	Av. gene size (kb)
(Zhu et al., 2012)	NP11	7,938	NA	2.6
(Morin et al., 2014)	CECT1137	212	NA	NA
(Paper I)	CBS14	8,550	5.9	1.8
(Li et al., 2020)	ZHK	NA	NA	1.8
(Bo et al., 2022)	X-20	NA	NA	1.8
(Paper II)	CBS 20	6,390	3.9	2.0
	CBS 7808	6,390	4.0	2.0
	DBVPG 8058	6,305	3.9	2.0

Strain names that are represented in red, blue or green fonts belong to *R. toruloides*, *R. glutinis* and *R. babjevae*, respectively; NA, not available.

Transcriptome-based annotation of coding sequences has shown a higher number of transcripts than protein-coding genes (Kumar et al., 2012; Zhu et al., 2012; paper **I**; paper **II**). That can be indicative of alternative splicing in

the group. For instance, Zhu et al. (2012) identified 1371 genes in *R. toruloides* NP11 with two or more annotated transcript isoforms. In addition, a high proportion of the annotated genes contained several introns. The presence of split genes has been widely acknowledged in genomes from *Rhodotorula* spp., as illustrated in Table 2 (Deligios et al., 2015; Firrincieli et al., 2015; Gan et al., 2017; Li et al., 2020; Morin et al., 2014; Sen et al., 2019; Tkavc et al., 2018; Zhu et al., 2012; paper **I**; paper **II**). Available assembled mitogenomes have also indicated the presence of intronic ORF (Gan et al., 2017; Zhao et al., 2013; Zhou et al., 2020). The ratio of exons per gene in *Rhodotorula* spp. varies between 3.9 and 7.1, with the highest numbers in *R. mucilaginosa*. The number of protein-coding genes and the average ratio of exons per gene in *Rhodotorula* spp. (Table 2) is usually higher than in other yeast species (Fakankun et al., 2021; Juneau et al., 2007). For instance, 5,762 ORF have been annotated in the *S. cerevisiae* nuclear genome, and only 5% contain introns (Juneau et al., 2007). Intronic sequences in yeast can regulate its transcriptome and modulate processes such as gene expression (Goebels et al., 2013; Schirman et al., 2021). Intron gain during evolution results from intron transposition events, among other suggested mechanisms such as intronization or intron insertion during DNA repair (Poverennaya & Roytberg, 2020). It is a mechanism of alternative splicing evolution that enables the production of a higher number of mRNA isoforms per gene, consequently increasing transcriptome and proteome plasticity and potentially providing higher phenotype variability (Singh & Ahi, 2022). RNA-seq data from *R. glutinis* ZHK contained 2.18% reads that were matching intronic regions (Li et al., 2020). That corroborated the occurrence of intron retention events or alternative splicing in *Rhodotorula* yeasts.

Natural antisense transcripts have also been identified in the transcriptome of strains from the *R. glutinis* cluster (Zhu et al., 2012; paper **I**; paper **II**). The number of detected antisense transcripts in *R. babjevae* CBS 7808 and DBVPG 8058 was 315 and 309, respectively (paper **II**). One example is *FAS21*, which was annotated in CBS 14 in the complementary strand to *FAS2* (paper **I**). Its expression was verified during CBS 14 cultivation in media containing CG as the principal carbon source at 60 h (paper **III**). Transcription interference may represent a mechanism for antisense-dependent gene regulation in *Rhodotorula* spp. and likewise other yeast species (Nevers et al., 2018; Wery et al., 2018).

3.4 Ploidy levels

Under favourable conditions, yeast cells can propagate by mitosis, growing in clonal populations of haploid, diploid, polyploids, or even aneuploid individuals (Figure 9) (Dujon, 2010; Steensels et al., 2021). Aneuploid cells are those with an imbalanced number of chromosomes (i.e., not a multiple of the haploid chromosome number) with various effects on cell fitness (Gilchrist & Stelkens, 2019). Aneuploidy has been reported as a mechanism of natural adaptative evolution (Figure 10f) (Gilchrist & Stelkens, 2019; Smukowski Heil et al., 2017).
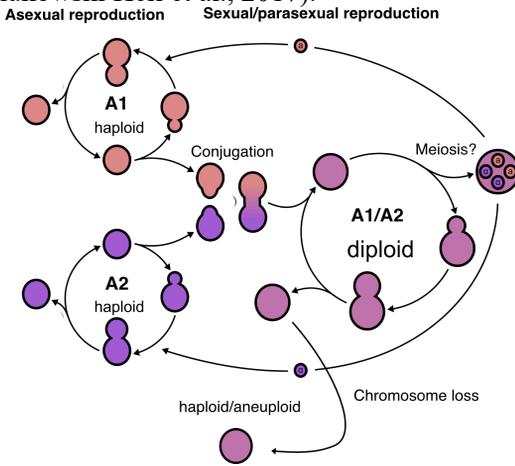

Figure 9. Schematic representation of *Rhodotorula* spp. life cycle. Figure modified from Usher (2019)

Besides asexual or clonal reproduction, yeast cells can additionally experience sexual and even parasexual mating (Figure 9) (Mishra et al., 2021). Many species can undergo sexual reproduction in various ways (Dujon, 2010; Steensels et al., 2021). Sexual reproduction generally involves mating of individuals with compatible mating types and meiosis for ploidy decrease, while parasexual mating can happen between cells of the same mating type, and ploidy decrease by mitotic chromosome loss (see below). After cell-cell fusion, a dikaryon is formed, containing both parental nuclei (Coelho et al., 2017). In some basidiomycetes, the step of karyogamy can be delayed, and consequently, cell lineages can experience an extended dikaryotic phase (Coelho et al., 2017). Moreover, interspecific hybridization events have been reported, including some examples between Basidiomycota yeast species (Bovers et al., 2006; Nakao et al., 2009; Sipiczki, 2019). The

multinucleate cells or heterokaryons can experience fitness defects and genetic instability if the parental strains are incompatible. They can subsequently revert to homokaryons or undergo parasexual chromosome loss producing aneuploid hybrid cells (Bennett, 2015). In some cases, aneuploid offspring is also a result of chromosomal nondisjunction and non-reciprocal genetic exchanges between the two parental genomes (Dujon, 2010; Gilchrist & Stelkens, 2019).

Abe and Sasakuma (1986) identified a diploid self-sporulating cycle in *R. toruloides*. In that study, cells were observed as diploid or aneuploid in addition to haploid progeny after a sexual cross between the compatible mating types (A1 and A2). Recently, the genome assembly from *R. toruloides* strain CGMCC 2.1609 supported the occurrence of aneuploidy in the species (Sambles et al., 2017). A higher proportion of CGMCC 2.1609 genome sequence (55.1%) was highly similar to CBS 14, while its 41.6% was highly similar to CBS 349 (Sambles et al., 2017). Moreover, *R. toruloides* strains NP11, ATCC 10788, and ATCC 10657 have been described as haploid (Hu & Ji, 2016; Liu et al., 2017; Zhu et al., 2012), while CBS 6016 and its domesticated strain Y4 as diploid (Liu et al., 2017).

R. babjevae strains CBS 7808 and DBVPG 8058 were identified as tetraploid using short-read sequencing data and the tools nQuire and KmerCountExact for assessing allele frequency values of single nucleotide polymorphisms (SNPs) and *k*-mer frequencies, respectively (paper **II**). In an earlier study, ploidy prediction using *k*-mer frequencies suggested that *R. mucilaginosa* JGTA-S1 is haploid, similar to strain C2.5t1 (Sen et al., 2019). However, when we reproduced the analysis with a reduced *k*-mer length, JGTA-S1 also showed tetraploidy (paper **II**).

Furthermore, NGS data is currently being used in our group to characterize the genomes of a broader range of strains from the *R. glutinis* cluster. An interesting preliminary result is the diploidy of *R. toruloides* CBS 14, previously described as haploid (Hu & Ji, 2016). Additionally, the hybrid strain CBS 6016 seems to be aneuploid with higher collinearity and proportion of matching syntenic blocks with CBS 14 than CBS 349 (Pappas, 2023), similar to the genome from CGMCC 2.1609 (Sambles et al., 2017).

3.5 Genome evolution

Genome analyses in natural and artificially evolved yeast populations can aid in inferring evolutionary relationships and identifying genome alterations that led to their physiological differentiation (Guillamón & Barrio, 2017; Sen et al., 2019). Extensive sequence divergences have been found between different yeast lineages (Dujon, 2010; Dujon et al., 2004). Dynamic mechanisms have generated these genetic polymorphisms with various evolutionary consequences (Guillamón & Barrio, 2017).

3.5.1 Mechanisms of genome evolution in yeast

Some examples of mechanisms involved in yeast genetic polymorphisms are illustrated in Figure 10. A common mechanism described in yeast is gene duplication (Dujon, 2010). Functional redundancy related to them is evolutionarily stable, and the retained paralogous genes can be preserved with the gain of functional specialization, dosage amplification, or backup compensation (Dean et al., 2008; Kuzmin et al., 2022). Examples of mechanisms that generate paralogous gene pairs are small-scale duplication events, including tandem or segmental duplications and the yeast WGD (Kellis et al., 2004; Kuzmin et al., 2022).
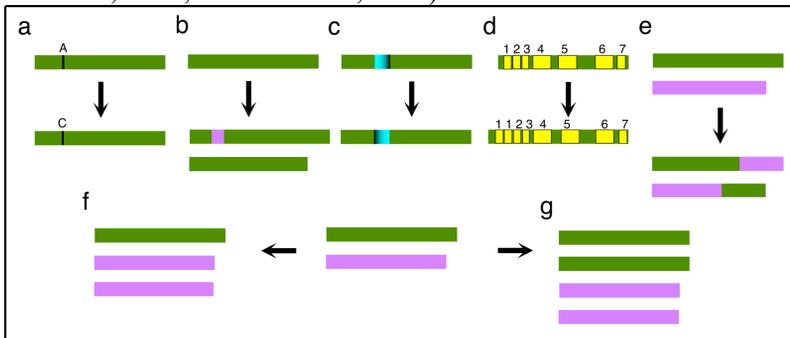
Figure 10. Examples of mechanisms generating genetic polymorphism in yeasts. (a) Single nucleotide polymorphisms; (b) short-sequence insertions and deletions; (c) inversions; (d) gene duplications; (e) translocations; and ploidy changes: (f) chromosomal duplication, generating aneuploidy and (g) the whole genome duplication.

Other mechanisms of genome evolution which also have the potential to generate novel gene functions are SNPs, inversions, and short-sequence insertions or deletions (Figure 10) (Dujon, 2010; Gorter et al., 2017; Guillamón & Barrio, 2017). They are generated due to polymerase errors in DNA replication or repair. Due to their high occurrence rate, SNPs are used

as genetic markers for estimating phylogenetic relationships (Guillamón & Barrio, 2017). Species from the same genus can have high sequence divergence even when sharing a highly conserved gene synteny (Dujon, 2010). SNPs producing synonymous codon changes are considered neutral genetic variants because they do not modify the encoded amino acid sequence and thus have little or no effect on phenotype (Doniger et al., 2008).

Furthermore, large chromosomal rearrangements, including translocations, inversions, deletions, and duplications of large chromosome regions are adaptative mechanisms that can affect gene order and dosage (Gorter et al., 2017). Hybridization is another driving source of evolutionary fate within yeast populations, which can additionally generate ploidy changes (Figure 10). Variations in ploidy levels have also been observed as an adaptation mechanism to different stressors (Gorter et al., 2017). In diploid cells, loss of heterozygosity mechanisms can remove sequence divergence between different alleles, producing chromosome pairs with both heterozygous and homozygous regions (Dujon, 2010; Smukowski Heil et al., 2017). Introgression events have also been reported in yeast (Kavanaugh et al., 2006; Peris et al., 2017).

3.5.2 Evolutive relationships in *Rhodotorula* spp.

The genetic heterogeneity of a high number of isolates belonging to *Rhodotorula* spp. was first investigated by Gadanho and Sampaio (2002). To delimitate *R. glutinis* and its related taxa, they compared all known *Rhodotorula* species at the time, including morphological and physiological characterization, sexual compatibility, PCR fingerprinting, rRNA sequence, and DNA-DNA reassociation analyses. Among their findings, Gadanho and Sampaio (2002) suggested designating strains from *R. babjevae*, *R. glutinis,* and *R. graminis* as *R. glutinis sensu stricto* due to their close relationship. Their close phylogenetic relationship based on the standard rDNA regions has been reported elsewhere (Figure 11) (Li et al., 2020; Sampaio, 2011b).
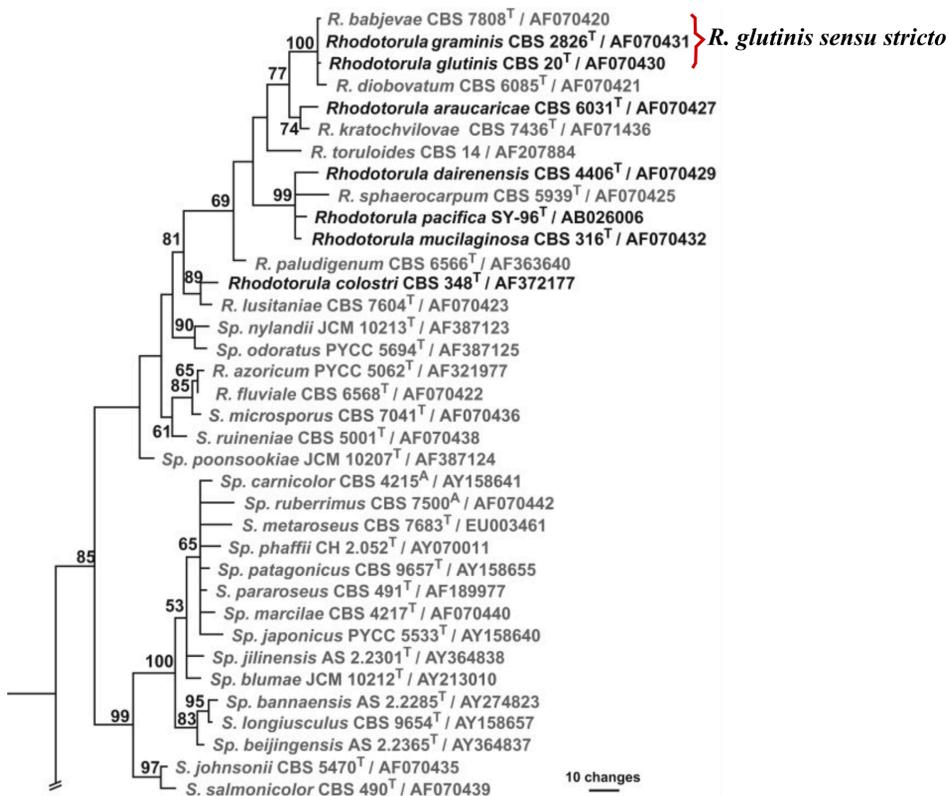
Figure 11. Phylogenetic placement of *Rhodotorula glutinis* cluster within the order Sporidiobolales, based on the D1/D2 regions of the LSU rRNA gene. Figure modified from Sampaio (2011b).

The availability of genome sequences from different *Rhodotorula* spp. has enabled the development of genome divergence analyses confirming the short evolutionary distance between *R. babjevae*, *R. glutinis,* and *R. graminis* (Figure 12) (paper **II**). For genome comparison, the number of shared orthologous clusters was evaluated, as well as DDH (DNA–DNA homology), ANI (Average Nucleotide Identity), and *K*r (paper **II**). For instance, the three species shared ANI values higher than 80% and 5933 out of 7223 orthologous clusters, representing putative common ancestral genes (Figure 12).

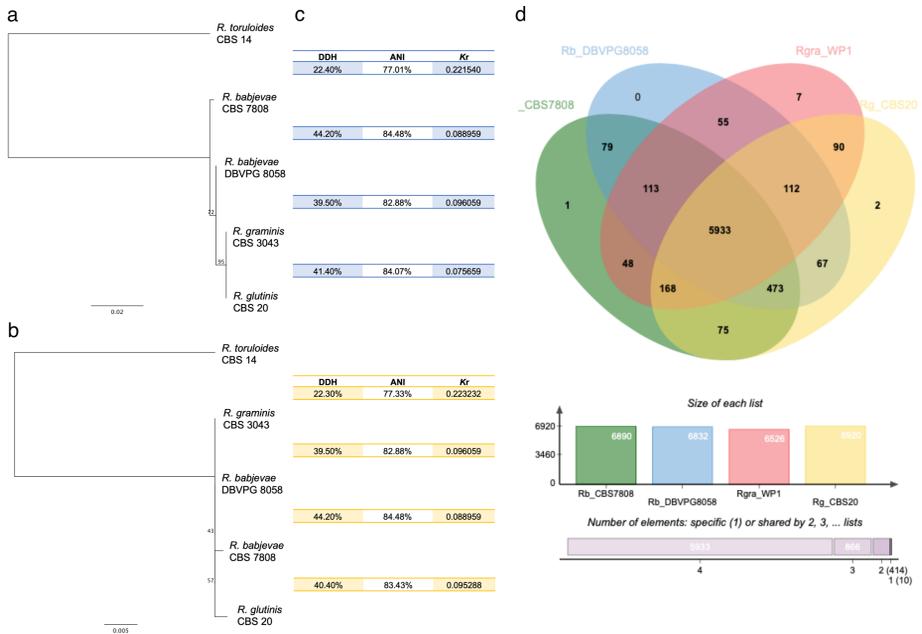Figure 12. Genome comparison analysis between species in the *Rhodotorula glutinis* cluster. Phylogenetic relationship between the closely related species *R. glutinis* (strain CBS 20), *R. graminis* (CBS 3043), *R. babjevae* (CBS 7808 and DBVPG 8058), and *R. toruloides* (CBS 14, outgroup) based on: (a) ITS; and (b) D1/D2 LSU of rRNA. The phylogenetic tree was inferred using PhyML with 100 bootstraps on Geneious prime v2021.0.1. (c) Genome divergence analysis based on the alignment-free distance measure *K*r, Average Nucleotide Identity (ANI), and DNA–DNA homology (DDH). (d) OrthoVenn2 diagram showing the distribution of shared orthologous clusters. For c) and d) the genome from *R. graminis* strain WP1 was used instead of CBS 3043. Figure modified from paper **II**.

In other studies, *R. glutinis* strains X-20 and ZHK were found to be closely related to *R. graminis* WP1 and *R. diobovata* (Figure 11) compared to other red yeast species (Bo et al., 2022; Li et al., 2020). From them, WP1 had the highest genome collinearity with *R. glutinis*. However, strains from *R. babjevae* were not included in the syntenic analysis (Bo et al., 2022; Li et al., 2020). In both studies, most orthologous genes were preserved through purifying natural selections, indicated by low non-synonymous substitution rates (Ka/Ks < 0.5) when comparing single-copy orthologous genes with those from WP1, *R. taiwanensis*, and *R. toruloides* NP11.

R. mucilaginosa strains JGTA-S1 and C2.5t1 were found to be more closely related to *Rhodotorula* sp. JG1b, then to *R. graminis* WP1 and *R. toruloides* NP11 based on the annotated orthologous proteins (Sen et al.,

2019). In the study, several copies of genes that encode putative anti-freeze glycoproteins were identified in the five assessed *Rhodotorula* genomes, suggesting a common ancestral evolution to cold tolerance (Sen et al., 2019). Furthermore, *R. mucilaginosa* strains showed to have collinearity through genome synteny analysis. Yet, the high number of assembled scaffolds in C2.5t1 precluded a more accurate identification of evolution mechanisms between the strains. That is a common issue when analyzing pairwise genome alignments with low-resolution genome assemblies from *Rhodotorula* spp. (Tkavc et al., 2018).

Interestingly, *R. babjevae* showed high inter-strain variability in terms of genome sequence and phenotype, which was comparable to interspecies variability in *R. glutinis sensu stricto* (Figure 12) (Brandenburg et al., 2021; paper **II**). Besides, homologous chromosomes in *R. babjevae* strains CBS 7808 and DBVPG 8058 shared pairwise identity values ranging between 82-87% (paper **II**). CBS 7808 had more duplicated genes than DBVPG 8058, suggesting a faster evolution. Most of these duplications were found near each other, indicating that they could have resulted from adaptative evolution to changing environments (Lallemand et al., 2020). In paper **II**, *SRRM2* was the most frequently duplicated gene. It encodes Ser/Arg repetitive matrix protein 2, which is involved in mRNA splicing. Thus, this gene could provide higher proteomic diversity and consequently higher levels of evolutionary plasticity as part of alternative splicing processes. Minor inversions and two translocation events were also identified. Moreover, several studies have reported a high intraspecific heterogeneity in *R. toruloides* (Gadanho & Sampaio, 2002; Hamamoto et al., 1986; paper **II**). It was suggested that nucleotide sequences and lipid production rates from *R. toruloides* strains belonging to different mating types have diversified highly (Hu & Ji, 2016; Zhang, Skerker, et al., 2016).

On the other hand, the mitochondrial genomes from *Rhodotorula* spp. have also been compared (Gan et al., 2017; Zhou et al., 2020). For instance, *R. toruloides* strain NP11 showed a high sequence identity between mitochondrial proteins with strain NBRC0880 (>85%) but lower to *R. taiwanensis* RS1 and *R. mucilaginosa* RIT389 (40–80%) (Zhou et al., 2020). However, in terms of mitochondrial gene synteny *R. toruloides* confirmed its high inter-strain variability when compared to the high collinearity between *R. taiwanensis* strains RS1 and MD1149 (Tkavc et al., 2018) or even between RS1 with *R. mucilaginosa* RIT389 (Gan et al., 2017; Zhou et al., 2020).

The standard rDNA regions, which are a common assessment tool in phylogenetic placement, have been found not sufficient for understanding yeast diversity (Chand Dakal et al., 2016; Conti et al., 2021; Libkind et al., 2020). Genome divergence values such as DDH, ANI, and Kr and evaluation of shared orthologous clusters were more sensitive methods for delineating *Rhodotorula* species (paper **II**). In general, a genome comparison study using high-quality genome sequences from a broad range of *Rhodotorula* strains is needed for re-assessing species delimitation and for understanding better their genome organization and evolution. However, due to population dynamics and the occurrence of heterospecific hybridization events, the definition of yeast species may remain a complex question (Dujon, 2010; Guillamón & Barrio, 2017).

4. Physiology of lipid production in *Rhodotorula* from residual products

Rhodotorula spp. can convert low-value residual products such as hemicellulose from the pulp-and-paper industry and CG from biodiesel production into lipids and carotenoids (Balan, 2019; Blomqvist et al., 2018; Brandenburg et al., 2018; Chmielarz et al., 2021; Nagaraj et al., 2022; Papanikolaou & Aggelis, 2020). Interestingly, increased lipid production rates were observed upon the cultivation of *R. toruloides* in CG when mixed with rather small amounts of hemicellulosic hydrolysate (HH) (Chmielarz et al., 2021). These high lipid production rates are desirable for increasing the sustainability and efficiency of bioconversion. That is because *R. toruloides* cultivation is aerobic and thus vastly energy-demanding (Karlsson et al., 2016).

Yeast cells can respond to their environment by triggering different molecular mechanisms to respond to inhibitors, assimilate carbon sources from media (e.g., glucose, xylose, glycerol, and acetate), and grow. Among others, these responses include transcriptome modulation, leading to the activation/deactivation of different metabolic pathways. *R. toruloides* transcriptome has been thus studied for describing metabolic pathways that are potentiated in response to different environments (Bommareddy et al., 2017; Jagtap et al., 2021; Touchette et al., 2022; Wang et al., 2018; Zhu et al., 2012; paper **III**). The generated data have also aided the reconstruction of more complete and accurate draft metabolic network models in *R. toruloides* (Kim et al., 2021; Wang et al., 2018).

4.1 Lipid metabolism

4.1.1 Biosynthesis of neutral lipids

Microbial conversion of residual substrates into TAG can be *de novo* or *ex novo*. The latter occurs when free FAs and glycerol, which resulted from the hydrolysis of hydrophobic substrates, are reassembled into TAG. On the other hand, *de novo* synthesis occurs when the cells are under carbon excess and stress, such as nutrient deficiency (i.e., nitrogen, phosphorus, sulphur, and others) (Granger et al., 1993). A simplified scheme of lipid and carotenoid metabolism under nitrogen starvation is illustrated in Figure 13.
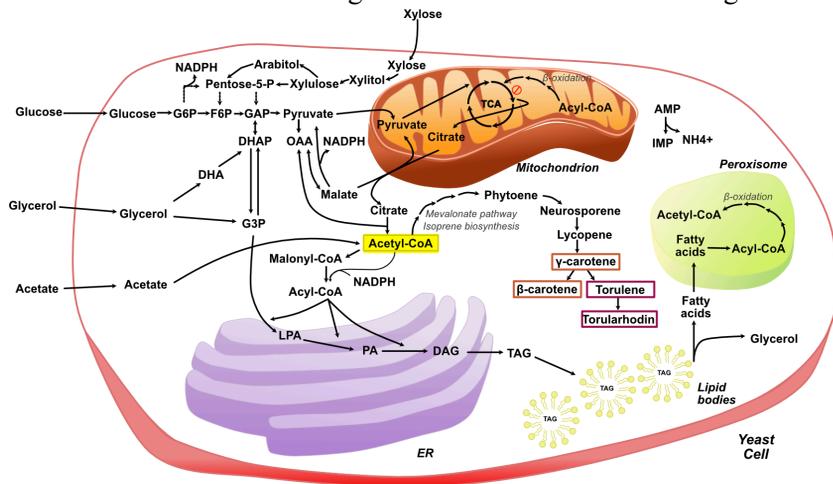

Figure 13. Simplified scheme of lipid and carotenoid metabolism in *Rhodotorula toruloides* under nitrogen starvation. Figure modified from Passoth et al. (2023).

Nitrogen starvation is a well-studied example of lipid biosynthesis triggered by nutrient deficiency in *R. toruloides*, as described in (Bommareddy et al., 2017; Passoth et al., 2023; Tiukova, Prigent, et al., 2019). Low cellular nitrogen levels are sensed by adenosine monophosphate (AMP) deaminase, which catalyses the conversion of AMP into inosine monophosphate (IMP) and ammonia. This reaction can potentially provide ammonium for cell maintenance. AMP is an allosteric activator of isocitrate dehydrogenase (IDH) which gets consequently deactivated with AMP rising conversion to IMP. As IDH is part of the tricarboxylic acid cycle (TCA), its rate also decreases. Isocitrate and citrate, which coexist in equilibrium, accumulate in mitochondria due to IDH deactivation. Citrate is then transported to the

cytoplasm, presumably through the citrate/malate shuttle. Once in the cytoplasm, citrate is converted into acetyl-CoA and oxaloacetate by ATP-citrate lyase (ACL). ACL represents the principal acetyl-CoA supplier for FA synthesis. Oxalacetate returns to mitochondria through the citrate/malate shuttle, countering the citrate efflux. FA synthesis involves the conversion of cytosolic acetyl-CoA into malonyl-CoA by acetyl-CoA carboxylase and then into acyl-CoA through the FA synthase (FAS) complex. Two C-atoms hereby are added at a time to the acyl-CoA chain, and CO_2 is released, leading to the formation of FAs of different lengths. These FAS reactions require NADPH as a co-factor (e.g., 16 mol NADPH for producing C_{18}-fatty acyl-CoA). NADPH can be generated through pentose-phosphate pathway (PPP) reactions (i.e., the steps catalysed by D-glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase) or by malic enzyme (ME). The cytosolic ME can generate NADPH when catalysing the oxidative decarboxylation of malate into pyruvate as part of the pyruvate-oxaloacetate-malate cycle (Bommareddy et al., 2015).

Under phosphate limitation, lipid accumulation is also enhanced in *R. toruloides* (Wang et al., 2018). AMP is converted through dephosphorylation into adenosine and phosphate, providing phosphate for cell maintenance. As described above, decreased levels of AMP results in IDH inactivation, mitochondrial citrate accumulation and transport to cytoplasm, acetyl-CoA formation and FA biosynthesis. During phosphate limitation, however, NADPH levels supplied by the PPP are lower.

The synthesised acyl-CoA chains during either of the two described cell physiological conditions are esterified with glycerol-3-phosphate in the endoplasmic reticulum. This step results in the formation of structural- and storage lipids or TAG (Passoth et al., 2023). They are then stored in lipid droplets (Figure 2) (Passoth et al., 2023).

Some differences in lipid metabolism have been observed between *R. toruloides* and the conventional yeast *S. cerevisiae*. For example, the origin of cytosolic acetyl-CoA used for FA synthesis in *R. toruloides* is citrate through ACL (encoded by *ACL1*), whereas it is acetate in *S. cerevisiae* as part of the pyruvate dehydrogenase bypass (Rodriguez et al., 2016). Interestingly, in CG and HH mixture (CGHH), when there is a surplus of acetate, the transcription of the gene encoding Acetyl-CoA synthetase (*ACS1*) is upregulated rather than *ACL1* (paper **III**). Moreover, FAs are degraded through β-oxidation by both peroxisomal and mitochondrial

enzymes in *R. toruloides*, but only by the former in *S. cerevisiae* (Kim et al., 2021; paper **III**). In addition, *R. diobovata* and other *Rhodotorula* spp. have one ME in the cytoplasm and another in mitochondria (Fakankun et al., 2021; paper **I**; paper **II**). Yet, only mitochondrial ME has been found in *L. starkeyi* and *Y. lipolytica*, being mostly NAD^+-dependent (Ratledge, 2014; Tang et al., 2010). NADPH supply in these oleaginous yeast species may be provided by cytosolic $NADP^+$ dependent-IDH instead (Ratledge, 2014).

4.1.2   Biosynthesis of carotenoids

Carotenoid biosynthesis in *Rhodotorula* spp. involves reactions within the mevalonate pathway, isoprene biosynthesis and carotenogenesis starting from cytosolic acetyl-CoA (Figure 13) (Igreja et al., 2021; Wei et al., 2019). Phytoene is the first colourless carotenoid formed. It is then converted into neurosporene as part of the carotenogenic pathway and subsequently into lycopene by phytoene desaturase (encoded by *crtI*). Lycopene represents a precursor of cyclic carotenoids such as γ-carotene, β-carotene, torulene and torularhodin (Igreja et al., 2021). Bifunctional lycopene cyclase/phytoene synthase (*crtYB*) can convert lycopene into γ-carotene, which may be further converted to β-carotene or torulene, a precursor of torularhodin (Wei et al., 2019).

4.2   Conversion of residual products into lipids

4.2.1   Hemicellulosic hydrolysate

Wheat straw and wood residues are examples of lignocellulosic materials with applications as a substrate for yeast fermentation (Passoth & Sandgren, 2019). Hemicellulose, cellulose and lignin are the three major polymers of plant cell walls (Figure 14). However, lignocellulosic biomass is recalcitrant to degradation and consequently requires pre-treatments to disrupt its structure and release sugar monomers. HH is the hemicellulosic fraction from lignocellulosic materials after hydrolytic pre-treatment (Passoth & Sandgren, 2019). HH composition varies depending on the hydrolysate preparation procedure. It can contain a mixture of xylose, acetic acid and glucose as major components, representing carbon sources for yeast growth (Brandenburg et al., 2021; Chmielarz et al., 2021; paper **III**).
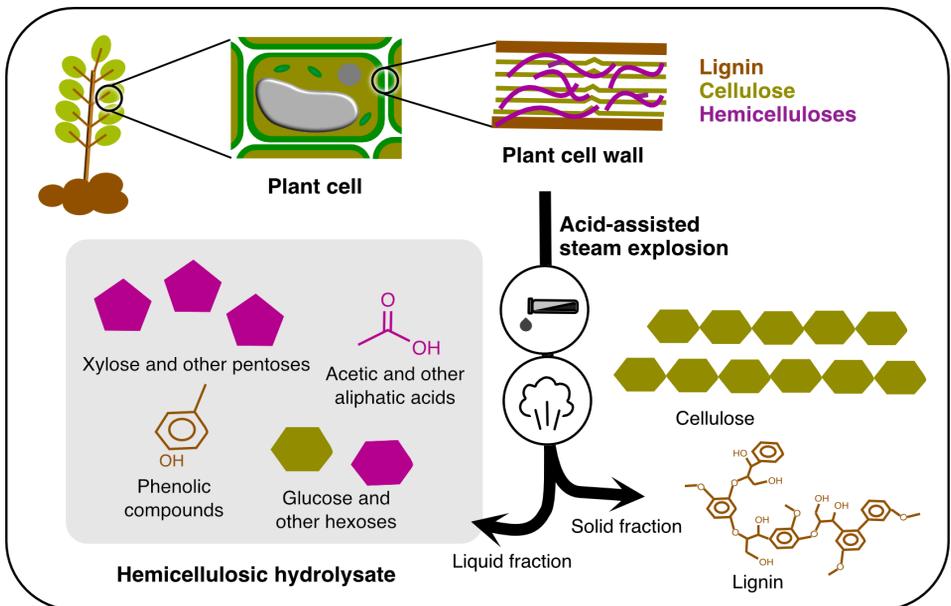
Figure 14. Simplified scheme of hemicellulosic hydrolysate origin.

Recently, Jagtap et al. (2021) studied genes and metabolic pathways that are differentially regulated when *R. toruloides* IFO0880 was cultivated on YP (10 g/L yeast extract and 20 g/L peptone) supplemented with xylose or acetate compared to YP supplemented with glucose (YPG). During growth on YPG compared to YP, transcription of genes involved in the lower half of glycolysis, pyruvate dehydrogenase subunits and the TCA was upregulated. The expression of genes within lipid biosynthesis was also increased in the study. According to concentrations of intracellular metabolites, Jagtap et al. (2021) suggested that glucose could also have been converted into galactose to synthesize other compounds such as polysaccharides, glycolipids or glycoproteins.

Xylose assimilation in *R. toruloides* involves the sequential steps catalysed by xylose reductase, xylitol dehydrogenase, and arabitol dehydrogenase (Jagtap et al., 2021; paper **III**). Instead of *R. toruloides* conversion of xylulose into arabitol, in other yeast species it is transformed to xylulose-5-phosphate by xylulokinase, which can then enter the PPP (Jagtap & Rao, 2018). Arabitol production during yeast growth on xylose could be an adaptive strategy but of unknown benefits to the cell (Jagtap et al., 2021; Pinheiro et al., 2020).

During *R. toruloides* growth on acetate, enhanced transcription of genes involved in gluconeogenesis, PPP and TCA was observed (Jagtap et al., 2021). Yet, gene expression within lipid production pathways was reduced compared to YPG. This reduction was not observed during *R. toruloides* CBS 14 cultivation on CGHH (paper **III**). In that study, acetate and xylose from HH enhanced lipid production to a certain extent, which could be a lower modulation than when glucose is in the medium. *ACS1* expression was enhanced, suggesting acetate assimilation via this reaction (Jagtap et al., 2021; paper **III**).

4.2.2 Crude glycerol

The valorisation of CG has generated interest for being a major by-product of biodiesel production (Figure 15) (da Silva et al., 2009). Several studies have evaluated the ability of yeast species to assimilate CG alone or in combination with other carbon sources (Bommareddy et al., 2017; Chmielarz et al., 2021; Filippousi et al., 2019). *Rhodotorula* spp. have shown a good performance compared to other yeast species in terms of increased biomass and lipid production when grown on glycerol-containing media (Abeln & Chuck, 2021; Chmielarz et al., 2021; Filippousi et al., 2019; Rakicka et al., 2015). For instance, *R. toruloides* achieved the highest lipid production among seven oleaginous yeast species, with lipid yields of 0.25 g per consumed g of carbon source (Chmielarz et al., 2021).
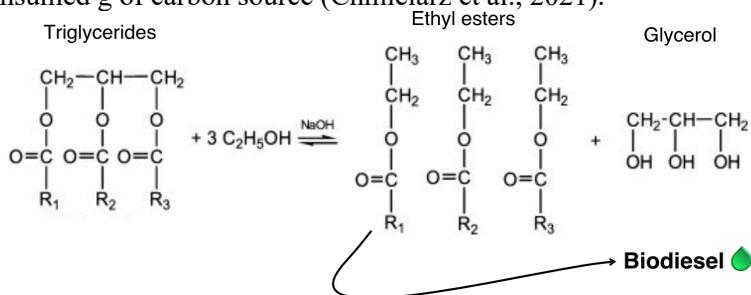
Figure 15. Transesterification reaction rendering biodiesel. Figure modified from da Silva et al. (2009).

Glycerol consumption in *R. toruloides* has only been observed after limiting glucose concentrations due to catabolite repression presumably (Bommareddy et al., 2015). Yet, a mixture of glycerol and glucose or the complex substrate HH enhanced lipid productivity in the strains DSMZ 4444 and CBS 14, respectively (Bommareddy et al., 2015; Chmielarz et al., 2021;

paper **III**). The catabolic L-glycerol 3-phosphate pathway (*GUT1* and *GUT2*) is the main route for glycerol assimilation in *S. cerevisiae* (Sprague & Cronan, 1977; Swinnen et al., 2013). *GUT1* and *GUT2* were transcribed in *R. toruloides* CBS 14 when cells were cultivated on CGHH and CG alone (paper **III**). Bommareddy et al. (2017) also detected *GUT1* and *GUT2* expression in a medium containing a 1:1 mixture of glycerol and glucose. The study showed that regulation of glycerol uptake via the catabolic L-glycerol 3-phosphate pathway at the transcriptional level could be different than in *S. cerevisiae* and not be affected by the presence of glucose (Bommareddy et al., 2017). Furthermore, genes encoding glycerol dehydrogenase (GDH) and dihydroxyacetone kinase enzymes, which are part of the dihydroxyacetone pathway (Klein et al., 2017; Matsuzawa et al., 2010), were also expressed in both transcriptomic studies (Bommareddy et al., 2017; paper **III**). That could represent an additional pathway of glycerol entrance to central metabolism and NADPH generation for biosynthetic reactions (Figure 13). Interestingly, the two expressed GDH in paper **III** were annotated as $NADP^+$-dependent, contradicting its previous description as part of glycerol anabolic reactions (Klein et al., 2017). They were upregulated when HH was added to CG media, and thus their transcription may be stimulated by molecular components of HH. Furthermore, transcription of genes involved in the catabolic glyceraldehyde pathway from *Neurospora crassa* (Klein et al., 2017; Tom et al., 1978; Viswanath-Reddy et al., 1977) was also observed (paper **III**). Further investigations are still needed to confirm the implications of this pathway in *R. toruloides* glycerol utilization.

On the other hand, glycerol production in response to stress was suggested in *R. toruloides* DSMZ 4444 (Bommareddy et al., 2017).

4.3   Genes and activated pathways associated with faster glycerol consumption

Faster growth, initial glycerol consumption and higher lipid productivity were observed in *R. toruloides* CBS 14 cultures when small amounts of HH were added to CG (Chmielarz et al., 2021). In paper **III**, RNA samples were collected at similar physiological situations in CGHH and CG media to investigate the metabolic pathways that are activated by HH and enhance glycerol utilization. At 10 h cultivations, transcription increased among

genes encoding mitochondrial enzymes (i.e., those involved in TCA and oxidative phosphorylation) in CGHH compared to CG (Figure 16). In addition, there was an enhanced transcription of genes involved in protein turnover, β-oxidation, handling oxidative stress and degradation of xylose and aromatic compounds in the presence of the HH. We suggested that the physiological reason for faster glycerol assimilation, growth and lipid production in CGHH could be the transcription of genes involved in pathways that generate metabolic energy (Figure 16) (paper **III**). High ATP levels could provide the cells with sufficient energy needed for the observed increase in protein turnover processes, which enabled an efficient adaption to the changing carbon source (paper **III**).
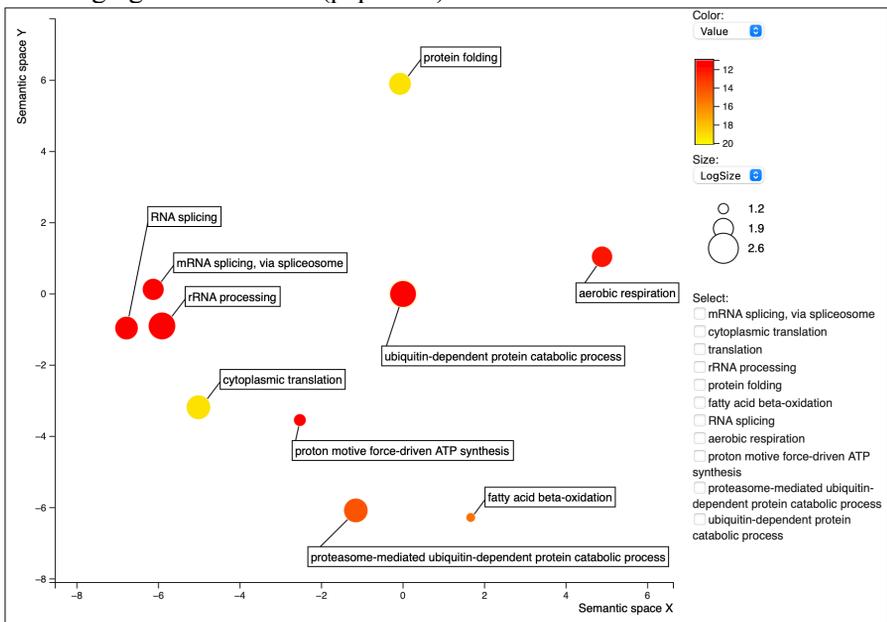

Figure 16. Gene ontology (GO) terms belonging to biological processes of upregulated genes (adjusted p-value < 0.05) in CGHH compared to CG at 10 h cultivations. The axes in the scatterplot have no intrinsic meaning. The colour of the bubbles indicates the number of upregulated genes in each GO term. The different sizes of the bubbles correspond to the frequencies of each GO term, which were calculated based on the *Saccharomyces cerevisiae* S288C reference database (larger bubbles belong to more general GO terms). GO terms that occurred less than 10 times were filtered out. CGHH, mixture of crude glycerol and hemicellulosic hydrolysate; CG, crude glycerol.

5. Conclusions and further perspectives

Oleaginous yeast species from the *Rhodotorula* genus have a great biotechnological potential for their ability to accumulate lipids and carotenoids, with a lipid profile similar to VO. Putative biotechnological applications from *Rhodotorula* microbial lipids include the production of feed and food additives, pharmaceuticals, biofuels and other products of commercial interest. Microbial lipid potential as a replacement of VO represents an opportunity for exploiting resources with reduced carbon footprint and food supply impairment. In addition, *Rhodotorula* spp. has the advantage over other oleaginous microorganisms of growing on residual products, making the bioprocess more economically and sustainably attractive. However, the lower prices and more established technologies of VO production decrease the industrial attractiveness of large-scale technological development of microbial lipids. Optimization of microbial lipid production and processing is needed to make the technologies more competitive. In my opinion, good marketing strategies can also play an important role in microbial lipids gaining a market share, for instance, as feed additives.

Genomic and transcriptomic data was recovered and used in the construction of hybrid genome assemblies and gene annotation from *R. toruloides*, *R. glutinis* and *R. babjevae* strains. The approach achieved high quality in terms of completeness, contiguity (near-chromosome level) and the high number of annotated genes and associated transcripts. We identified mitochondrial genomes and putative extrachromosomal elements, including the first-time detection of a potential plasmid in basidiomycetes. We also predicted ploidy and karyotype from some *Rhodotorula* strains. The annotation approach indicated a high proportion of split genes, a high number of exons per gene and the presence of antisense transcripts, the

expression of one of which was detected. Overall, the results presented in this thesis expand our understanding of the genome organization and gene expression in the *R. glutinis* cluster. It would be interesting to see a continuation in this study, including the identification of telomeric and centromeric regions, for instance, allowing the establishment of *Rhodotorula* spp. as models for understanding eukaryotic genomes. Our results can also help establish more efficient genetic engineering methods that convert *Rhodotorula* spp. in versatile microbial cell factories. To give an idea, further analysis of extrachromosomal elements in *Rhodotorula* spp. can lead to the identification of a replication origin, with applications as part of efficient episomal vectors. Ploidy and karyotype identification can also be valuable for strain selection.

On the other hand, gene expression data from *R. toruloides* gave insights into its transcriptional regulation when grown on CG supplemented with HH. We suggested that the higher lipid productivity rates observed in CGHH are due to earlier transcription of genes involved in energy-generating pathways. Genomics and transcriptomics studies within this thesis are useful for identifying new genetic engineering targets to develop *Rhodotorula* strains with higher lipid productivity on low-cost substrates. At the same time, they can deepen our understanding of the physiological bases for *Rhodotorula* spp. conversion of residual products into lipids. In addition, they can also be of great help for updating draft metabolic network models, making them completer and more accurate. In the future, it would be interesting to investigate the contribution of each glycerol assimilation pathway suggested in paper **III** through targeted gene deletion or overexpression. Multi-omic studies could also be considered as a tool for supporting some of the results in addition to whether glycerol synthesis is activated in response to stress, as suggested by Bommareddy et al. (2017).

Genome evolution analysis revealed that *R. babjevae* strains have a closer relationship to *R. glutinis* than to *R. toruloides* and a high intraspecific divergence. The high genetic variability within strains from the *R. glutinis* cluster could make them valuable genome evolution models.

The optimized methods and pipelines within *Rhodotorula* spp. genomics and gene expression studies can be further used in a broader spectrum of yeast species. For instance, an increasing number of *Rhodotorula* isolates haven´t been studied at a genomic level yet. In that sense, hybrid assembly approaches have been proven effective for generating high-quality genome

assemblies with near-chromosome level resolution. These assemblies could represent good reference genomes and high-quality input data for other omics and metabolic networks analysis in non-conventional yeast species. In addition, annotation approaches using good reference genomes and expression data from different cultivation conditions showed to be successful in identifying a higher number of transcript isoforms and accurate predictions of intron/exon boundaries in *Rhodotorula* yeasts. NGS technologies, such as Nanopore sequencing, are continuously becoming more affordable and accurate. Perhaps new yeast isolates could be explored using WGS methods to perform genome assembly, strain characterization and more precise phylogenetic placements. Ultimately, genome comparison studies using a higher number of genome sequences and the described genome divergence values could contribute to the identification of novel yeast taxa.

References

Abe, K., & Sasakuma, T. (1986). Identification of a Diploid Self-sporulating Cycle in the Basidiomycetous Yeast *Rhodosporidium toruloides*. *Microbiology*, *132*, 1459-1465.

Abeln, F., & Chuck, C. J. (2021). The history, state of the art and future prospects for oleaginous yeast research. *Microbial Cell Factories*, *20*(1), 221. https://doi.org/10.1186/s12934-021-01712-1

Ahmad, K. M., Kokošar, J., Guo, X., Gu, Z., Ishchuk, O. P., & Piškur, J. (2014). Genome structure and dynamics of the yeast pathogen *Candida glabrata*. *FEMS Yeast Research*, *14*(4), 529-535. https://doi.org/10.1111/1567-1364.12145

Alcock, T. D., Salt, D. E., Wilson, P., & Ramsden, S. J. (2022). More sustainable vegetable oil: Balancing productivity with carbon storage opportunities. *Science of The Total Environment*, *829*, 154539. https://doi.org/10.1016/j.scitotenv.2022.154539

Alvarez, H., & Steinbüchel, A. (2002). Triacylglycerols in prokaryotic microorganisms. *Applied Microbiology and Biotechnology*, *60*(4), 367-376. https://doi.org/10.1007/s00253-002-1135-0

Alvarez, H. M., Pucci, O. H., & Steinbüchel, A. (1997). Lipid storage compounds in marine bacteria. *Applied Microbiology and Biotechnology*, *47*(2), 132-139. https://doi.org/10.1007/s002530050901

Ayadi, I., Belghith, H., Gargouri, A., & Guerfali, M. (2018). Screening of new oleaginous yeasts for single cell oil production, hydrolytic potential exploitation and agro-industrial by-products valorization. *Process Safety and Environmental Protection*, *119*, 104-114. https://doi.org/10.1016/j.psep.2018.07.012

Balan, V. (2019). *Microbial Lipid Production* (1 ed.). Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9484-7

Banno, I. (1967). STUDIES ON THE SEXUALITY OF *RHODOTORULA*. *The Journal of General and Applied Microbiology*, *13*(2), 167-196. https://doi.org/10.2323/jgam.13.167

Bastos Lima, M. G. (2021). The Contested Sustainability of Biofuels in a North-South Context. In M. G. Bastos Lima (Ed.), *The Politics of Bioeconomy and Sustainability: Lessons from Biofuel Governance, Policies and Production Strategies in the Emerging World* (pp. 23-47). Springer International Publishing. https://doi.org/10.1007/978-3-030-66838-9_2

Bennett, R. J. (2015). The parasexual lifestyle of *Candida albicans*. *Curr Opin Microbiol*, *28*, 10-17. https://doi.org/10.1016/j.mib.2015.06.017

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., . . . Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53-59. https://doi.org/10.1038/nature07517

Blomqvist, J., Pickova, J., Tilami, S. K., Sampels, S., Mikkelsen, N., Brandenburg, J., Sandgren, M., & Passoth, V. (2018). Oleaginous yeast as a component in fish feed. *Sci Rep*, *8*(1), 15945. https://doi.org/10.1038/s41598-018-34232-x

Bo, S., Ni, X., Guo, J., Liu, Z., Wang, X., Sheng, Y., Zhang, G., & Yang, J. (2022). Carotenoid Biosynthesis: Genome-Wide Profiling, Pathway Identification in *Rhodotorula glutinis* X-20, and High-Level Production. *Front Nutr*, *9*, 918240. https://doi.org/10.3389/fnut.2022.918240

Bommareddy, R. R., Sabra, W., Maheshwari, G., & Zeng, A.-P. (2015). Metabolic network analysis and experimental study of lipid production in *Rhodosporidium toruloides* grown on single and mixed substrates. *Microbial Cell Factories*, *14*(1), 36. https://doi.org/10.1186/s12934-015-0217-5

Bommareddy, R. R., Sabra, W., & Zeng, A.-P. (2017). Glucose-mediated regulation of glycerol uptake in *Rhodosporidium toruloides*: Insights through transcriptomic analysis on dual substrate fermentation. *Engineering in Life Sciences*, *17*(3), 282-291. https://doi.org/10.1002/elsc.201600010

Bonturi, N., Crucello, A., Viana, A. J. C., & Miranda, E. A. (2017). Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a *Rhodosporidium toruloides* adapted strain. *Process Biochemistry*, *57*, 16-25. https://doi.org/10.1016/j.procbio.2017.03.007

Bonturi, N., Pinheiro, M. J., de Oliveira, P. M., Rusadze, E., Eichinger, T., Liudžiūtė, G., De Biaggi, J. S., Brauer, A., Remm, M., Miranda, E. A., Ledesma-Amaro, R., & Lahtvee, P.-J. (2022). Development of a dedicated Golden Gate Assembly Platform (RtGGA) for *Rhodotorula toruloides*. *Metabolic Engineering Communications*, *15*, e00200. https://doi.org/10.1016/j.mec.2022.e00200

Borovkova, A. N., Michailova, Y. V., & Naumova, E. S. (2020). Molecular Genetic Features of Biological Species of the Genus *Saccharomyces*. *Microbiology*, *89*(4), 387-395. https://doi.org/10.1134/S0026261720040037

Bovers, M., Hagen, F., Kuramae, E. E., Diaz, M. R., Spanjaard, L., Dromer, F., Hoogveld, H. L., & Boekhout, T. (2006). Unique hybrids between the fungal pathogens *Cryptococcus neoformans* and *Cryptococcus gattii*. *FEMS Yeast Research*, *6*(4), 599-607. https://doi.org/10.1111/j.1567-1364.2006.00082.x

Boviatsi, E., Papadaki, A., Efthymiou, M.-N., Nychas, G.-J. E., Papanikolaou, S., da Silva, J. A. C., Freire, D. M. G., & Koutinas, A. (2020). Valorisation of sugarcane molasses for the production of microbial lipids via fermentation of two *Rhodosporidium* strains for enzymatic synthesis of polyol esters. *Journal of Chemical Technology & Biotechnology*, *95*(2), 402-407. https://doi.org/10.1002/jctb.5985

Brancaccio, R. N., Robitaille, A., Dutta, S., Rollison, D. E., Tommasino, M., & Gheit, T. (2021). MinION nanopore sequencing and assembly of a complete human papillomavirus genome. *Journal of Virological Methods*, *294*, 114180. https://doi.org/10.1016/j.jviromet.2021.114180

Brandenburg, J., Blomqvist, J., Pickova, J., Bonturi, N., Sandgren, M., & Passoth, V. (2016). Lipid production from hemicellulose with *Lipomyces starkeyi* in a pH regulated fed-batch cultivation. *Yeast*, *33*(8), 451-462. https://doi.org/10.1002/yea.3160

Brandenburg, J., Blomqvist, J., Shapaval, V., Kohler, A., Sampels, S., Sandgren, M., & Passoth, V. (2021). Oleaginous yeasts respond differently to carbon sources present in lignocellulose hydrolysate. *Biotechnology for Biofuels*, *14*(1), 124. https://doi.org/10.1186/s13068-021-01974-2

Brandenburg, J., Poppele, I., Blomqvist, J., Puke, M., Pickova, J., Sandgren, M., Rapoport, A., Vedernikovs, N., & Passoth, V. (2018). Bioethanol and lipid production from the enzymatic hydrolysate of wheat straw after furfural extraction. *Applied Microbiology and Biotechnology*, *102*(14), 6269-6277. https://doi.org/10.1007/s00253-018-9081-7

Brandt, C., Bongcam-Rudloff, E., & Müller, B. (2020). Abundance Tracking by Long-Read Nanopore Sequencing of Complex Microbial Communities in Samples from 20 Different Biogas/Wastewater Plants. *Applied Sciences*, *10*(21), 7518. https://www.mdpi.com/2076-3417/10/21/7518

Bravo Ruiz, G., Ross, Z. K., Holmes, E., Schelenz, S., Gow, N. A. R., & Lorenz, A. (2019). Rapid and extensive karyotype diversification in haploid clinical *Candida auris* isolates. *Current Genetics*, *65*(5), 1217-1228. https://doi.org/10.1007/s00294-019-00976-w

Brown, W. R. A., Liti, G., Rosa, C., James, S., Roberts, I., Robert, V., Jolly, N., Tang, W., Baumann, P., Green, C., Schlegel, K., Young, J., Hirchaud, F., Leek, S., Thomas, G., Blomberg, A., & Warringer, J. (2011). A Geographically Diverse Collection of *Schizosaccharomyces pombe* Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. *G3 Genes|Genomes|Genetics*, *1*(7), 615-626. https://doi.org/10.1534/g3.111.001123

Brunel, M., Burkina, V., Pickova, J., Sampels, S., & Moazzami, A. A. (2022). Oleaginous yeast *Rhodotorula toruloides* biomass effect on the metabolism of Arctic char (*Salvelinus alpinus*). *Frontiers in Molecular Biosciences*, *9*. https://www.frontiersin.org/articles/10.3389/fmolb.2022.931946

Buedenbender, L., Kumar, A., Blümel, M., Kempken, F., & Tasdemir, D. (2020). Genomics-and metabolomics-based investigation of the deep-sea sediment-

derived yeast, *Rhodotorula mucilaginosa* 50-3-19/20B. *Marine drugs*, *19*(1), 14. https://doi.org/10.3390/md19010014

Cahan, P., & Kennell, J. C. (2005). Identification and distribution of sequences having similarity to mitochondrial plasmids in mitochondrial genomes of filamentous fungi. *Molecular Genetics and Genomics*, *273*(6), 462-473. https://doi.org/10.1007/s00438-005-1133-x

Carmona-Cabello, M., García, I. L., Papadaki, A., Tsouko, E., Koutinas, A., & Dorado, M. P. (2021). Biodiesel production using microbial lipids derived from food waste discarded by catering services. *Bioresour Technol*, *323*, 124597. https://doi.org/10.1016/j.biortech.2020.124597

Castañeda, M. T., Nuñez, S., Garelli, F., Voget, C., & De Battista, H. (2018). Comprehensive analysis of a metabolic model for lipid production in *Rhodosporidium toruloides*. *Journal of Biotechnology*, *280*, 11-18. https://doi.org/10.1016/j.jbiotec.2018.05.010

Chand Dakal, T., Giudici, P., & Solieri, L. (2016). Contrasting Patterns of rDNA Homogenization within the *Zygosaccharomyces rouxii* Species Complex. *PLoS One*, *11*(8), e0160744. https://doi.org/10.1371/journal.pone.0160744

Chmielarz, M., Blomqvist, J., Sampels, S., Sandgren, M., & Passoth, V. (2021). Microbial lipid production from crude glycerol and hemicellulosic hydrolysate with oleaginous yeasts. *Biotechnol Biofuels*, *14*(1), 65. https://doi.org/10.1186/s13068-021-01916-y

Coelho, M. A., Bakkeren, G., Sun, S., Hood, M. E., & Giraud, T. (2017). Fungal Sex: The Basidiomycota. *Microbiology Spectrum*, *5*(3), 5.3.12. https://doi.org/doi:10.1128/microbiolspec.FUNK-0046-2016

Coelho, M. A., Rosa, A., Rodrigues, N., Fonseca, Á., & Gonçalves, P. (2008). Identification of Mating Type Genes in the Bipolar Basidiomycetous Yeast *Rhodosporidium toruloides*: First Insight into the *MAT* Locus Structure of the *Sporidiobolales*. *Eukaryotic Cell*, *7*(6), 1053-1061. https://doi.org/doi:10.1128/EC.00025-08

Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., & Walters, L. (1998). New Goals for the U.S. Human Genome Project: 1998-2003. *Science*, *282*(5389), 682-689. https://doi.org/doi:10.1126/science.282.5389.682

Conti, A., Corte, L., Casagrande Pierantoni, D., Robert, V., & Cardinali, G. (2021). What is the best lens? Comparing the resolution power of genome-derived markers and standard barcodes. *Microorganisms*, *9*(2), 299.

Coradetti, S. T., Pinel, D., Geiselman, G. M., Ito, M., Mondo, S. J., Reilly, M. C., Cheng, Y.-F., Bauer, S., Grigoriev, I. V., Gladden, J. M., Simmons, B. A., Brem, R. B., Arkin, A. P., & Skerker, J. M. (2018). Functional genomics of lipid metabolism in the oleaginous yeast *Rhodosporidium toruloides*. *eLife*, *7*, e32110. https://doi.org/10.7554/eLife.32110

da Silva, G. P., Mack, M., & Contiero, J. (2009). Glycerol: A promising and abundant carbon source for industrial microbiology. *Biotechnology Advances*, *27*(1), 30-39. https://doi.org/10.1016/j.biotechadv.2008.07.006

Daudu, R., Parker, C. W., Singh, N. K., Wood, J. M., Debieu, M., O'Hara, N. B., Mason, C. E., & Venkateswaran, K. (2020). Draft Genome Sequences of *Rhodotorula mucilaginosa* Strains Isolated from the International Space Station. *Microbiology Resource Announcements*, *9*(31), e00570-00520. https://doi.org/doi:10.1128/MRA.00570-20

de Jonge, P., de Jongh, F. C., Meijers, R., Steensma, H. Y., & Scheffers, W. A. (1986). Orthogonal-field-alternation gel electrophoresis banding patterns of DNA from yeasts. *Yeast*, *2*(3), 193-204. https://doi.org/10.1002/yea.320020307

De Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development*, *19*(18), 2100-2110.

Dean, E. J., Davis, J. C., Davis, R. W., & Petrov, D. A. (2008). Pervasive and Persistent Redundancy among Duplicated Genes in Yeast. *PLOS Genetics*, *4*(7), e1000113. https://doi.org/10.1371/journal.pgen.1000113

Deligios, M., Fraumene, C., Abbondio, M., Mannazzu, I., Tanca, A., Addis, M. F., & Uzzau, S. (2015). Draft Genome Sequence of *Rhodotorula mucilaginosa*, an Emergent Opportunistic Pathogen. *Genome Announc*, *3*(2). https://doi.org/10.1128/genomeA.00201-15

Dhar, M. K., Sehgal, S., & Kaul, S. (2012). Structure, replication efficiency and fragility of yeast ARS elements. *Research in Microbiology*, *163*(4), 243-253. https://doi.org/10.1016/j.resmic.2012.03.003

Diamantopoulou, P., Filippousi, R., Antoniou, D., Varfi, E., Xenopoulos, E., Sarris, D., & Papanikolaou, S. (2020). Production of added-value microbial metabolites during growth of yeast strains on media composed of biodiesel-derived crude glycerol and glycerol/xylose blends. *FEMS Microbiology Letters*, *367*(10). https://doi.org/10.1093/femsle/fnaa063

Díaz, T., Fillet, S., Campoy, S., Vázquez, R., Viña, J., Murillo, J., & Adrio, J. L. (2018). Combining evolutionary and metabolic engineering in *Rhodosporidium toruloides* for lipid production with non-detoxified wheat straw hydrolysates. *Applied Microbiology and Biotechnology*, *102*(7), 3287-3300. https://doi.org/10.1007/s00253-018-8810-2

Ding, Y., Owen, M., Le, J., Batalov, S., Chau, K., Kwon, Y. H., Van Der Kraan, L., Bezares-Orin, Z., Zhu, Z., Veeraraghavan, N., Nahas, S., Bainbridge, M., Gleeson, J., Baer, R. J., Bandoli, G., Chambers, C., & Kingsmore, S. F. (2023). Scalable, high quality, whole genome sequencing from archived, newborn, dried blood spots. *npj Genomic Medicine*, *8*(1), 5. https://doi.org/10.1038/s41525-023-00349-w

Dinh, H. V., Suthers, P. F., Chan, S. H. J., Shen, Y., Xiao, T., Deewan, A., Jagtap, S. S., Zhao, H., Rao, C. V., Rabinowitz, J. D., & Maranas, C. D. (2019). A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. *Metab Eng Commun*, *9*, e00101. https://doi.org/10.1016/j.mec.2019.e00101

Doniger, S. W., Kim, H. S., Swain, D., Corcuera, D., Williams, M., Yang, S.-P., & Fay, J. C. (2008). A Catalog of Neutral and Deleterious Polymorphism in

Yeast. *PLOS Genetics*, *4*(8), e1000183. https://doi.org/10.1371/journal.pgen.1000183

Douglass, A. P., O'Brien, C. E., Offei, B., Coughlan, A. Y., Ortiz-Merino, R. A., Butler, G., Byrne, K. P., & Wolfe, K. H. (2019). Coverage-Versus-Length Plots, a Simple Quality Control Step for de Novo Yeast Genome Sequence Assemblies. *G3 Genes|Genomes|Genetics*, *9*(3), 879-887. https://doi.org/10.1534/g3.118.200745

Duan, H., Jones, A. W., Hewitt, T., Mackenzie, A., Hu, Y., Sharp, A., Lewis, D., Mago, R., Upadhyaya, N. M., Rathjen, J. P., Stone, E. A., Schwessinger, B., Figueroa, M., Dodds, P. N., Periyannan, S., & Sperschneider, J. (2022). Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. *Genome Biology*, *23*(1), 84. https://doi.org/10.1186/s13059-022-02658-2

Dujon, B. (2010). Yeast evolutionary genomics. *Nature Reviews Genetics*, *11*(7), 512-524. https://doi.org/10.1038/nrg2811

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., de Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.-M., . . . Souciet, J.-L. (2004). Genome evolution in yeasts. *Nature*, *430*(6995), 35-44. https://doi.org/10.1038/nature02579

El Bilali, H., Callenius, C., Strassner, C., & Probst, L. (2019). Food and nutrition security and sustainability transitions in food systems. *Food and Energy Security*, *8*(2), e00154. https://doi.org/10.1002/fes3.154

Fakankun, I., Fristensky, B., & Levin, D. B. (2021). Genome sequence analysis of the oleaginous yeast, *Rhodotorula diobovata*, and comparison of the carotenogenic and oleaginous pathway genes and gene products with other oleaginous yeasts. *Journal of Fungi*, *7*(4), 320.

Fawzy, S., Osman, A. I., Doran, J., & Rooney, D. W. (2020). Strategies for mitigation of climate change: a review. *Environmental Chemistry Letters*, *18*(6), 2069-2094. https://doi.org/10.1007/s10311-020-01059-w

Fidan, O., & Zhan, J. (2015). Recent advances in engineering yeast for pharmaceutical protein production. *RSC advances*, *5*(105), 86665-86674.

Filippousi, R., Antoniou, D., Tryfinopoulou, P., Nisiotou, A. A., Nychas, G. J., Koutinas, A. A., & Papanikolaou, S. (2019). Isolation, identification and screening of yeasts towards their ability to assimilate biodiesel-derived crude glycerol: microbial production of polyols, endopolysaccharides and lipid. *Journal of Applied Microbiology*, *127*(4), 1080-1100. https://doi.org/10.1111/jam.14373

Firrincieli, A., Otillar, R., Salamov, A., Schmutz, J., Khan, Z., Redman, R., Fleck, N., Lindquist, E., Grigoriev, I., & Doty, S. (2015). Genome sequence of the plant growth promoting endophytic yeast *Rhodotorula graminis* WP1. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.00978

Flach B., L. S., Bolla S. (2022). *Biofuels Annual*. G. A. I. Network.

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Svardal, H., Theofanopoulou, C., . . . Zammit, G. (2022). The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution*, *37*(3), 197-202. https://doi.org/10.1016/j.tree.2021.11.008

Gadanho, M., & Sampaio, J. P. (2002). Polyphasic taxonomy of the basidiomycetous yeast genus *Rhodotorula*: *Rh. glutinis* sensu stricto and *Rh. dairenensis* comb. nov. *FEMS Yeast Research*, *2*(1), 47-58. https://doi.org/10.1111/j.1567-1364.2002.tb00068.x

Gan, H. M., Thomas, B. N., Cavanaugh, N. T., Morales, G. H., Mayers, A. N., Savka, M. A., & Hudson, A. O. (2017). Whole genome sequencing of *Rhodotorula mucilaginosa* isolated from the chewing stick (*Distemonanthus benthamianus)*: insights into *Rhodotorula* phylogeny, mitogenome dynamics and carotenoid biosynthesis. *PeerJ*, *5*, e4030. https://doi.org/10.7717/peerj.4030

Garay, L. A., Sitepu, I. R., Cajka, T., Fiehn, O., Cathcart, E., Fry, R. W., Kanti, A., Joko Nugroho, A., Faulina, S. A., Stephanandra, S., German, J. B., & Boundy-Mills, K. L. (2017). Discovery of synthesis and secretion of polyol esters of fatty acids by four basidiomycetous yeast species in the order Sporidiobolales. *Journal of Industrial Microbiology and Biotechnology*, *44*(6), 923-936. https://doi.org/10.1007/s10295-017-1919-y

Geiselman, G. M., Kirby, J., Landera, A., Otoupal, P., Papa, G., Barcelos, C., Sundstrom, E. R., Das, L., Magurudeniya, H. D., Wehrs, M., Rodriguez, A., Simmons, B. A., Magnuson, J. K., Mukhopadhyay, A., Lee, T. S., George, A., & Gladden, J. M. (2020). Conversion of poplar biomass into high-energy density tricyclic sesquiterpene jet fuel blendstocks. *Microbial Cell Factories*, *19*(1), 208. https://doi.org/10.1186/s12934-020-01456-4

Gibbs, R. A. (2020). The Human Genome Project changed everything. *Nature Reviews Genetics*, *21*(10), 575-576. https://doi.org/10.1038/s41576-020-0275-3

Gilchrist, C., & Stelkens, R. (2019). Aneuploidy in yeast: Segregation error or adaptation mechanism? *Yeast*, *36*(9), 525-539. https://doi.org/10.1002/yea.3427

Goebels, C., Thonn, A., Gonzalez-Hilarion, S., Rolland, O., Moyrand, F., Beilharz, T. H., & Janbon, G. (2013). Introns Regulate Gene Expression in *Cryptococcus neoformans* in a Pab2p Dependent Pathway. *PLOS Genetics*, *9*(8), e1003686. https://doi.org/10.1371/journal.pgen.1003686

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 Genes. *Science*, *274*(5287), 546-567. https://doi.org/doi:10.1126/science.274.5287.546

Gong, G., Liu, L., Zhang, X., & Tan, T. (2019). Multi-omics metabolism analysis on irradiation-induced oxidative stress to *Rhodotorula glutinis*. *Applied Microbiology and Biotechnology*, *103*(1), 361-374. https://doi.org/10.1007/s00253-018-9448-9

Gong, Z., Zhou, W., Shen, H., Zhao, Z. K., Yang, Z., Yan, J., & Zhao, M. (2016). Co-utilization of corn stover hydrolysates and biodiesel-derived glycerol by *Cryptococcus curvatus* for lipid production. *Bioresource Technology*, *219*, 552-558. https://doi.org/10.1016/j.biortech.2016.08.021

Goordial, J., Raymond-Bouchard, I., Riley, R., Ronholm, J., Shapiro, N., Woyke, T., LaButti, K. M., Tice, H., Amirebrahimi, M., Grigoriev, I. V., Greer, C., Bakermans, C., & Whyte, L. (2016). Improved High-Quality Draft Genome Sequence of the Eurypsychrophile *Rhodotorula* sp. JG1b, Isolated from Permafrost in the Hyperarid Upper-Elevation McMurdo Dry Valleys, Antarctica. *Genome Announcements*, *4*(2), e00069-00016. https://doi.org/doi:10.1128/genomeA.00069-16

Gordon, J. L., Byrne, K. P., & Wolfe, K. H. (2009). Additions, Losses, and Rearrangements on the Evolutionary Route from a Reconstructed Ancestor to the Modern *Saccharomyces cerevisiae* Genome. *PLOS Genetics*, *5*(5), e1000485. https://doi.org/10.1371/journal.pgen.1000485

Gorter, F. A., Derks, M. F. L., van den Heuvel, J., Aarts, M. G. M., Zwaan, B. J., de Ridder, D., & de Visser, J. A. G. M. (2017). Genomics of Adaptation Depends on the Rate of Environmental Change in Experimental Yeast Populations. *Molecular Biology and Evolution*, *34*(10), 2613-2626. https://doi.org/10.1093/molbev/msx185

Granger, L. M., Perlot, P., Goma, G., & Pareilleux, A. (1993). Effect of various nutrient limitations on fatty acid production by *Rhodotorula glutinis*. *Applied Microbiology and Biotechnology*, *38*(6), 784-789. https://doi.org/10.1007/BF00167145

Grdic, Z., Rudan, E., & Nizic, M. (2020). Circular economy concept in the context of economic development in EU countries. *Sustainability*, *12*(7), 3060. https://doi.org/10.3390/su12073060

Gresham, D., Usaite, R., Germann, S. M., Lisby, M., Botstein, D., & Regenberg, B. (2010). Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the *GAP1* locus. *Proceedings of the National Academy of Sciences*, *107*(43), 18551-18556. https://doi.org/doi:10.1073/pnas.1014023107

Griffiths, A. J. (1995). Natural plasmids of filamentous fungi. *Microbiological Reviews*, *59*(4), 673-685. https://doi.org/doi:10.1128/mr.59.4.673-685.1995

Guerfali, M., Ayadi, I., Ayadi, W., Smaoui, S., Elhadef, K., Zaghden, H., Jlaiel, L., Sahli, E., Belghith, H., & Gargouri, A. (2022). Concomitant production of multifunctional metabolites on biodiesel-derived crude glycerol by the oleaginous yeast *Rhodotorula babjevae* Y-SL7. *Biomass Conversion and Biorefinery*. https://doi.org/10.1007/s13399-022-03028-5

Guerfali, M., Ayadi, I., Mohamed, N., Ayadi, W., Belghith, H., Bronze, M. R., Ribeiro, M. H. L., & Gargouri, A. (2019). Triacylglycerols accumulation and glycolipids secretion by the oleaginous yeast *Rhodotorula babjevae* Y-SL7: Structural identification and biotechnological applications. *Bioresource Technology*, *273*, 326-334. https://doi.org/10.1016/j.biortech.2018.11.036

Guillamón, J. M., & Barrio, E. (2017). Genetic Polymorphism in Wine Yeasts: Mechanisms and Methods for Its Detection [Review]. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00806

Halkos, G. E., & Gkampoura, E.-C. (2020). Reviewing usage, potentials, and limitations of renewable energy sources. *Energies*, *13*(11), 2906.

Hamamoto, M., Sugiyama, J., GOTO, S., & KOMAGATA, K. (1986). Numerical taxonomy based on the electrophoretic mobility of enzymes in the genera *Rhodosporidium*, *Cystofilobasidium*, and *Rhodotorula*. *The Journal of General and Applied Microbiology*, *32*(2), 89-99.

Hao, W. (2022). From Genome Variation to Molecular Mechanisms: What we Have Learned From Yeast Mitochondrial Genomes? [Mini Review]. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.806575

Harari, Y., Ram, Y., & Kupiec, M. (2018). Frequent ploidy changes in growing yeast cultures. *Current Genetics*, *64*(5), 1001-1004. https://doi.org/10.1007/s00294-018-0823-y

Hawksworth, D. L., Crous, P. W., Redhead, S. A., Reynolds, D. R., Samson, R. A., Seifert, K. A., Taylor, J. W., Wingfield, M. J., Abaci, O., Aime, C., Asan, A., Bai, F. Y., de Beer, Z. W., Begerow, D., Berikten, D., Boekhout, T., Buchanan, P. K., Burgess, T., Buzina, W., . . . Zhang, N. (2011). The amsterdam declaration on fungal nomenclature. *IMA Fungus*, *2*(1), 105-112. https://doi.org/10.5598/imafungus.2011.02.01.14

Hoang, N. T., & Kanemoto, K. (2021). Mapping the deforestation footprint of nations reveals growing threat to tropical forests. *Nat Ecol Evol*, *5*(6), 845-853. https://doi.org/10.1038/s41559-021-01417-z

Hu, J., & Ji, L. (2016). Draft Genome Sequences of *Rhodosporidium toruloides* Strains ATCC 10788 and ATCC 10657 with Compatible Mating Types. *Genome Announcements*, *4*(2), e00098-00016. https://doi.org/doi:10.1128/genomeA.00098-16

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801-811. https://doi.org/10.1016/j.humimm.2021.02.012

Hull, R. M., King, M., Pizza, G., Krueger, F., Vergara, X., & Houseley, J. (2019). Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLOS Biology*, *17*(12), e3000471. https://doi.org/10.1371/journal.pbio.3000471

Igreja, W. S., Maia, F. d. A., Lopes, A. S., & Chisté, R. C. (2021). Biotechnological production of carotenoids using low cost-substrates is influenced by

cultivation parameters: A review. *International Journal of Molecular Sciences*, *22*(16), 8819. https://doi.org/10.3390/ijms22168819

Iwakiri, R., Eguchi, S., Noda, Y., Adachi, H., & Yoda, K. (2005). Isolation and structural analysis of efficient autonomously replicating sequences (ARSs) of the yeast *Candida utilis*. *Yeast*, *22*(13), 1049-1060. https://doi.org/10.1002/yea.1296

Jagtap, S. S., Deewan, A., Liu, J. J., Walukiewicz, H. E., Yun, E. J., Jin, Y. S., & Rao, C. V. (2021). Integrating transcriptomic and metabolomic analysis of the oleaginous yeast *Rhodosporidium toruloides* IFO0880 during growth under different carbon sources. *Appl Microbiol Biotechnol*, *105*(19), 7411-7425. https://doi.org/10.1007/s00253-021-11549-8

Jagtap, S. S., & Rao, C. V. (2018). Production of d-arabitol from d-xylose by the oleaginous yeast *Rhodosporidium toruloides* IFO0880. *Applied Microbiology and Biotechnology*, *102*(1), 143-151. https://doi.org/10.1007/s00253-017-8581-1

Jiao, X., Zhang, Q., Zhang, S., Yang, X., Wang, Q., & Zhao, Z. K. (2018). Efficient co-expression of multiple enzymes from a single promoter mediated by virus 2A sequence in the oleaginous yeast *Rhodosporidium toruloides*. *FEMS Yeast Research*, *18*(8). https://doi.org/10.1093/femsyr/foy086

Johns, A. M. B., Love, J., & Aves, S. J. (2016). Four Inducible Promoters for Controlled Gene Expression in the Oleaginous Yeast *Rhodotorula toruloides* [Original Research]. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01666

Jones, A. D., Boundy-Mills, K. L., Barla, G. F., Kumar, S., Ubanwa, B., & Balan, V. (2019). Microbial Lipid Alternatives to Plant Lipids. In V. Balan (Ed.), *Microbial Lipid Production: Methods and Protocols* (pp. 1-32). Springer New York. https://doi.org/10.1007/978-1-4939-9484-7_1

Juanssilfero, A. B., Kahar, P., Amza, R. L., Miyamoto, N., Otsuka, H., Matsumoto, H., Kihira, C., Thontowi, A., Yopi, Ogino, C., Prasetya, B., & Kondo, A. (2018). Selection of oleaginous yeasts capable of high lipid accumulation during challenges from inhibitory chemical compounds. *Biochemical Engineering Journal*, *137*, 182-191. https://doi.org/10.1016/j.bej.2018.05.024

Juneau, K., Palm, C., Miranda, M., & Davis, R. W. (2007). High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proceedings of the National Academy of Sciences*, *104*(5), 1522-1527. https://doi.org/doi:10.1073/pnas.0610354104

Kajiya-Kanegae, H., Nagasaki, H., Kaga, A., Hirano, K., Ogiso-Tanaka, E., Matsuoka, M., Ishimori, M., Ishimoto, M., Hashiguchi, M., Tanaka, H., Akashi, R., Isobe, S., & Iwata, H. (2021). Whole-genome sequence diversity and association analysis of 198 soybean accessions in mini-core collections. *DNA Research*, *28*(1). https://doi.org/10.1093/dnares/dsaa032

Kaneda, T. (1991). Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiological Reviews*, *55*(2), 288-302. https://doi.org/doi:10.1128/mr.55.2.288-302.1991

Karlsson, H., Ahlgren, S., Sandgren, M., Passoth, V., Wallberg, O., & Hansson, P. A. (2016). A systems analysis of biodiesel production from wheat straw using oleaginous yeast: process design, mass and energy balances. *Biotechnol Biofuels*, *9*, 229. https://doi.org/10.1186/s13068-016-0640-9

Karlsson, H., Ahlgren, S., Sandgren, M., Passoth, V., Wallberg, O., & Hansson, P. A. (2017). Greenhouse gas performance of biochemical biodiesel production from straw: soil organic carbon changes and time-dependent climate impact. *Biotechnol Biofuels*, *10*, 217. https://doi.org/10.1186/s13068-017-0907-9

Kavanaugh, L. A., Fraser, J. A., & Dietrich, F. S. (2006). Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. *Molecular biology and evolution*, *23*(10), 1879-1890.

Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, *428*(6983), 617-624. https://doi.org/10.1038/nature02424

Kim, J., Coradetti, S. T., Kim, Y. M., Gao, Y., Yaegashi, J., Zucker, J. D., Munoz, N., Zink, E. M., Burnum-Johnson, K. E., Baker, S. E., Simmons, B. A., Skerker, J. M., Gladden, J. M., & Magnuson, J. K. (2021). Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in *Rhodosporidium toruloides*. *Frontiers in Bioengineering Biotechnology*, *8*, 612832. https://doi.org/10.3389/fbioe.2020.612832

Kingan, S. B., Heaton, H., Cudini, J., Lambert, C. C., Baybayan, P., Galvin, B. D., Durbin, R., Korlach, J., & Lawniczak, M. K. N. (2019). A High-Quality De novo Genome Assembly from a Single Mosquito Using PacBio Sequencing. *Genes*, *10*(1), 62. https://www.mdpi.com/2073-4425/10/1/62

Klein, M., Swinnen, S., Thevelein, J. M., & Nevoigt, E. (2017). Glycerol metabolism and transport in yeast and fungi: established knowledge and ambiguities. *Environ Microbiol*, *19*(3), 878-893. https://doi.org/10.1111/1462-2920.13617

Koh, C. M. J., Liu, Y., Moehninsi, Du, M., & Ji, L. (2014). Molecular characterization of KU70 and KU80 homologues and exploitation of a KU70-deficient mutant for improving gene deletion frequency in *Rhodosporidium toruloides*. *BMC Microbiology*, *14*(1), 50. https://doi.org/10.1186/1471-2180-14-50

Koh, L. P., & Wilcove, D. S. (2008). Is oil palm agriculture really destroying tropical biodiversity? *Conservation Letters*, *1*(2), 60-64. https://doi.org/10.1111/j.1755-263X.2008.00011.x

Kumar, S., Kushwaha, H., Bachhawat, A. K., Raghava, G. P. S., & Ganesan, K. (2012). Genome Sequence of the Oleaginous Red Yeast *Rhodosporidium*

toruloides MTCC 457. *Eukaryotic Cell*, *11*(8), 1083-1084. https://doi.org/doi:10.1128/EC.00156-12

Kuzmin, E., Taylor, J. S., & Boone, C. (2022). Retention of duplicated genes in evolution. *Trends in Genetics*, *38*(1), 59-72. https://doi.org/10.1016/j.tig.2021.06.016

Lallemand, T., Leduc, M., Landès, C., Rizzon, C., & Lerat, E. (2020). An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. *Genes*, *11*(9), 1046. https://doi.org/10.3390/genes11091046

Lang, D., Zhang, S., Ren, P., Liang, F., Sun, Z., Meng, G., Tan, Y., Li, X., Lai, Q., Han, L., Wang, D., Hu, F., Wang, W., & Liu, S. (2020). Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore. *GigaScience*, *9*(12). https://doi.org/10.1093/gigascience/giaa123

Legrand, M., Jaitly, P., Feri, A., d'Enfert, C., & Sanyal, K. (2019). *Candida albicans*: An Emerging Yeast Model to Study Eukaryotic Genome Plasticity. *Trends in Genetics*, *35*(4), 292-307. https://doi.org/10.1016/j.tig.2019.01.005

Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., Axelrod, N., Huang, J., Kirkness, E. F., Denisov, G., Lin, Y., MacDonald, J. R., Pang, A. W. C., Shago, M., Stockwell, T. B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S. A., . . . Venter, J. C. (2007). The Diploid Genome Sequence of an Individual Human. *PLOS Biology*, *5*(10), e254. https://doi.org/10.1371/journal.pbio.0050254

Li, C.-J., Zhao, D., Cheng, P., Zheng, L., & Yu, G.-H. (2020). Genomics and lipidomics analysis of the biotechnologically important oleaginous red yeast *Rhodotorula glutinis* ZHK provides new insights into its lipid and carotenoid metabolism. *BMC Genomics*, *21*(1), 834. https://doi.org/10.1186/s12864-020-07244-z

Liao, J. C., Mi, L., Pontrelli, S., & Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology*, *14*(5), 288-304. https://doi.org/10.1038/nrmicro.2016.32

Libkind, D., Čadež, N., Opulente, D. A., Langdon, Q. K., Rosa, C. A., Sampaio, J. P., Gonçalves, P., Hittinger, C. T., & Lachance, M. A. (2020). Towards yeast taxogenomics: lessons from novel species descriptions based on complete genome sequences. *FEMS Yeast Research*, *20*(6). https://doi.org/10.1093/femsyr/foaa042

Lin, X., Wang, Y., Zhang, S., Zhu, Z., Zhou, Y. J., Yang, F., Sun, W., Wang, X., & Zhao, Z. K. (2014). Functional integration of multiple genes into the genome of the oleaginous yeast *Rhodosporidium toruloides*. *FEMS Yeast Research*, *14*(4), 547-555. https://doi.org/10.1111/1567-1364.12140

Liu, D., Geiselman, G. M., Coradetti, S., Cheng, Y.-F., Kirby, J., Prahl, J.-P., Jacobson, O., Sundstrom, E. R., Tanjore, D., Skerker, J. M., & Gladden, J.

(2020). Exploiting nonionic surfactants to enhance fatty alcohol production in *Rhodosporidium toruloides*. *Biotechnology and Bioengineering*, *117*(5), 1418-1425. https://doi.org/10.1002/bit.27285

Liu, H., Jiao, X., Wang, Y., Yang, X., Sun, W., Wang, J., Zhang, S., & Zhao, Z. K. (2017). Fast and efficient genetic transformation of oleaginous yeast *Rhodosporidium toruloides* by using electroporation. *FEMS Yeast Research*, *17*(2). https://doi.org/10.1093/femsyr/fox017

Liu, M., Zhang, X., & Tan, T. (2016). The effect of amino acids on lipid production and nutrient removal by *Rhodotorula glutinis* cultivation in starch wastewater. *Bioresource Technology*, *218*, 712-717. https://doi.org/10.1016/j.biortech.2016.07.027

Liu, X., Zhang, Y., Liu, H., Jiao, X., Zhang, Q., Zhang, S., & Zhao, Z. K. (2019). RNA interference in the oleaginous yeast *Rhodosporidium toruloides*. *FEMS Yeast Research*, *19*(3). https://doi.org/10.1093/femsyr/foz031

Liu, Y., Koh, C. M. J., Sun, L., Hlaing, M. M., Du, M., Peng, N., & Ji, L. (2013). Characterization of glyceraldehyde-3-phosphate dehydrogenase gene Rt*GPD1* and development of genetic transformation method by dominant selection in oleaginous yeast *Rhodosporidium toruloides*. *Applied Microbiology and Biotechnology*, *97*(2), 719-729. https://doi.org/10.1007/s00253-012-4223-9

Liu, Y., Koh, C. M. J., Yap, S. A., Cai, L., & Ji, L. (2021). Understanding and exploiting the fatty acid desaturation system in *Rhodotorula toruloides*. *Biotechnology for Biofuels*, *14*(1), 73. https://doi.org/10.1186/s13068-021-01924-y

Liu, Y., Yap, S. A., Koh, C. M. J., & Ji, L. (2016). Developing a set of strong intronic promoters for robust metabolic engineering in oleaginous *Rhodotorula* (*Rhodosporidium*) yeast species. *Microbial Cell Factories*, *15*(1), 200. https://doi.org/10.1186/s12934-016-0600-x

Lopes, H. J. S., Bonturi, N., Kerkhoven, E. J., Miranda, E. A., & Lahtvee, P.-J. (2020). C/N ratio and carbon source-dependent lipid production profiling in *Rhodotorula toruloides*. *Applied Microbiology and Biotechnology*, *104*(6), 2639-2649. https://doi.org/10.1007/s00253-020-10386-5

Luo, W., Yan, J., Luo, S., Liu, W., Xie, D., & Jiang, B. (2023). A chromosome-level reference genome of the wax gourd (*Benincasa hispida*). *Scientific Data*, *10*(1), 78. https://doi.org/10.1038/s41597-023-01986-7

Marchant, R., & Banat, I. M. (2012). Microbial biosurfactants: challenges and opportunities for future exploitation. *Trends in Biotechnology*, *30*(11), 558-565. https://doi.org/10.1016/j.tibtech.2012.07.003

Matsuzawa, T., Ohashi, T., Hosomi, A., Tanaka, N., Tohda, H., & Takegawa, K. (2010). The gld1+ gene encoding glycerol dehydrogenase is required for glycerol metabolism in *Schizosaccharomyces pombe*. *Applied Microbiology and Biotechnology*, *87*(2), 715-727. https://doi.org/10.1007/s00253-010-2586-3

Menegazzo, M. L., & Fonseca, G. G. (2019). Biomass recovery and lipid extraction processes for microalgae biofuels production: A review. *Renewable and Sustainable Energy Reviews*, *107*, 87-107. https://doi.org/10.1016/j.rser.2019.01.064

Mirzaei Seveiri, R., Hamidi, M., Delattre, C., Sedighian, H., Pierre, G., Rahmani, B., Darzi, S., Brasselet, C., Karimitabar, F., Razaghpoor, A., & Amani, J. (2020). Characterization and Prospective Applications of the Exopolysaccharides Produced by *Rhodosporidium babjevae*. *Adv Pharm Bull*, *10*(2), 254-263. https://doi.org/10.34172/apb.2020.030

Mishra, A., Forche, A., & Anderson, M. Z. (2021). Parasexuality of Candida Species. *Front Cell Infect Microbiol*, *11*, 796929. https://doi.org/10.3389/fcimb.2021.796929

Møller, H. D., Parsons, L., Jørgensen, T. S., Botstein, D., & Regenberg, B. (2015). Extrachromosomal circular DNA is common in yeast. *Proceedings of the National Academy of Sciences*, *112*(24), E3114-E3122. https://doi.org/doi:10.1073/pnas.1508825112

Morin, N., Calcas, X., Devillers, H., Durrens, P., Sherman, D. J., Nicaud, J.-M., & Neuvéglise, C. (2014). Draft Genome Sequence of *Rhodosporidium toruloides* CECT1137, an Oleaginous Yeast of Biotechnological Interest. *Genome Announcements*, *2*(4), e00641-00614. https://doi.org/doi:10.1128/genomeA.00641-14

Morrow, C. A., & Fraser, J. A. (2009). Sexual reproduction and dimorphism in the pathogenic basidiomycetes. *FEMS Yeast Res*, *9*(2), 161-177. https://doi.org/10.1111/j.1567-1364.2008.00475.x

Mussagy, C. U., Guimarães, A. A. C., Rocha, L. V. F., Winterburn, J., Santos-Ebinuma, V. d. C., & Pereira, J. F. B. (2021). Improvement of carotenoids production from *Rhodotorula glutinis* CCT-2186. *Biochemical Engineering Journal*, *165*, 107827. https://doi.org/10.1016/j.bej.2020.107827

Nagaraj, Y. N., Burkina, V., Okmane, L., Blomqvist, J., Rapoport, A., Sandgren, M., Pickova, J., Sampels, S., & Passoth, V. (2022). Identification, Quantification and Kinetic Study of Carotenoids and Lipids in *Rhodotorula toruloides* CBS 14 Cultivated on Wheat Straw Hydrolysate. *Fermentation*, *8*(7), 300. https://www.mdpi.com/2311-5637/8/7/300

Nakao, Y., Kanamori, T., Itoh, T., Kodama, Y., Rainieri, S., Nakamura, N., Shimonaga, T., Hattori, M., & Ashikari, T. (2009). Genome Sequence of the Lager Brewing Yeast, an Interspecies Hybrid. *DNA Research*, *16*(2), 115-129. https://doi.org/10.1093/dnares/dsp003

Nevers, A., Doyen, A., Malabat, C., Néron, B., Kergrohen, T., Jacquier, A., & Badis, G. (2018). Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. *Nucleic Acids Research*, *46*(12), 6009-6025. https://doi.org/10.1093/nar/gky342

Nora, L. C., Cassiano, M. H. A., Santana, Í. P., Guazzaroni, M.-E., Silva-Rocha, R., & da Silva, R. R. (2023). Mining novel cis-regulatory elements from the

emergent host *Rhodosporidium toruloides* using transcriptomic data [Technology and Code]. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.1069443

Olsen, R.-A., Bunikis, I., Tiukova, I., Holmberg, K., Lötstedt, B., Pettersson, O. V., Passoth, V., Käller, M., & Vezzi, F. (2015). De novo assembly of *Dekkera bruxellensis*: a multi technology approach using short and long-read sequencing and optical mapping. *GigaScience*, *4*(1). https://doi.org/10.1186/s13742-015-0094-1

Otoupal, P. B., Ito, M., Arkin, A. P., Magnuson, J. K., Gladden, J. M., & Skerker, J. M. (2019). Multiplexed CRISPR-Cas9-Based Genome Editing of *Rhodosporidium toruloides*. *mSphere*, *4*(2), e00099-00019. https://doi.org/doi:10.1128/mSphere.00099-19

Papanikolaou, S., & Aggelis, G. (2020). Microbial products from wastes and residues. *FEMS Microbiol Lett*, *367*(19). https://doi.org/10.1093/femsle/fnaa156

Pappas, F. (2023). *Genomic scaffolding and comparative genomics analysis of 23 oleaginous yeast strains belonging to the Rhodotorula genus.* SLU]. Uppsala. http://urn.kb.se/resolve?urn=urn:nbn:se:slu:epsilon-s-18788

Passoth, V., Brandenburg, J., Chmielarz, M., Martín-Hernández, G. C., Nagaraj, Y., Müller, B., & Blomqvist, J. (2023). Oleaginous yeasts for biochemicals, biofuels and food from lignocellulose-hydrolysate and crude glycerol. *Yeast*(1-13). https://doi.org/10.1002/yea.3838

Passoth, V., & Sandgren, M. (2019). Biofuel production from straw hydrolysates: current achievements and perspectives. *Appl Microbiol Biotechnol*, *103*(13), 5105-5116. https://doi.org/10.1007/s00253-019-09863-3

Patel, A., Arora, N., Sartaj, K., Pruthi, V., & Pruthi, P. A. (2016). Sustainable biodiesel production from oleaginous yeasts utilizing hydrolysates of various non-edible lignocellulosic biomasses. *Renewable and Sustainable Energy Reviews*, *62*, 836-855. https://doi.org/10.1016/j.rser.2016.05.014

Patel A, K. D., Rova E, Katapodis P, Rova U, Christakopoulos P, Matsakas L. . (2020). An Overview of Potential Oleaginous Microorganisms and Their Role in Biodiesel and Omega-3 Fatty Acid-Based Industries. *Microorganisms*, *8*(3), 434.

Patra, P., Das, M., Kundu, P., & Ghosh, A. (2021). Recent advances in systems and synthetic biology approaches for developing novel cell-factories in non-conventional yeasts. *Biotechnology Advances*, *47*, 107695. https://doi.org/10.1016/j.biotechadv.2021.107695

Paul, D., Magbanua, Z., Arick, M., French, T., Bridges, S. M., Burgess, S. C., & Lawrence, M. L. (2014). Genome Sequence of the Oleaginous Yeast *Rhodotorula glutinis* ATCC 204091. *Genome Announcements*, *2*(1), e00046-00014. https://doi.org/doi:10.1128/genomeA.00046-14

Peris, D., Arias, A., Orlić, S., Belloch, C., Pérez-Través, L., Querol, A., & Barrio, E. (2017). Mitochondrial introgression suggests extensive ancestral hybridization events among Saccharomyces species. *Molecular*

Phylogenetics and Evolution, *108*, 49-60. https://doi.org/10.1016/j.ympev.2017.02.008

Pervez, M. T., Hasnain, M. J. u., Abbas, S. H., Moustafa, M. F., Aslam, N., & Shah, S. S. M. (2022). A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *BioMed Research International*, *2022*, 3457806. https://doi.org/10.1155/2022/3457806

Pinheiro, M. J., Bonturi, N., Belouah, I., Miranda, E. A., & Lahtvee, P.-J. (2020). Xylose Metabolism and the Effect of Oxidative Stress on Lipid and Carotenoid Production in *Rhodotorula toruloides*: Insights for Future Biorefinery [Original Research]. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.01008

Pinu, F. R., Beale, D. J., Paten, A. M., Kouremenos, K., Swarup, S., Schirra, H. J., & Wishart, D. (2019). Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites*, *9*(4). https://doi.org/10.3390/metabo9040076

Poverennaya, I. V., & Roytberg, M. A. (2020). Spliceosomal Introns: Features, Functions, and Evolution. *Biochemistry (Moscow)*, *85*(7), 725-734. https://doi.org/10.1134/S0006297920070019

Rakicka, M., Lazar, Z., Dulermo, T., Fickers, P., & Nicaud, J. M. (2015). Lipid production by the oleaginous yeast *Yarrowia lipolytica* using industrial by-products under different culture conditions. *Biotechnology for Biofuels*, *8*(1), 104. https://doi.org/10.1186/s13068-015-0286-z

Rapoport, A., Guzhova, I., Bernetti, L., Buzzini, P., Kieliszek, M., & Kot, A. M. (2021). Carotenoids and Some Other Pigments from Fungi and Yeasts. *Metabolites*, *11*(2), 92. https://www.mdpi.com/2218-1989/11/2/92

Ratledge, C. (2014). The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: a reappraisal and unsolved problems. *Biotechnology Letters*, *36*(8), 1557-1568. https://doi.org/10.1007/s10529-014-1532-3

Ratledge, C., & Wynn, J. P. (2002). The Biochemistry and Molecular Biology of Lipid Accumulation in Oleaginous Microorganisms. In A. I. Laskin, J. W. Bennett, & G. M. Gadd (Eds.), *Advances in Applied Microbiology* (Vol. 51, pp. 1-52). Academic Press. https://doi.org/10.1016/S0065-2164(02)51000-5

Rodriguez, A., Ersig, N., Geiselman, G. M., Seibel, K., Simmons, B. A., Magnuson, J. K., Eudes, A., & Gladden, J. M. (2019). Conversion of depolymerized sugars and aromatics from engineered feedstocks by two oleaginous red yeasts. *Bioresource Technology*, *286*, 121365. https://doi.org/10.1016/j.biortech.2019.121365

Rodriguez, S., Denby, C. M., Van Vu, T., Baidoo, E. E. K., Wang, G., & Keasling, J. D. (2016). ATP citrate lyase mediated cytosolic acetyl-CoA biosynthesis increases mevalonate production in *Saccharomyces cerevisiae*. *Microbial Cell Factories*, *15*(1), 48. https://doi.org/10.1186/s12934-016-0447-1

Sambles, C., Middelhaufe, S., Soanes, D., Kolak, D., Lux, T., Moore, K., Matoušková, P., Parker, D., Lee, R., Love, J., & Aves, S. J. (2017). Genome sequence of the oleaginous yeast *Rhodotorula toruloides* strain CGMCC 2.1609. *Genomics Data*, *13*, 1-2. https://doi.org/10.1016/j.gdata.2017.05.009

Sampaio, J. P. (2011a). Chapter 127 - Rhodosporidium Banno (1967). In C. P. Kurtzman, J. W. Fell, & T. Boekhout (Eds.), *The Yeasts (Fifth Edition)* (pp. 1523-1539). Elsevier. https://doi.org/10.1016/B978-0-444-52149-1.00127-0

Sampaio, J. P. (2011b). Chapter 155 - Rhodotorula Harrison (1928). In C. P. Kurtzman, J. W. Fell, & T. Boekhout (Eds.), *The Yeasts (Fifth Edition)* (pp. 1873-1927). Elsevier. https://doi.org/10.1016/B978-0-444-52149-1.00155-5

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463-5467. https://doi.org/doi:10.1073/pnas.74.12.5463

Schindler, D. (2020). Genetic engineering and synthetic genomics in yeast to understand life and boost biotechnology. *Bioengineering*, *7*(4), 137. https://doi.org/10.3390/bioengineering7040137

Schirman, D., Yakhini, Z., Pilpel, Y., & Dahan, O. (2021). A broad analysis of splicing regulation in yeast using a large library of synthetic introns. *PLOS Genetics*, *17*(9), e1009805. https://doi.org/10.1371/journal.pgen.1009805

Schmidt, J. H. (2015). Life cycle assessment of five vegetable oils. *Journal of Cleaner Production*, *87*, 130-138. https://doi.org/10.1016/j.jclepro.2014.10.011

Schultz, J. C., Cao, M., & Zhao, H. (2019). Development of a CRISPR/Cas9 system for high efficiency multiplexed gene deletion in *Rhodosporidium toruloides*. *Biotechnology and Bioengineering*, *116*(8), 2103-2109. https://doi.org/10.1002/bit.27001

Sen, D., Paul, K., Saha, C., Mukherjee, G., Nag, M., Ghosh, S., Das, A., Seal, A., & Tripathy, S. (2019). A unique life-strategy of an endophytic yeast *Rhodotorula mucilaginosa* JGTA-S1—a comparative genomics viewpoint. *DNA Research*, *26*(2), 131-146. https://doi.org/10.1093/dnares/dsy044

Sen, S., Borah, S. N., Bora, A., & Deka, S. (2017). Production, characterization, and antifungal activity of a biosurfactant produced by *Rhodotorula babjevae* YS3. *Microbial Cell Factories*, *16*(1), 95. https://doi.org/10.1186/s12934-017-0711-z

Sen, S., Borah, S. N., Kandimalla, R., Bora, A., & Deka, S. (2020). Sophorolipid Biosurfactant Can Control Cutaneous Dermatophytosis Caused by *Trichophyton mentagrophytes* [Original Research]. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00329

Shapaval, V., Brandenburg, J., Blomqvist, J., Tafintseva, V., Passoth, V., Sandgren, M., & Kohler, A. (2019). Biochemical profiling, prediction of total lipid

content and fatty acid profile in oleaginous yeasts by FTIR spectroscopy. *Biotechnol Biofuels*, *12*, 140. https://doi.org/10.1186/s13068-019-1481-0

Shayesteh, H., Vadiveloo, A., Bahri, P. A., & Moheimani, N. R. (2022). Long term outdoor microalgal phycoremediation of anaerobically digested abattoir effluent. *Journal of Environmental Management*, *323*, 116322. https://doi.org/10.1016/j.jenvman.2022.116322

Sinclair, D. A., & Guarente, L. (1997). Extrachromosomal rDNA Circles— A Cause of Aging in Yeast. *Cell*, *91*(7), 1033-1042. https://doi.org/10.1016/S0092-8674(00)80493-6

Sineli, P. E., Maza, D. D., Aybar, M. J., Figueroa, L. I. C., & Viñarta, S. C. (2022). Bioconversion of sugarcane molasses and waste glycerol on single cell oils for biodiesel by the red yeast *Rhodotorula glutinis* R4 from Antarctica. *Energy Conversion and Management: X*, *16*, 100331. https://doi.org/10.1016/j.ecmx.2022.100331

Singh, P., & Ahi, E. P. (2022). The importance of alternative splicing in adaptive evolution. *Molecular Ecology*, *31*(7), 1928-1938. https://doi.org/10.1111/mec.16377

Sipiczki, M. (2019). Yeast two- and three-species hybrids and high-sugar fermentation. *Microbial Biotechnology*, *12*(6), 1101-1108. https://doi.org/10.1111/1751-7915.13390

Smukowski Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., & Dunham, M. J. (2017). Loss of Heterozygosity Drives Adaptation in Hybrid Yeast. *Molecular Biology and Evolution*, *34*(7), 1596-1612. https://doi.org/10.1093/molbev/msx098

Sprague, G. F., & Cronan, J. E. (1977). Isolation and characterization of *Saccharomyces cerevisiae* mutants defective in glycerol catabolism. *J Bacteriol*, *129*(3), 1335-1342. https://doi.org/10.1128/jb.129.3.1335-1342.1977

Steensels, J., Gallone, B., & Verstrepen, K. J. (2021). Interspecific hybridization as a driver of fungal evolution and adaptation. *Nature Reviews Microbiology*, *19*(8), 485-500. https://doi.org/10.1038/s41579-021-00537-4

Strope, P. K., Kozmin, S. G., Skelly, D. A., Magwene, P. M., Dietrich, F. S., & McCusker, J. H. (2015). 2μ plasmid in *Saccharomyces* species and in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, *15*(8). https://doi.org/10.1093/femsyr/fov090

Su, X., Zhou, M., Hu, P., Xiao, Y., Wang, Z., Mei, R., Hashmi, M. Z., Lin, H., Chen, J., & Sun, F. (2019). Whole-genome sequencing of an acidophilic *Rhodotorula* sp. ZM1 and its phenol-degrading capability under acidic conditions. *Chemosphere*, *232*, 76-86. https://doi.org/10.1016/j.chemosphere.2019.05.195

Swinnen, S., Klein, M., Carrillo, M., McInnes, J., Nguyen, H. T. T., & Nevoigt, E. (2013). Re-evaluation of glycerol utilization in *Saccharomyces cerevisiae*: characterization of an isolate that grows on glycerol without supporting

supplements. *Biotechnol Biofuels*, *6*(1), 157. https://doi.org/10.1186/1754-6834-6-157

Tang, W., Zhang, S., Tan, H., & Zhao, Z. K. (2010). Molecular Cloning and Characterization of a Malic Enzyme Gene from the Oleaginous Yeast *Lipomyces starkeyi*. *Molecular Biotechnology*, *45*(2), 121-128. https://doi.org/10.1007/s12033-010-9255-8

Tiukova, I. A., Brandenburg, J., Blomqvist, J., Sampels, S., Mikkelsen, N., Skaugen, M., Arntzen, M. Ø., Nielsen, J., Sandgren, M., & Kerkhoven, E. J. (2019). Proteome analysis of xylose metabolism in *Rhodotorula toruloides* during lipid production. *Biotechnology for Biofuels*, *12*(1), 137. https://doi.org/10.1186/s13068-019-1478-8

Tiukova, I. A., Pettersson, M. E., Hoeppner, M. P., Olsen, R.-A., Käller, M., Nielsen, J., Dainat, J., Lantz, H., Söderberg, J., & Passoth, V. (2019). Chromosomal genome assembly of the ethanol production strain CBS 11270 indicates a highly dynamic genome structure in the yeast species *Brettanomyces bruxellensis*. *PLoS One*, *14*(5), e0215077. https://doi.org/10.1371/journal.pone.0215077

Tiukova, I. A., Prigent, S., Nielsen, J., Sandgren, M., & Kerkhoven, E. J. (2019). Genome-scale model of *Rhodotorula toruloides* metabolism [https://doi.org/10.1002/bit.27162]. *Biotechnology and Bioengineering*, *116*(12), 3396-3408. https://doi.org/10.1002/bit.27162

Tkavc, R., Matrosova, V. Y., Grichenko, O. E., Gostinčar, C., Volpe, R. P., Klimenkova, P., Gaidamakova, E. K., Zhou, C. E., Stewart, B. J., Lyman, M. G., Malfatti, S. A., Rubinfeld, B., Courtot, M., Singh, J., Dalgard, C. L., Hamilton, T., Frey, K. G., Gunde-Cimerman, N., Dugan, L., & Daly, M. J. (2018). Prospects for Fungal Bioremediation of Acidic Radioactive Waste Sites: Characterization and Genome Sequence of *Rhodotorula taiwanensis* MD1149 [Original Research]. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.02528

Tokuda, N., Terada, Tomoki P., & Sasai, M. (2012). Dynamical Modeling of Three-Dimensional Genome Organization in Interphase Budding Yeast. *Biophysical Journal*, *102*(2), 296-304. https://doi.org/10.1016/j.bpj.2011.12.005

Tom, G. D., Viswanath-Reddy, M., & Howe, H. B., Jr. (1978). Effect of carbon source on enzymes involved in glycerol metabolism in *Neurospora crassa*. *Arch Microbiol*, *117*(3), 259-263. https://doi.org/10.1007/BF00738544

Touchette, D., Altshuler, I., Gostincar, C., Zalar, P., Raymond-Bouchard, I., Zajc, J., McKay, C. P., Gunde-Cimerman, N., & Whyte, L. G. (2022). Novel Antarctic yeast adapts to cold by switching energy metabolism and increasing small RNA synthesis. *ISME J*, *16*(1), 221-232. https://doi.org/10.1038/s41396-021-01030-9

Tran, T. N., Ngo, D.-H., Nguyen, N. T., & Ngo, D.-N. (2019). Draft genome sequence data of *Rhodosporidium toruloides* VN1, a strain capable of

producing natural astaxanthin. *Data in Brief*, *26*, 104443. https://doi.org/10.1016/j.dib.2019.104443

Tsai, Y.-Y., Ohashi, T., Kanazawa, T., Polburee, P., Misaki, R., Limtong, S., & Fujiyama, K. (2017). Development of a sufficient and effective procedure for transformation of an oleaginous yeast, *Rhodosporidium toruloides* DMKU3-TK16. *Current Genetics*, *63*(2), 359-371. https://doi.org/10.1007/s00294-016-0629-8

Tully, M., & Gilbert, H. J. (1985). Transformation of *Rhodosporidium toruloides*. *Gene*, *36*(3), 235-240. https://doi.org/10.1016/0378-1119(85)90178-7

Tuon, F. F., & Costa, S. F. (2008). Infección por Rhodotorula. Revisión de 128 casos. *Revista Iberoamericana de Micología*, *25*(3), 135-140. https://doi.org/10.1016/S1130-1406(08)70032-9

Usher, J. (2019). The mechanisms of mating in pathogenic fungi—a plastic trait. *Genes*, *10*(10), 831. https://doi.org/10.3390/genes10100831

Vadiveloo, A., Foster, L., Kwambai, C., Bahri, P. A., & Moheimani, N. R. (2021). Microalgae cultivation for the treatment of anaerobically digested municipal centrate (ADMC) and anaerobically digested abattoir effluent (ADAE). *Science of The Total Environment*, *775*, 145853. https://doi.org/10.1016/j.scitotenv.2021.145853

Venkata Subhash, G., & Venkata Mohan, S. (2011). Biodiesel production from isolated oleaginous fungi *Aspergillus* sp. using corncob waste liquor as a substrate. *Bioresource Technology*, *102*(19), 9286-9290. https://doi.org/10.1016/j.biortech.2011.06.084

Vidakovic, A., Huyben, D., Sundh, H., Nyman, A., Vielma, J., Passoth, V., Kiessling, A., & Lundh, T. (2020). Growth performance, nutrient digestibility and intestinal morphology of rainbow trout (Oncorhynchus mykiss) fed graded levels of the yeasts Saccharomyces cerevisiae and Wickerhamomyces anomalus. *Aquaculture Nutrition*, *26*(2), 275-286. https://doi.org/10.1111/anu.12988

Viswanath-Reddy, M., Bennett, S. N., & Howe, H. B., Jr. (1977). Characterization of glycerol nonutilizing and protoperithecial mutants of *Neurospora*. *Mol Gen Genet*, *153*(1), 29-38. https://doi.org/10.1007/BF01035993

Wang, J., Gao, Q., Zhang, H., & Bao, J. (2016). Inhibitor degradation and lipid accumulation potentials of oleaginous yeast *Trichosporon cutaneum* using lignocellulose feedstock. *Bioresource Technology*, *218*, 892-901. https://doi.org/10.1016/j.biortech.2016.06.130

Wang, Q.-M., Yurkov, A., Göker, M., Lumbsch, H., Leavitt, S., Groenewald, M., Theelen, B., Liu, X.-Z., Boekhout, T., & Bai, F.-Y. (2015). Phylogenetic classification of yeasts and related taxa within *Pucciniomycotina*. *Studies in mycology*, *81*(1), 149-189. https://doi.org/ https://doi.org/10.1016/j.simyco.2015.12.002

Wang, Y., Lin, X., Zhang, S., Sun, W., Ma, S., & Zhao, Z. K. (2016). Cloning and evaluation of different constitutive promoters in the oleaginous yeast

Rhodosporidium toruloides. *Yeast*, *33*(3), 99-106. https://doi.org/10.1002/yea.3145

Wang, Y., Zeng, F., Hon, C. C., Zhang, Y., & Leung, F. C. C. (2008). The mitochondrial genome of the Basidiomycete fungus *Pleurotus ostreatus* (oyster mushroom). *FEMS Microbiology Letters*, *280*(1), 34-41. https://doi.org/10.1111/j.1574-6968.2007.01048.x

Wang, Y., Zhang, S., Zhu, Z., Shen, H., Lin, X., Jin, X., Jiao, X., & Zhao, Z. K. (2018). Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast *Rhodosporidium toruloides*. *Biotechnology for Biofuels*, *11*(1), 148. https://doi.org/10.1186/s13068-018-1134-8

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), 737-738. https://doi.org/10.1038/171737a0

Wei, T., lt, sup, gt, lt, sup, gt, Yue, W., lt, sup, gt, lt, sup, gt, Jun, Z., lt, sup, gt, lt, . . . gt. (2019). Biosynthetic Pathway of Carotenoids in *Rhodotorula* and Strategies for Enhanced Their Production. *Journal of Microbiology and Biotechnology*, *29*(4), 507-517. https://doi.org/10.4014/jmb.1801.01022

Wendland, J., & Walther, A. (2014). Chromosome Number Reduction in *Eremothecium coryli* by Two Telomere-to-Telomere Fusions. *Genome Biology and Evolution*, *6*(5), 1186-1198. https://doi.org/10.1093/gbe/evu089

Wery, M., Gautier, C., Descrimes, M., Yoda, M., Vennin-Rendos, H., Migeot, V., Gautheret, D., Hermand, D., & Morillon, A. (2018). Native elongating transcript sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast. *RNA*, *24*(2), 196-208. https://doi.org/10.1261/rna.063446.117

Willenbücher, K., Wibberg, D., Huang, L., Conrady, M., Ramm, P., Gätcke, J., Busche, T., Brandt, C., Szewzyk, U., Schlüter, A., Barrero Canosa, J., & Maus, I. (2022). Phage Genome Diversity in a Biogas-Producing Microbiome Analyzed by Illumina and Nanopore GridION Sequencing. *Microorganisms*, *10*(2), 368. https://www.mdpi.com/2076-2607/10/2/368

Wolfe, K. H., & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, *387*(6634), 708-713. https://doi.org/10.1038/42711

Xu, S., Zhao, L., Xiao, S., & Gao, T. (2019). Whole genome resequencing data for three rockfish species of Sebastes. *Scientific Data*, *6*(1), 97. https://doi.org/10.1038/s41597-019-0100-z

Zhang, G.-C., Kong, I. I., Kim, H., Liu, J.-J., Cate, J. H. D., & Jin, Y.-S. (2014). Construction of a Quadruple Auxotrophic Mutant of an Industrial Polyploid Saccharomyces cerevisiae Strain by Using RNA-Guided Cas9 Nuclease. *Applied and Environmental Microbiology*, *80*(24), 7694-7701. https://doi.org/doi:10.1128/AEM.02310-14

Zhang, L., Chao, B., & Zhang, X. (2020). Modeling and optimization of microbial lipid fermentation from cellulosic ethanol wastewater by *Rhodotorula glutinis* based on the support vector machine. *Bioresource Technology*, *301*, 122781. https://doi.org/10.1016/j.biortech.2020.122781

Zhang, S., Ito, M., Skerker, J. M., Arkin, A. P., & Rao, C. V. (2016). Metabolic engineering of the oleaginous yeast *Rhodosporidium toruloides* IFO0880 for lipid overproduction during high-density fermentation. *Applied Microbiology and Biotechnology*, *100*(21), 9393-9405. https://doi.org/10.1007/s00253-016-7815-y

Zhang, S., Skerker, J. M., Rutter, C. D., Maurer, M. J., Arkin, A. P., & Rao, C. V. (2016). Engineering *Rhodosporidium toruloides* for increased lipid production. *Biotechnology and Bioengineering*, *113*(5), 1056-1066. https://doi.org/10.1002/bit.25864

Zhang, Z., Guo, J., Cai, X., Li, Y., Xi, X., Lin, R., Liang, J., Wang, X., & Wu, J. (2022). Improved Reference Genome Annotation of *Brassica rapa* by Pacific Biosciences RNA Sequencing [Original Research]. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.841618

Zhao, D., & Li, C. (2022). Multi-omics profiling reveals potential mechanisms of culture temperature modulating biosynthesis of carotenoids, lipids, and exopolysaccharides in oleaginous red yeast *Rhodotorula glutinis* ZHK. *LWT*, *171*, 114103. https://doi.org/10.1016/j.lwt.2022.114103

Zhao, X. Q., Aizawa, T., Schneider, J., Wang, C., Shen, R. F., & Sunairi, M. (2013). Complete mitochondrial genome of the aluminum-tolerant fungus *Rhodotorula taiwanensis* RS1 and comparative analysis of *Basidiomycota* mitochondrial genomes. *Microbiologyopen*, *2*(2), 308-317. https://doi.org/10.1002/mbo3.74

Zhao, Y., Song, B., Li, J., & Zhang, J. (2021). *Rhodotorula toruloides*: an ideal microbial cell factory to produce oleochemicals, carotenoids, and other products. *World Journal of Microbiology and Biotechnology*, *38*(1), 13. https://doi.org/10.1007/s11274-021-03201-4

Zhou, R., Zhu, Z., Zhang, S., & Zhao, Z. K. (2020). The complete mitochondrial genome of the lipid-producing yeast *Rhodotorula toruloides*. *FEMS Yeast Research*, *20*(6). https://doi.org/10.1093/femsyr/foaa048

Zhu, L., Wang, Z., Takala, J., Hiltunen, E., Qin, L., Xu, Z., Qin, X., & Yuan, Z. (2013). Scale-up potential of cultivating *Chlorella zofingiensis* in piggery wastewater for biodiesel production. *Bioresource Technology*, *137*, 318-325. https://doi.org/10.1016/j.biortech.2013.03.144

Zhu, L. Y., Zong, M. H., & Wu, H. (2008). Efficient lipid production with *Trichosporon fermentans* and its use for biodiesel preparation. *Bioresource Technology*, *99*(16), 7881-7885. https://doi.org/10.1016/j.biortech.2008.02.033

Zhu, Z., Zhang, S., Liu, H., Shen, H., Lin, X., Yang, F., Zhou, Y. J., Jin, G., Ye, M., Zou, H., & Zhao, Z. K. (2012). A multi-omic map of the lipid-producing

yeast *Rhodosporidium toruloides*. *Nature Communications*, *3*(1), 1112. https://doi.org/10.1038/ncomms2112

Zimmer, C., & Fabre, E. (2011). Principles of chromosomal organization: lessons from yeast. *Journal of Cell Biology*, *192*(5), 723-733. https://doi.org/10.1083/jcb.201010058

Zininga, J. T., Puri, A. K., Govender, A., Singh, S., & Permaul, K. (2019). Concomitant production of chitosan and lipids from a newly isolated *Mucor circinelloides* ZSKP for biodiesel production. *Bioresource Technology*, *272*, 545-551. https://doi.org/10.1016/j.biortech.2018.10.035

Popular science summary

Humans have used yeasts as an active "ingredient" in food production for centuries. A well-known example is the usage of *Saccharomyces cerevisiae* in bakeries and breweries. However, yeast lipid bioprocesses have also been studied as a microbial source of edible fats for more than a century ago and impelled from food shortages associated with the First and Second World Wars. *Rhodotorula* spp. are oleaginous yeasts because they can accumulate lipids in a proportion higher than 20% of their dry matter. Lipids and carotenoids from *Rhodotorula* spp. have potential biotechnological applications as food and feed additives. They could also be used as an alternative source of lipidic biomass for biodiesel production, with a reduced carbon footprint and food supply impairment compared to vegetable oils. *Rhodotorula* spp. have the advantage over other oleaginous microorganisms of having high lipid production rates, even when cultivated in residual products (e.g., wheat straw and woody residues from forestry). That makes the bioprocess more sustainable, aiding a circular economy concept, as illustrated in Figure 1.

Sequencing methods have generated an increasing availability of genomics and transcriptomics studies, revolutionizing the study of yeast biology and evolution. Baker's yeast was the first completely sequenced eukaryote. Genomics studies genes and genomes, which in Eukaryotes represent organized physical structures in which each DNA molecule is hierarchically folded into chromosome units. Genomics is a powerful tool for expanding the availability of genome sequences, studying evolution and providing a comprehensive framework for characterizing genome structure and organization. On the other hand, transcriptomics is the study of transcriptomes, which include all RNA molecules in a cell or organism. They provide insights into gene expression regulation in response to different

environments. It is worth mentioning that yeast divergence is quite broad. For example, closely related species from *Saccharomyces sensu stricto* have a comparable divergence to that among different orders of mammals. From *Rhodotorula* spp., *R. toruloides* gathers most of the genomics and transcriptomics studies, while little or none have been studied about other biotechnologically interesting species such as *R. babjevae*.

This thesis provided genome sequence and its annotation for different *Rhodotorula* strains. Ideally, a good genome assembly approach retrieves DNA sequences, each of them representing a complete chromosome. Gene annotation refers to the identification of regions that can be transcribed into RNA molecules. The comparative analysis of *Rhodotorula* genomes, assembled in a near-chromosome level resolution, gave some insights into their genome organization. One interesting finding was the presence of circular extrachromosomal DNA in *R. toruloides*. This discovery has potential applications for achieving higher efficiencies in strain genetic manipulation though further investigations are required. Another aspect of the work was evaluating the evolutive relationships between *Rhodotorula* strains using the recovered whole-genome DNA sequences.

The thesis investigated differential gene transcription of *R. toruloides* when hemicellulosic hydrolysate (HH) is present in crude glycerol (CG) media. CG is a major by-product of biodiesel production and a substrate in which *Rhodotorula* spp. can naturally grow. HH is a fraction of pretreated agriculture and forestry residues. Both HH and CG provide carbon sources for *Rhodotorula* spp. fermentation. When small amounts of HH were added to CG, *R. toruloides* glycerol utilization and lipid production rates increased. That is a desirable feature for enhancing the sustainability and efficiency of yeast lipid production. We suggested that the reason behind it is the activation of metabolic pathways that provide the cells with the necessary metabolic energy to adapt to the changing carbon source.

Overall, this thesis deepens our knowledge of the genomes, gene expression and physiology of lipid production of *Rhodotorula* yeasts. The results can be useful for selecting or even developing *Rhodotorula* strains with higher lipid productivity on low-cost residual substrates and ultimately establishing them as lipid cell factories.

Populärvetenskaplig sammanfattning

Människor har använt jäst som en aktiv "ingrediens" i livsmedelsproduktion under århundraden. Ett välkänt exempel är användningen av *Saccharomyces cerevisiae* i bagerier och bryggerier. Emellertid har lipidackumulerande jästarter också studerats som en mikrobiell källa till ätliga fetter för mer än ett sekel sedan och drivits av livsmedelsbrist i samband med första och andra världskrigen. *Rhodotorula* spp. är lipidackumulerande jästarter eftersom de kan ackumulera lipider till mer än 20 % av torrsubstansen. Lipider och karotenoider från *Rhodotorula*-arter har potentiella biotekniska tillämpningar som livsmedel och fodertillsatser. De skulle också kunna användas som en alternativ källa till biodieselproduktion, med ett minskat koldioxidavtryck och utan konkurrens med livsmedelsproduktion jämfört med vegetabiliska oljor. Arter inom släktet *Rhodotorula* har fördelen jämfört med andra lipidackumulerande mikroorganismer att de har höga lipidproduktionshastigheter, även när de odlas i restprodukter (t.ex. vetehalm och vedrester från skogsbruk). Det gör bioprocessen mer hållbar, vilket underlättar ett cirkulärt ekonomikoncept, som illustreras i Figur 1.

Sekvenseringsmetoder har genererat en ökande tillgänglighet av genom- och transkriptomstudier, vilket revolutionerar studier av jästbiologi och evolution. Bagerjäst var den första helt sekvenserade eukaryoten. Genomik studerar gener och genom, som i eukaryoter representerar organiserade fysiska strukturer där varje DNA-molekyl är hierarkiskt veckad till kromosomenheter. Genomik är ett kraftfullt verktyg för att utöka tillgängligheten av genomsekvenser, studera evolution och tillhandahålla ett heltäckande ramverk för att karakterisera genomets struktur och organisation. Å andra sidan finns transkriptomik, dvs studier av transkriptom, som inkluderar alla RNA-molekyler i en cell eller organism. Dessa ger insikter i regleringen av genuttryck som svar på olika miljöer. Det

är värt att nämna att skillnaden bland jästarter är ganska stor. Till exempel har närbesläktade arter från *Saccharomyces sensu stricto* en jämförbar divergens med den mellan olika däggdjursordningar. Inom släktet *Rhodotorula* är det främst arten *R. toruloides* som använts för genom- och transkriptomstudier, medan det finns få eller inga studier om andra bioteknologiskt intressanta arter som *R. babjevae*.

I denna avhandling gav sekvenserades och annoterades genom av olika *Rhodotorula*-stammar. DNA-sekvenserna ska helst representera en komplett kromosom. Genannoteringar är identifieringen av regioner som kan transkriberas till RNA-molekyler. När olika *Rhodotorula*-genom jämfördes, sammansatta i en upplösning på nästan kromosomnivå, gavs några insikter i deras genomorganisation. En intressant upptäckt var närvaron av cirkulär extrakromosomalt DNA i *R. toruloides*. Denna upptäckt har potentiella tillämpningar för att uppnå högre effektivitet vid genetisk manipulation av stammar även om ytterligare undersökningar krävs. En annan aspekt av arbetet var att undersöka de evolutionära relationerna mellan *Rhodotorula*-stammar.

Avhandlingen undersökte differentiell gentranskription av *R. toruloides* odlad i råglycerol (CG) blandad med en mindre mängd hemicellulosahydrolysat (HH). CG är en viktig biprodukt från biodieselproduktion och kan användas som substrat till *Rhodotorula* spp. HH är en del av förbehandlade jord- och skogsbruksrester. Både HH och CG tillhandahåller kolkällor för *Rhodotorula* spp.. När små mängder HH tillsattes till CG ökade *R. toruloides* glycerolutnyttjande och lipidproduktionshastighet. Det är en önskvärd egenskap för att förbättra hållbarheten och effektiviteten av jästlipidproduktion. Vi föreslog att orsaken bakom det är aktiveringen av metaboliska vägar som ger cellerna den nödvändiga metaboliska energin för att anpassa sig till den tillgängliga kolkällan.

Sammantaget fördjupar denna avhandling vår kunskap om genom, genuttryck och fysiologi för lipidproduktion hos jästarter inom släktet *Rhodotorula*. Resultaten kan vara användbara för att välja ut eller till och med utveckla *Rhodotorula*-stammar med högre lipidproduktivitet på billiga restsubstrat och slutligen etablera dem som lipidcellfabriker.

Acknowledgements

My PhD studies at SLU have been an amazing experience though a big challenge for me coming from Cuba. Besides academic education, I have gone through a long road of learning, such as language, social and cultural norms, heavy dependence on the internet, and dealing with Swedish winters. Though frustrating sometimes, it has been an amusing and rewarding process. I thus feel deeply blessed for this experience and for the support I got from all of you.

First of all, I would like to acknowledge my main supervisor Volkmar. Thank you for selecting me as a PhD student. You were very patient with the Cuban internet connection during my interview but active in solving the bureaucracies for my registration. Thank you for encouraging me to do a research internship and presenting my results at conferences abroad. They have been extremely valuable experiences for me. I am also thankful for your broad knowledge, guidance and approachability. Mats, I am also grateful for your share in my acceptance as a PhD student and all the struggles finding a flight back to Sweden when covid restrictions started. Your advice has always been very welcome. Bettina, it was great working with you. Your instructions have been priceless. Due to your encouragement and support, I have learnt more Bioinformatics, and it is something that I have gotten to enjoy very much. Sabine, I wish I had gone further with my project to learn more about yeast lipids and food chemistry. Yet, I got some good bits of advice about fats in healthy diets during lunches. Johanna, thank you for your kind help during my cloning attempts, for supervising me and commenting on my thesis, and for fun discussions during lunch.

I want to acknowledge the co-authors Christian, Martin and Adrian for their valuable contribution and help. Christian, I treasure your introduction

and guidance on Nanopore sequencing. I am also grateful for all the support and collaboration from Leila, He, and Anna Darlene for Nanopore sequencing and bioinformatics tips from Renaud.

A warm thanks to Jerry for your enthusiasm, pieces of advice, and for your patience when teaching me how to ice skate.

I also want to thank all my colleagues within the Lipodrive/Microfungi group for your warm welcome, shared discussions and help, including Yasha, Mikolaj, Mathilde, Nils, Jule, Jana, Alejandra and all the students and guests who spent some time with us. I want to thank especially Nils and Mikolaj for your kind support with computers and lab issues. Yasha, I really appreciate your positivity and goodwill. You were a top companion for travelling to conferences and workshops.

To my colleagues in Estonia, thank you for the collaboration during our joint project. Special thanks to Nemailla and Petri-Jaan for your supervision during my stay in Tartu.

I would like to give a special thanks to my office mates Laura and Topi. Your help to get around in the labs, with computer work, and even when preparing my thesis has been priceless. It has been great getting to know you. Though maybe I should apologize for distracting you too much. I would like to thank you for all the discussions, fun after-works and fellow PhD-student support, Sanjana, Florentine, Ani, Pernilla, Piera, Naike, Thomas, and Andreia. I also appreciated your kind help with confocal microscopy, Sanjana and Adrian. Thank you for your tips and discussions during fika and lunches, Igor, Shamik, Alyona, Hasi, Jonas, Gustav, Yong, and Simon. Also, thanks to Monika and other PhD students from the department for their help in PhD related matters and for organizing social activities together. I feel very grateful for sharing this PhD journey with all of you.

I am also thankful to Amanda, Jany and Arletys for all the shared moments and the little piece of home that I have found around you.

Quiero agradecer a toda mi familia. Los quiero un montón. Papi, gracias por insistirme en estudiar acá y por todos tus consejos, tanto académicos como sobre Suecia. Es realmente genial tenerte cerca. Mami, tío Jesús, nana y Vikita, no ha sido fácil estar lejos de ustedes, pero es un aliciente poder comunicarnos seguido y sentir su apoyo. Mami, también valoro mucho tu fuerza, tu nobleza y tus amorosos cuidados con abuela. Todos ustedes son mi modelo a seguir.

Axel, thank you for being so supportive and loving with me. You can really make me laugh even after the toughest days at work. I hope by now you have gotten to acknowledge yeasts as living organisms and not just "fermentation powder". Jag älskar dig, amor.

GENOMICS
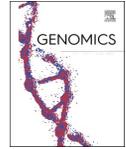
Short Communication

Chromosome-level genome assembly and transcriptome-based annotation of the oleaginous yeast *Rhodotorula toruloides* CBS 14

Giselle C. Martín-Hernández [a], [1], Bettina Müller [a], [1], Mikołaj Chmielarz [a], Christian Brandt [b], [c], Martin Hölzer [b], [d], Adrian Viehweger [b], [e], Volkmar Passoth [a], [*]

[a] *Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden*
[b] *nanozoo GmbH, Leipzig, Germany*
[c] *Institute for Infectious Diseases and Infection Control, Jena University Hospital, Jena, Germany*
[d] *RNA Bioinformatics and High-Throughput Analysis, Friedrich Schiller University Jena, Jena, Germany*
[e] *Department of Medical Microbiology, University Hospital Leipzig, Germany*

ARTICLE INFO

ABSTRACT

Rhodotorula toruloides is an oleaginous yeast with high biotechnological potential. In order to understand the molecular physiology of lipid synthesis in *R. toruloides* and to advance metabolic engineering, a high-resolution genome is required. We constructed a genome draft of *R. toruloides* CBS 14, using a hybrid assembly approach, consisting of short and long reads generated by Illumina and Nanopore sequencing, respectively. The genome draft consists of 23 contigs and 3 scaffolds, with a N50 length of 1,529,952 bp, thus largely representing chromosomal organization. The total size of the genome is 20,534,857 bp and the overall GC content is 61.83%. Transcriptomic data from different growth conditions was used to aid species-specific gene annotation. We annotated 9464 genes and identified 11,691 transcripts. Furthermore, we demonstrated the presence of a potential plasmid, an extrachromosomal circular structure of about 11 kb with a copy number about three times as high as the other chromosomes.

1. Introduction

The basidiomycete yeast *Rhodotorula toruloides* is an oleaginous microorganism with high biotechnological potential for lipid and carotenoid production. This yeast can naturally accumulate lipids up to 70% of dry cell weight, and a number of carotenoids [20,42]. *R. toruloides* can be cultivated to high cell densities on a wide range of substrates, including lignocellulose hydrolysate and other residual products [11,26,48,51,61]. This makes *R. toruloides* a promising host for the production of single-cell oil, as a sustainable and less controversial alternative to plant-derived oils for the production of biofuels, food and feed additives [42].

The molecular physiology behind lipid synthesis in *R. toruloides* has been relatively little explored, which hinders effective metabolic engineering for improved lipid production. Some draft genome sequences from *R. toruloides* strains have been determined using short-read sequencing technologies, however a complete picture of the *R. toruloides* genome on chromosomal level is still lacking [23,33,38,43,47,55,62,64]. The combination of short- and long-read sequencing strategies has been shown to improve the accuracy of genome sequences in yeast [41,54].

Thus, we combined Nanopore long-read sequencing and Illumina short-read sequencing to obtain a chromosome-level genome assembly. We further generated comprehensive transcriptomic data from different growth conditions to aid species-specific annotation. The results provide a valuable resource for pathway analysis and manipulation of *R. toruloides* and enable better understanding of genome biology and evolution of basidiomycetous yeasts.

* Corresponding author at: Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden.
E-mail address: Volkmar.Passoth@slu.se (V. Passoth).
[1] Giselle C. Martín-Hernández and Bettina Müller equally contributed to this paper and shall both be regarded as first authors.

2. Materials and methods

2.1. DNA extraction

R. toruloides CBS 14 (Westerdijk Fungal Biodiversity Institute, Utrecht, the Netherlands) was grown in 50 mL YPD medium [10]. Cells were harvested during exponential growth and the cell wall was digested according to Pi et al. [45]. DNA was extracted from the protoplasts using NucleoBond® CB 20 Kit (Macherey-Nagel, Germany). Concentration, purity and integrity was assessed with Qubit™ 4 Fluorometer, NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, USA) and agarose gel electrophoresis, respectively.

2.2. RNA extraction

R. toruloides was cultivated in 500 mL bioreactors (Multifors, Infors HT, Bottmingen, Switzerland) with either a mixture of lignocellulose hydrolysate and crude glycerol or crude glycerol only as carbon source [5,11]. Additionally, the media contained 0.75 g/L yeast extract (BactoTM Yeast Extract, BD, France), 1 g/L MgCl$_2$ 6xH$_2$O (Merck KGaA, Germany), 2 g/L (NH$_4$)$_2$HPO$_4$ ($\geq$98%, Sigma-Aldrich, USA), and 1.7 g/L YNB without amino acids and ammonium sulfate (DifcoTM, Becton Dickinson, France). Cultivations were performed in triplicates, at 25 °C, pH 6 and oxygen tension of 21%.

Total RNA was extracted in triplicates from 5 mL samples withdrawn after 10, 40, 72 and 96 h, respectively from bioreactors with mixed carbon source, and 10, 30 and 60 h, respectively from bioreactors with glycerol as sole carbon source, using the Quick-RNA™ Fecal/Soil Microbe MicroPrep Kit (Zymo Research, USA) according to the manufacturer's instructions with some modifications [37]. Briefly, the cells were harvested and resuspended in 1 mL Trizol and disrupted in a FastPrep −24 bead beater (M.P. Biomedicals, USA) at speed 6.0 m/s for 40 s. The homogenate was separated into layers by adding 0.2 mL chloroform and centrifugation. 400 μL of the upper layer containing the RNA was further processed as described in the manufacturer's manual. DNAse I treatment was performed as described in the manual applying 26 U/mL at 37° for 15 min. The technical replicates of the retrieved RNA were pooled prior rRNA depletion, which was performed using the human riboPOOL™ (siTOOLsBiotech, Germany) and Streptavidin-coated magnetic beads (Thermo Fisher Scientific, Norway) following the two-step protocol of the manufacturer. rRNA-depleted samples were purified by ethanol precipitation. RNAse inhibitor (1 U/μL, Thermo Fisher) was added before storage. Total RNA and rRNA-depleted samples were tested for integrity and quantity on RNA Nano Chips (Agilent 2100 Bioanalyzer System, Agilent Technologies, Germany).

2.3. Library preparation and sequencing

The extracted DNA was divided into two samples and sequenced using either MinION (Oxford Nanopore Technologies) or Illumina sequencing platform. Before Nanopore library preparation, 5 μg of the retrieved DNA were "pre-cleaned" using 31.5 μL of AMPure magnetic beads, for removing short DNA fragments [6]. The DNA library was prepared using the Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies, Oxford, UK) and a modified protocol [6]. The library was loaded onto a FLO-MIN106 flow cell attached to the MinION device (Oxford Nanopore Technologies). Sequencing was performed using the MinKNOW software (Oxford Nanopore Technologies) according to the protocol from Brandt et al. [6]. The retrieved long reads had a mean length of 7889 bp and a read length N50 of 11,739 bp. The total number of reads was 124,417 and the total number of base pairs sequenced was 981.6 Mb.

Short-read paired-end sequencing of DNA and rRNA-depleted samples was performed on the Illumina Novaseq platform (S prime, 2 × 150 bp) using TruSeq PCR free DNA library preparation kit (Illumina Inc.).

In order to verify the circular structure of contig_63, Sanger sequencing (Macrogen Europe B·V, the Netherlands) was performed from PCR amplicons using the primers shown in supplementary table s1. The PCR mixture consisted of 0.3 μL DNA, 1.25 μL primer, 12.5 μL Dream Taq Green PCR Mix (Thermo Scientific, Lithuania) and 0.8 μL DMSO in a total volume of 25 μL. Amplification was conducted as follows: initial denaturation at 95 °C for 5 min, 35 cycles of denaturation (95 °C for 30 s), annealing (1 min) and elongation (72 °C) followed by a final extension at 72 °C for 10 min. Annealing temperatures and elongation times were adapted to the respective primer combination (table S2). Amplification was assessed on agarose gel electrophoresis at 8.2 V/cm electric field strength. The gel was prepared using GelRed® Nucleic Acid Gel Stain (Biotium), 1.0% agarose and TAE 1× buffer (VWR Life Science). The PCR products from the corresponding sizes were purified using GeneJet Gel Extraction kit (Thermo Scientific, Lithuania).

Geneious prime version 2021.0.1 (Biomatters Ltd.) was used for assembly of the Sanger sequences obtained and to align *FAS2* and *FAS21* from *R. toruloides* CBS 14 and *FAS2* from *R. toruloides* NP11. Whole genome alignment was performed using the LASTZ sequence alignment tool (version 7.0.2) implemented in Geneious [22].

2.4. Quality control and genome assembly

Short and long reads generated by Illumina and Nanopore sequencing, respectively, were combined for hybrid *de-novo* assembly. Before, the quality of the long reads was ensured using NanoPlot v1.25.0 [15] and all short reads were removed until reaching a target base coverage with Filtlong v0.2.0 (https://github.com/rrwick/Filtlong) [59] applying the parameters --target_bases 5,000,000,000 and --length_weight 8. Short reads were quality-trimmed and adapter-clipped with fastp v0.20.0 [9] using parameters -5 -3 -W 4 -M 20 -l 25 -x -n 5 -z 6. To achieve high contiguity in the initial assembly step, a draft assembly was produced from the preprocessed long reads using Flye v2.8 [30], setting the suggested genome size to 20 Mbp and keeping the plasmid and meta options activated. The Flye draft assembly was further subjected to long-read-based polishing using two rounds of Racon v1.4.7 [56] followed by one round of Medaka v0.10.0 (https://github.com/nanoporetech/medaka) [53] using the model r941_min_high. We used minimap2 (−x map-ont) v2.17 [34] and samtools v.1.10 [36] to prepare alignment files for the Racon polishing. The assembly graph was visualized and investigated via Bandage v0.8.1 [60]. The long-read assembly was finally subjected to two Pilon v1.23 [57] rounds, in which the quality-filtered short reads were used for final polishing. Alignment files for Pilon were prepared using BWA v0.7.17 [35] and samtools. We mapped the Illumina and Nanopore reads with BWA and minimap2, respectively, to the final assembly and used the pileup.sh script from BBMap v38.86 [8] to calculate coverage histograms for each contig with a bin size of 1000 nt and plotted them with ggplot2. Besides the visual inspection of coverage patterns, we used the samtools coverage function and the faidx command on the long-read-mapped data to filter out contigs violating at least one of the following cutoffs: minimum read number = 100, minimum bases covered by reads = 5000, minimum read coverage = 15, and minimum read coverage depth (amount of bases covered by reads in comparison to contig length) = 10. The resulting quality-controlled and cleaned assembly file was used for annotation and chromosome analyses.

2.5. Annotation

The final assembly was first screened for repetitive regions using RepeatMasker v4.0.9 [40], checked for completeness using BUSCO v3.0.2 [50,58] with the fungi_odb9 database as reference and visualized using chromoMap v0.2 [2]. Contigs, larger than 10 kb but smaller than 250 kb were checked for circularization using a python script (https://github.com/Kzra/Simple-Circularise, v1.0) [29].

Taxonomic classification was performed using sourmash v2.0.1 [46] and its "LCA" method. Both indexing the tree and querying genomes

used a k-mer size of 31 and a sampling fraction of 10^{-4}. The LCA index was derived from publicly available genomes (GenBank, https://osf.io/4f8n3/).

The repeat-masked assembly was further annotated using a combination of homology-based gene comparison and RNA-Seq-derived transcript information. First, we used MetaEuk v1.ea903e5 [27] with the "easy-predict" subcommand for draft gene annotation providing all proteins (filtered models; best) obtained from the JGI fungal genome portal MycoCosm (http://jgi.doe.gov/fungi; downloaded April 2020) as a database. Next, we combined and improved the MetaEuk annotation with the obtained RNA-Seq data. Prior to this, RNA-Seq datasets were quality-checked using FastQC v0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [3] and trimmed and adapter-clipped using fastp with the aforementioned parameters. Potential residual rRNA was removed by SortMeRNA v2.1b [31]. The processed RNA-Seq reads were per-sample-mapped to the assembly using HISAT v2.2.0 [28] and subsequently provided to StringTie v2.1.1 [32,44] for transcript annotation, guided by the initial MetaEuk gene annotation. Finally, the MetaEuk-guided StringTie information from each individual RNA-Seq sample were merged into a final single annotation using the StringTie merge functionality. In this step, we again provided the full MycoCosm-based MetaEuk annotation next to the individual RNA-Seq annotations to also include genes that were not recovered in the RNA-Seq data. Lastly, we extracted all annotated gene sequences and used the dammit pipeline v1.2 (http://www.camillescott.org/dammit) [49] for functional annotation. Dammit unites different databases for annotation: Pfam-A, Rfam, OrthoDB, and again BUSCO (fungi_odb9) that we have decorated with additional information derived from UniProt/Swiss-Prot. We extracted the GO terms annotated by dammit and counted their appearance to apply a weight per GO term. GO terms that occured at least 10 times were subjected to REVIGO (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021800) to summarize them by removing redundancy and to visualize the results in semantic similarity-based scatterplots and interactive graphs.

We used a custom python script to identify potential telomeres by using known motifs at contig ends. This information was provided along with the Nanopore reads to Tapestry v1.0.0 [14].

3. Results and discussion

The hybrid *de-novo* assembly for *R. toruloides* CBS 14 (Fig. 1) resulted in a 20,534,857 bp genome, which is in line with the 20.4 Mb median reported for this and other *R. toruloides* strains (Table 1). The overall GC content of the obtained genome is 61.83%, which corresponds to previously reported values (61.9% on average). The identified repetitive sequences represent 2.01% of the total genome length, of which 1.56% are single repeats and 0.45% low complexity regions. The assembled genome resulted in 23 contigs and 3 scaffolds ranging from 5778 to 1,965,970 bp, and an N50 length of 1,529,952 bp (Table 2). This number of contigs and scaffolds was significantly lower than that obtained using short-read sequencing only: the lowest number achieved in previous studies is 186 contigs (Table 1). This clearly shows the higher accuracy, contiguity and completeness of the genome presented here, achieved through the improved coverage using a hybrid assembly approach. 56.37% of the genome is represented by 4 contigs and 3 scaffolds, each larger than 1 Mb. The rest is allocated in 10 medium-size contigs (0.5–1 Mb) and in 9 small contigs (<0.5 Mb), four of which are circularized. Telomere sequences could be detected at both termini of two contigs and at one terminus each of 15 contigs (Table 2). De Jonge et al. [16] identified at least 11 chromosomes in *R. toruloides* CBS 14 by pulsed field gel electrophoresis (PFGE). Zhu et al. [64] predicted 16 chromosomes in *R. toruloides* NP11 applying Illumina sequencing and gap closing using genome walking. When comparing the size of the obtained contigs and scaffolds with the corresponding bands found by PFGE, we identified at least 18 chromosomes in *R. toruloides* CBS 14.

The mitochondrial genome was recovered in one contig with a length of 157 kb. However, the final step of circularizing sequences in an assembly remains challenging and thus can result in shorter or even longer contigs, since assembly programs cannot clearly define where a circular DNA ends, even despite longer reads [24]. Thus, tandem repeats at the end of a contig can increase the length of a linear representation of a circular sequence. Such repeats can be identified by overlapping sequences at the ends of a contig. To account for this, in an additional step, we aligned the contig against itself to identify and remove circular repeats, ultimately resulting in an actual mitochondrial genome size of 69 kb. The mitochondrial genome has a GC content of 40.9%, which agrees with the GC values previously reported for *R. toruloides* strains NP11 and NBRC 0880 [13,63].

In addition to the mitochondrial genome, three further circular contigs (contig_64, contig_49, contig_63) were predicted. They have a comparable GC content of about 62% (Table 2). Among these, contig 63 was particularly noticeable because it showed a read depth approximately three times higher than the other chromosomes, which may indicate relaxed replication regulation (Table 2). We confirmed the circular structure of contig 63 by overlapping amplification using sequence-specific primers and subsequent Sanger sequencing of the amplicons obtained (Supplementary Fig. S1, supplementary file 2). This showed that the circular structure is about 1143 bp larger than the one predicted (Supplementary Fig. S1). As explained above, circular sequence structures possess a certain challenge for assembly tools and
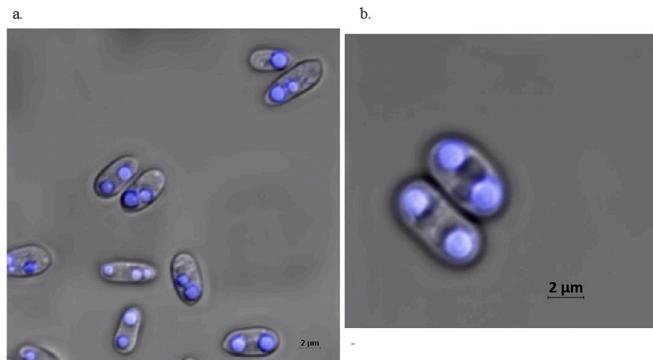

Fig. 1. Confocal microscope picture of Nile Red-stained lipid droplets of *Rhodotorula toruloides* CBS 14. (a) Group of cells at stationary phase (b) Magnified view of cells.

Table 1

Genome assembly statistics of *Rhodotorula toruloides* strains and sequencing platforms used.

Reference	Strain	Sequencing platform	Coverage	Genome size (Mb)	GC content (%)	Contigs	Scaffolds	coding sequences
[33]	MTCC 457	Illumina	121 X	20.09	62	689	644	NA
[64]	NP11	Illumina	96 X	20.2	NA	17,814	34	8171 a
[38]	CECT1137	Illumina, Roche 454 FLX	NA	20.45	61.9	NA	62	8206 a
[43]	ATCC 204091	Illumina, Roche 454 FLX, Sanger	NA	20.48	61.9	186	29	3359 b
[62]	IFO0559	Illumina	170 X	20.28	NA	NA	246	8100 c
[62]	IFO0880	Illumina	139 X	20.36	NA	NA	219	7920 c
[23]	ATCC 10788	Illumina	241 X	20.75	62.01	NA	61	7730 a
[23]	ATCC 10657	Illumina	233 X	21.49	61.81	NA	137	7800 a
[47]	CGMCC 2.1609	Illumina, Roche 454 FLX	13 X	33.4	61.9	868	365	9820
[55]	VN1	Illumina	125 X	20.02	61.8	424	NA	8021 a
This study	CBS 14	Illumina, Oxford Nanopore	1000 X	20.53	61.83	23	3	9464 a

NA, not available; a, b and c refer to protein-coding sequences, putative genes and predicted proteins, respectively.

Table 2

Characteristics of the assembled contigs and scaffolds.

Contig/ scaffold	Size (bp)	GC content (%)	Read depth	No. telomeric regions
contig_1	666,847	62.12	21	1
contig_19	763,076	61.61	20	0
contig_20	906,560	61.88	21	2
contig_28	659,395	61.82	21	1
contig_29	785,970	61.85	19	0
contig_30	1,712,111	62.15	21	1
contig_31	5778	39.10	0	1
contig_33	1,538,587	62.03	25	1
contig_34	1,529,952	62.08	21	1
contig_35	930,820	61.92	22	1
contig_36	1,965,970	62.07	21	0
contig_40	309,449	61.55	17	1
contig_43	618,889	62.24	20	0
contig_49	63,359	62.14	14	1
contig_5	577,483	61.83	20	1
contig_53	670,355	61.86	20	2
contig_58	477,283	62.04	22	0
contig_60	668,592	62.28	22	1
contig_63	11,496	61.73	63	1
contig_64	152,705	62.29	20	1
contig_65	265,355	60.97	20	0
contig_67	268,355	62.24	20	0
contig_8	157,597	40.90	438	0
scaffold_27	1,925,638	61.87	22	1
scaffold_3	1,219,652	62.22	21	0
scaffold_32	1,683,583	62.10	22	1

can particularly result in longer but also shorter contigs that miss parts of the actual sequence at the end of a contig [24]. The genes annotated in contig 63 are *UTP22*, *H2A* and *H2B* which are encoding for RNA-associated protein 22, Histone H2A and Histone H2B, respectively (Supplementary Table S3). A truncated copy of *UTP22* from *R. toruloides* NP11 was also annotated in contig 63 and identified through Blast search (Supplementary Table S3). The CDS of *UTP22*, *H2A* and *H2B* are also found in scaffold 32 sharing a pairwise identity of 99.8%, 96.7%, and 98.5%, respectively (Supplementary Figs. S2, S3). *H2B* and *UTP22* are localized next to each other, *H2A* is localized elsewhere on scaffold 32 (Supplementary Fig. S3). Both of the *UTP22* genes encoded in contig 63 are shortened compared to the *UTP22* gene that we found encoded on scaffold 32. The *UTP22* gene on scaffold 32 shares 100% sequence identity with the *UTP22* gene identified in strain NP11. In addition to scaffold 32, the CDS of *H2B* aligned partially to contig_29 and to scaffold 30 sharing a pairwise identity of 88% (Supplementary Fig. S2). Five transcripts were annotated within the circular contig 63 (Supplementary Table S3). All of them were expressed under the experimental conditions. Extrachromosomal endogenous DNA have been previously found in *Saccharomyces cerevisiae* [39,52]. The presence of DNA mitochondrial plasmids in filamentous fungi, including some Basidiomycota species, have also been widely acknowledged before ([21]; [65]; [12]). Within them, mitochondrial circular DNA plasmids are less frequently found

than linear. They are typically within a length range of 2.5–5 kb, and encoding enzymes involved in their replication such as a DNA polymerase or a reverse transcriptase ([12]; [65]). In *S. cerevisiae* DNA replication initiate at certain autonomously replicating sequences (ARS) [17]. An ARS consensus sequence (ACS), or ACS like pattern, was not found in contig 63. AT-rich ACS elements of ARS have been found to be degenerate and in *Schizosaccharymces pombe*, which is closer related to *S. cerevisiae* than *R. toruloides*, the absence of such consensus sequence has been reported [17,25]. This is the first time that such extrachromosomal circular DNA has been detected in Basidiomycetes. The replication origin of such structures may be useful for developing efficient episomal vectors for the manipulation of *Rhodotorula* yeasts.

Analysis of the RNA-Seq data of *R. toruloides* CBS 14 using StringTie software and guided by the MetaEuk annotation (based on all protein sequences from MycoCosm) resulted in the annotation of 9464 genes with 11,691 transcripts. This number is significantly higher than previously reported for *R. toruloides* genomes (Table 1). The higher number of transcripts than protein-coding genes can be explained by alternative splicing and non-coding RNAs. Zhu et al. [64] found 1371 genes encoding two or more transcript isoforms in *R. toruloides* NP11. The average number of exons per gene is 5.95 in *R. toruloides* CBS 14 (supplementary table 4). Confirming the findings of Zhu et al. [64], we observed a predominance of split genes in the genome, with a total of 8550 for *R. toruloides* CBS 14. After functional annotation of coding sequences through Gene Ontology (GO) terms we clustered annotations into the categories: biological processes, cellular components and molecular functions (Supplementary Figs. S4–S6). Genes encoding enzymes crucial for lipid and carotenoid metabolisms such as *CDC19*, *MAE1*, *MAE2*, *ACL1*, *ACC1*, *FAS1*, *FAS2*, *OLE1*, *ACAD10*, *ACAD11*, *crtYB*, *crtI* and *BTS1* are present in the genome (supplementary table s5). Similar to the situation in *S. cerevisiae* [7], the genes encoding the α- and β-subunits of FAS are located in different chromosomes (Supplementary Table S5).

The MetaEuk annotation identified a gene (*FAS21*) on the opposite strand of the *FAS2*- gene (Supplementary Fig. S7), which encodes the α-subunit of the fatty acid synthase complex (FAS) in *R. toruloides* NP11 [64]. *FAS21* contains 2 exons and its mRNA would be complementary to parts of the *FAS2*-sequence. There is a growing number of natural antisense transcripts identified in fungal transcriptome analyses [18,19]. *FAS21* transcript might be involved in controlling *FAS2* expression. However, we did not identify a *FAS21* transcript under the experimental conditions carried out, thus the involvement of *FAS21* in the regulation of *FAS2*- expression and thus fatty acid synthesis still needs to be verified. Three different transcripts of *FAS2* were identified containing 3, 12 and 16 exons, respectively (Supplementary Table S6).

BUSCO orthologs analysis showed that 96.9% of the assessed genes were identified and complete (96.6% single-copy and 0.3% duplicated), 1.7% fragmented and only 1.4% were lacking, demonstrating the high quality of the hybrid genome assembly (Supplementary Fig. S8).

In the case of the transcriptomic data, 84.1% of the genes were indicated as complete (69.3% single copy,14.8% duplicated), 14.1% as

fragmented and 1.8% genes as missing.

The genome annotation and assembly correspond to the following taxonomic classification: Eukaryota superkingdom, Basidiomycota phylum, Microbotryomycetes class, Sporidiobolales order, Sporidiobolaceae family, *Rhodotorula* genus, *Rhodotorula toruloides* species, CBS 14 strain.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.10.006.

References

[2] L. Anand, ChromoMap: an R package for interactive visualization and annotation of chromosomes, BioRxiv 605600 (2019), https://doi.org/10.1101/605600.

[3] S. Andrews, et al., FastQC: a quality control tool for high throughput sequence data, in: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

[5] J. Blomqvist, et al., Oleaginous yeast as a component in fish feed, Sci. Rep. 8 (1) (2018) 1–8, https://doi.org/10.1038/s41598-018-34232-x.

[6] C. Brandt, E. Bongcam-Rudloff, B. Müller, Abundance tracking by long-read nanopore sequencing of complex microbial communities in samples from 20 different biogas/wastewater plants, Appl. Sci. 10 (21) (2020) 1–14, https://doi.org/10.3390/app10217518.

[7] G. Burkl, H. Castorph, E. Schweizer, Mapping of a complex gene locus coding for part of the *Saccharomyces cerevisiae* fatty acid synthetase multienzyme complex, MGG Mol. Gen. Genet. 119 (4) (1972) 315–322, https://doi.org/10.1007/BF00272089.

[8] B. Bushnell, BBMap: A Fast, Accurate, Splice-Aware Aligner. United States. https://www.osti.gov/servlets/purl/1241166, 2014.

[9] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890, https://doi.org/10.1093/bioinformatics/bty560.

[10] M. Chmielarz, et al., FT-NIR: a tool for rapid intracellular lipid quantification in oleaginous yeasts, Biotechnol. Biofuels 12 (1) (2019) 1–9, https://doi.org/10.1186/s13068-019-1513-9.

[11] M. Chmielarz, J. Blomqvist, S. Sampels, M. Sandgren, V. Passoth, Microbial lipid production from crude glycerol and hemicellulosic hydrolysate with oleaginous yeasts, Biotechnol. Biofuels 14 (1) (2021) 1–12, https://doi.org/10.1186/s13068-021-01916-y.

[12] P. Cahan, J.C. Kennell, Identification and distribution of sequences having similarity to mitochondrial plasmids in mitochondrial genomes of filamentous fungi, Mol. Gen. Genomics. 273 (6) (2005) 462–473, https://doi.org/10.1007/s00438-005-1133-x.

[13] S.T. Coradetti, et al., Functional genomics of lipid metabolism in the oleaginous yeast *Rhodosporidium toruloides*, ELife 7 (2018) 1–55, https://doi.org/10.7554/eLife.32110.

[14] J.W. Davey, S.J. Davis, J.C. Mottram, P.D. Ashton, Tapestry: validate and edit small eukaryotic genome assemblies with long reads, BioRxiv (2020), https://doi.org/10.1101/2020.04.24.059402.

[15] W. De Coster, S. D'Hert, D.T. Schultz, M. Cruts, C. Van Broeckhoven, NanoPack: visualizing and processing long-read sequencing data, Bioinformatics 34 (15) (2018) 2666–2669, https://doi.org/10.1093/bioinformatics/bty149.

[16] P. De Jonge, F.C.M. De Jongh, R. Meijers, H.Y. Steensma, W.A. Scheffers, Orthogonal-field-alternation gel electrophoresis banding patterns of DNA from yeasts, Yeast 2 (3) (1986) 193–204, https://doi.org/10.1002/yea.320020307.

[17] M.K. Dhar, S. Sehgal, S. Kaul, Structure, replication efficiency and fragility of yeast *ARS* elements, Res. Microbiol. 163 (4) (2012) 243–253, https://doi.org/10.1016/j.resmic.2012.03.003.

[18] M.E. Donaldson, B.J. Saville, Natural antisense transcripts in fungi, Mol. Microbiol. 85 (3) (2012) 405–417, https://doi.org/10.1111/j.1365-2958.2012.08125.x.

[19] N. Chacko, X. Lin, Non-coding RNAs in the development and pathogenesis of eukaryotic microbes, Appl. Microbiol. 97 (18) (2013) 7989–7997, https://doi.org/10.1007/s00253-013-5160-y.

[20] Y. González-García, et al., High lipids accumulation in *Rhodosporidium toruloides* by applying single and multiple nutrients limitation in a simple chemically defined

medium, Ann. Microbiol. 67 (8) (2017) 519–527, https://doi.org/10.1007/s13213-017-1282-2.

[21] A.J.F. Griffiths, Natural plasmids of filamentous fungi, Microbiol. Rev. 59 (4) (1995) 673–685, https://doi.org/10.1128/mmbr.59.4.673-685.1995.

[22] R.S. Harris, Improved Pairwise Alignment of Genomic DNA, Ph.D. Thesis, The Pennsylvania State University, 2007.

[23] J. Hu, L. Ji, Draft genome sequences of *Rhodosporidium toruloides* strains ATCC 10788 and ATCC 10657 with compatible mating types, Genome Announc. 4 (2) (2016) 1083–1084, https://doi.org/10.1128/genomeA.00098-16.

[24] M. Hunt, et al., Circlator: automated circularization of genome assemblies using long sequencing reads, Genome Biol. 16 (1) (2015) 1–10, https://doi.org/10.1186/s13059-015-0849-0.

[25] R. Iwakiri, S. Eguchi, Y. Noda, H. Adachi, K. Yoda, Isolation and structural analysis of efficient autonomously replicating sequences (ARSs) of the yeast *Candida utilis*, Yeast 22 (13) (2005) 1049–1060, https://doi.org/10.1002/yea.1296.

[26] H.S.J. Lopes, N. Bonturi, E.J. Kerkhoven, E.A. Miranda, P. Lahtvee, C/N ratio and carbon source-dependent lipid production profiling in *Rhodotorula toruloides*, Appl. Microbiol. Biotechnol. 104 (6) (2020) 2639–2649, https://doi.org/10.1007/s00253-020-10386-5.

[27] E.L. Karin, M. Mirdita, J. Söding, MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics, Microbiome 8 (48) (2020), https://doi.org/10.1186/s40168-020-00808-x.

[28] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (4) (2015) 357–360, https://doi.org/10.1038/nmeth.3317.

[29] Kitson, E. (n.d.). SimpleCircularise. https://github.com/Kzra/Simple-Circularise.

[30] M. Kolmogorov, J. Yuan, Y. Lin, P.A. Pevzner, Assembly of long, error-prone reads using repeat graphs, Nat. Biotechnol. 37 (5) (2019) 540–546, https://doi.org/10.1038/s41587-019-0072-8.

[31] E. Kopylova, L. Noé, H. Touzet, SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data, Bioinformatics 28 (24) (2012) 3211–3217, https://doi.org/10.1093/bioinformatics/bts611.

[32] S. Kovaka, et al., Transcriptome assembly from long-read RNA-seq alignments with StringTie2, Genome Biol. 20 (1) (2019) 1–13, https://doi.org/10.1186/s13059-019-1910-1.

[33] S. Kumar, H. Kushwaha, A.K. Bachhawat, G.P.S. Raghava, K. Ganesan, Genome sequence of the oleaginous red yeast *Rhodosporidium toruloides* MTCC 457, Eukaryot. Cell 11 (8) (2012) 1083–1084, https://doi.org/10.1128/EC.00156-12.

[34] H. Li, Minimap2: pairwise alignment for nucleotide sequences, Bioinformatics 34 (18) (2018) 3094–3100, https://doi.org/10.1093/bioinformatics/bty191.

[35] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows—Wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.

[36] H. Li, et al., The sequence alignment/map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.

[37] S. Manzoor, E. Bongcam-Rudloff, A. Schnürer, B. Müller, Genome-guided analysis and whole transcriptome profiling of the mesophilic syntrophic acetate oxidising bacterium *Syntrophaceticus schinkii*, PLoS One 11 (11) (2016) 1–24, https://doi.org/10.1371/journal.pone.0166520.

[38] N. Morin, et al., Draft genome sequence of *Rhodosporidium toruloides* CECT1137, an oleaginous yeast of biotechnological interest, Genome Announc. 2 (4) (2014) 578–579, https://doi.org/10.1128/genomeA.00641-14.

[39] H.D. Møller, L. Parsons, T.S. Jørgensen, D. Botstein, B. Regenberg, Extrachromosomal circular DNA is common in yeast, Proc. Natl. Acad. Sci. U. S. A. 112 (24) (2015) E3114–E3122, https://doi.org/10.1073/pnas.1508825112.

[40] D. Nishimura, RepeatMasker. Biotech Software & Internet Report 1(1–2), 2000, pp. 36–39, https://doi.org/10.1089/152791600319259.

[41] R.A. Olsen, et al., De novo assembly of *Dekkera bruxellensis*: a multi technology approach using short and long-read sequencing and optical mapping, GigaScience 4 (1) (2015), https://doi.org/10.1186/s13742-015-0094-1.

[42] Y.K. Park, J.M. Nicaud, R. Ledesma-Amaro, The engineering potential of *Rhodosporidium toruloides* as a workhorse for biotechnological applications, Trends Biotechnol. 36 (3) (2018) 304–317, https://doi.org/10.1016/j.tibtech.2017.10.013.

[43] D. Paul, et al., Genome sequence of the oleaginous yeast *Rhodotorula glutinis* ATCC, Genome Announc. 2 (1) (2014) 2009–2010, https://doi.org/10.1128/genomeA.00046-14.

[44] M. Pertea, et al., StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. 33 (3) (2015) 290, https://doi.org/10.1038/nbt.3122.

[45] H.W. Pi, et al., Engineering the oleaginous red yeast *Rhodotorula glutinis* for simultaneous β-carotene and cellulase production, Sci. Rep. 8 (1) (2018) 2–11, https://doi.org/10.1038/s41598-018-29194-z.

[46] N.T. Pierce, L. Irber, T. Reiter, P. Brooks, C.T. Brown, Large-scale sequence comparisons with sourmash, F1000 Res. 8 (2019) 1006, https://doi.org/10.12688/f1000research.19675.1.

[47] C. Sambles, et al., Genome sequence of the oleaginous yeast *Rhodotorula toruloides* strain CGMCC 2.1609, Genomics Data 13 (2017) 1–2, https://doi.org/10.1016/j.gdata.2017.05.009.

[48] V. Sànchez Nogué, et al., Integrated diesel production from lignocellulosic sugars: via oleaginous yeast, Green Chem. 20 (18) (2018) 4349–4365, https://doi.org/10.1039/c8gc01905c.

[49] Scott, C. (n.d.). dammit: An Open and Accessible De Novo Transcriptome Annotator. http://www.camillescott.org/dammit.

[50] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy

orthologs, Bioinformatics 31 (19) (2015) 3210–3212, https://doi.org/10.1093/bioinformatics/btv351.

[51] G. Singh, S. Sinha, K.K. Bandyopadhyay, M. Lawrence, D. Paul, Triauxic growth of an oleaginous red yeast *Rhodosporidium toruloides* on waste "extract" for enhanced and concomitant lipid and β-carotene production, Microb. Cell Factories 17 (1) (2018) 1–10, https://doi.org/10.1186/s12934-018-1026-4.

[52] P.K. Strope, et al., 2μ plasmid in *Saccharomyces* species and in *Saccharomyces cerevisiae*, FEMS Yeast Res. 15 (8) (2015) 1–8, https://doi.org/10.1093/femsyr/fov090.

[53] Technologies, O. N. (n.d.). Medaka: Long-read Polishing Pipeline. https://github.com/nanoporetech/medaka.

[54] I.A. Tiukova, et al., Chromosomal genome assembly of the ethanol production strain CBS 11270 indicates a highly dynamic genome structure in the yeast species *Brettanomyces bruxellensis*, PLoS One 14 (5) (2019) 1–20, https://doi.org/10.1371/journal.pone.0215077.

[55] T.N. Tran, D.H. Ngo, N.T. Nguyen, D.N. Ngo, Draft genome sequence data of *Rhodosporidium toruloides* VN1, a strain capable of producing natural astaxanthin, Data Brief 26 (2019) 104443, https://doi.org/10.1016/j.dib.2019.104443.

[56] R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate *de novo* genome assembly from long uncorrected reads, Genome Res. 27 (5) (2017) 737–746, https://doi.org/10.1101/gr.214270.116.

[57] B.J. Walker, et al., Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS One 9 (11) (2014), https://doi.org/10.1371/journal.pone.0112963.

[58] R.M. Waterhouse, et al., BUSCO applications from quality assessments to gene prediction and phylogenomics, Mol. Biol. Evol. 35 (3) (2018) 543–548, https://doi.org/10.1093/molbev/msx319.

[59] Wick, R. (n.d.). Filtlong: Quality Filtering Tool for Long Reads. https://github.com/rrwick/Filtlong.

[60] R.R. Wick, M.B. Schultz, J. Zobel, K.E. Holt, Bandage: interactive visualization of de novo genome assemblies, Bioinformatics 31 (20) (2015) 3350–3352, https://doi.org/10.1093/bioinformatics/btv383.

[61] M.G. Wiebe, K. Koivuranta, M. Penttilä, L. Ruohonen, Lipid production in batch and fed-batch cultures of *Rhodosporidium toruloides* from 5 and 6 carbon carbohydrates, BMC Biotechnol. 12 (2012), https://doi.org/10.1186/1472-6750-12-26.

[62] S. Zhang, et al., Engineering *Rhodosporidium toruloides* for increased lipid production, Biotechnol. Bioeng. 113 (5) (2016) 1056–1066, https://doi.org/10.1002/bit.25864.

[63] R. Zhou, Z. Zhu, S. Zhang, Z.K. Zhao, The complete mitochondrial genome of the lipid-producing yeast *Rhodotorula toruloides*, FEMS Yeast Res. 20 (6) (2020), https://doi.org/10.1093/femsyr/foaa048.

[64] Z. Zhu, et al., A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*, Nat. Commun. 3 (2012) 1111–1112, https://doi.org/10.1038/ncomms2112.

[65] Y. Wang, F. Zeng, C.C. Hon, Y. Zhang, F.C.C. Leung, The mitochondrial genome of the Basidiomycete fungus Pleurotus ostreatus (oyster mushroom), FEMS Microbiol. Lett. 280 (1) (2008) 34–41, https://doi.org/10.1111/j.1574-6968.2007.01048.x.

II

Article

Near Chromosome-Level Genome Assembly and Annotation of *Rhodotorula babjevae* Strains Reveals High Intraspecific Divergence

Giselle C. Martín-Hernández [1,†], Bettina Müller [1,†], Christian Brandt [2], Martin Hölzer [3], Adrian Viehweger [4] and Volkmar Passoth [1,*]

1 Department of Molecular Sciences, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden; giselle.martin@slu.se (G.C.M.-H.); bettina.muller@slu.se (B.M.)
2 Institute for Infectious Diseases and Infection Control, Jena University Hospital, 07743 Jena, Germany; christian.brandt@med.uni-jena.de
3 Method Development and Research Infrastructure, MF1 Bioinformatics, Robert Koch Institute, 13353 Berlin, Germany; hoelzerm@rki.de
4 Institute of Medical Microbiology and Virology, University Hospital Leipzig, 04103 Leipzig, Germany; adrian.viehweger@medizin.uni-leipzig.de
* Correspondence: volkmar.passoth@slu.se; Tel.: +46-1867-3380
† These authors contributed equally to this work.

Abstract: The genus *Rhodotorula* includes basidiomycetous oleaginous yeast species. *Rhodotorula babjevae* can produce compounds of biotechnological interest such as lipids, carotenoids, and biosurfactants from low value substrates such as lignocellulose hydrolysate. High-quality genome assemblies are needed to develop genetic tools and to understand fungal evolution and genetics. Here, we combined short- and long-read sequencing to resolve the genomes of two *R. babjevae* strains, CBS 7808 (type strain) and DBVPG 8058, at chromosomal level. Both genomes are 21 Mbp in size and have a GC content of 68.2%. Allele frequency analysis indicates that both strains are tetraploid. The genomes consist of a maximum of 21 chromosomes with a size of 0.4 to 2.4 Mbp. In both assemblies, the mitochondrial genome was recovered in a single contig, that shared 97% pairwise identity. Pairwise identity between most chromosomes ranges from 82 to 87%. We also found indications for strain-specific extrachromosomal endogenous DNA. A total of 7591 and 7481 protein-coding genes were annotated in CBS 7808 and DBVPG 8058, respectively. CBS 7808 accumulated a higher number of tandem duplications than DBVPG 8058. We identified large translocation events between putative chromosomes. Genome divergence values between the two strains indicate that they may belong to different species.

Keywords: *Rhodotorula babjevae*; de novo hybrid assembly; nanopore sequencing; genome divergence; ploidy

1. Introduction

Oleaginous yeasts have received considerable attention in recent years due to many potential biotechnological applications of microbial lipids. *Rhodotorula* species are basidiomycetous oleaginous yeasts whose lipid production accounts for more than 70% of dry cell weight. They show high tolerance to inhibitors, enabling them to convert lignocellulosic hydrolysates into lipids [1–4]. Microbial lipids from *R. babjevae* and other oleaginous yeasts have a fatty acid composition similar to vegetable oils and represent an environmentally and ethically suitable alternative raw material for the production of biofuels, oleochemicals, feed, and food additives [2,5,6]. Under nitrogen-limited conditions, *R. babjevae* can simultaneously accumulate biotechnologically important enzymes, glycolipids, and carotenoids [5]. Glycolipids from *R. babjevae* have promising environmental applications in biodegrading hydrocarbon pollutants and replacing synthetic compounds and chemical surfactants [7–9]. They are also attractive for other applications in various industrial sectors due to their

antifungal, antibacterial, antiviral, and anti-carcinogenic activities [7–10]. However, it is desirable to obtain more robust *R. babjevae* strains to overcome the high production costs of microbial lipids and biosurfactants.

There are currently no methods described for the molecular manipulation of *R. babjevae* strains. To date, several genomes from *Rhodotorula* sp. have been sequenced including different strains of *R. toruloides*, *R. graminis* WP1, and *R. glutinis* ZHK. Of these, some have only been determined using short-read sequencing technologies or lack gene annotation [3,11–18]. To the best of our knowledge, no genome sequences are available for *R. babjevae*. The aim of this study was to obtain high-quality genome assemblies for *R. babjevae* as a prerequisite for the development of genetic tools, and to deepen our understanding of the biology and evolution of *Rhodotorula* species. To achieve this, we used a combination of short and long reads. This has previously been used successfully to generate high quality genome hybrid assemblies in terms of completeness, contiguity, and chromosome reconstruction [3,12,19,20]. We present here the de novo genome assemblies and annotations of two *R. babjevae* species strains, CBS 7808 (type strain) and DBVPG 8058, based on short- and long-read sequencing technologies. We also performed a genome divergence and ploidy analysis of both *R. babjevae* strains.

2. Materials and Methods

2.1. Yeast Strains

The type strain of *R. babjevae* (CBS 7808) was obtained from the CBS-KNAW collection (Utrecht, The Netherlands). Strain DBVPG 8058 was isolated and identified at the Swedish University of Agricultural Sciences, Uppsala (strain number in the strain collection of the Department of Molecular Sciences is J195) [2] and deposited in the Industrial Yeasts Collection (Perugia, Italy).

2.2. DNA Purification

The yeasts were cultivated in 50 mL Yeast–Peptone–Dextrose medium (YPD) until reaching exponential growth phase [21]. Cell wall degradation was performed according to [22] with some modifications. Briefly, the cells were suspended in 1 M sorbitol, 0.1 M sodium citrate, 0.01 M EDTA, and 0.03 M β-mercaptoethanol (SCEM), pH 5.8 after harvesting. Lyticase solution was added to the cell suspensions (100 U/mL) of CBS 7808 and DBVPG 8058, which were then incubated for 9 h or overnight, respectively. After Lyticase digestion, cells were harvested at $1200\times g$, suspended in SCEM buffer, and incubated overnight with Zymolyase (200 U/mL). Genomic DNA extraction from protoplasts was performed using the NucleoBond® CB 20 Kit (Macherey-Nagel, Düren, Germany). DNA concentration, purity, and quality were confirmed through Qubit™ 4 Fluorometer (Thermo Fisher Scientific, Singapore), NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and agarose gel electrophoresis, respectively.

2.3. Library Preparation and Sequencing

The extracted DNA samples were sequenced using MinION (Oxford Nanopore Technologies, Oxford, UK) and the Illumina sequencing platform. Nanopore DNA libraries were prepared according to [23]. Briefly, 31.5 μL of AMPure magnetic beads were added to 5 μg of DNA for a "pre-cleaning" step. Library preparation was then performed according to a modified protocol [23] using a Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies, Oxford, UK). Each DNA library was loaded onto a FLO-MIN106 flow cell mounted on a MINION device (Oxford Nanopore Technologies). MinKNOW software (version 19.06.8) was used for sequencing as described by [23]. The base calling was run using Guppy version 3.2.4-1—195590e and model HAC-mod (modified base sensitive high accuracy model).

From the 6,665,174 long reads recovered from the CBS 7808 DNA library, the mean read length was 2789.7 bases and the read length N50 5553 bases yielding a total of 18,593 Mbp. For DBVPG 8058, 2,953,255 long reads were retrieved containing a total of 15,702 Mbp.

The mean read length was 5317 bases and the read length N50 7411 bases. Aliquots of the extracted DNA from both *R. babjevae* strains were also subjected to short-read paired-end sequencing using the Illumina Novaseq platform (S prime, 2× 150 bp) and the TruSeq PCR free DNA library preparation kit (Illumina Inc., San Diego, CA, USA). 179,163,622 short reads were recovered from CBS 7808 DNA library, corresponding to a total of 27,053 Mbp. For DBVPG 8058, 203,873,550 short reads were retrieved containing a total of 30,784 Mbp.

2.4. Genome Assembly and Annotation

Genome assembly and annotation was performed using a custom pipeline described elsewhere [3], applying the program versions listed in Table S1. To further improve the annotation of transcripts and exon–intron boundaries, we additionally mapped RNA-Seq data from the closely related *R. toruloides* CBS 14 (PRJEB40807) to the *R. babjevae* genomes as previously described [3]. We used nQuire (v0.0) based on minimap2 short-read mappings (v2.17; no secondary alignments option) and the KmerCountExact script from the BBMap package (https://sourceforge.net/projects/bbmap/; accessed on 25 November 2021) (v38.86) to estimate the ploidy level of the *R. babjevae* strains [24,25]. To compare these methods and our ploidy results of the two *R. babjevae* strains with already published results, we also performed nQuire and KmerCountExact on Illumina sequencing data from *Rhodotorula mucilaginosa* JGTA-S1, accession number SRR5821556 [12].

The reconstruction of lipid metabolic pathway maps was performed using KEGG Mapper version 4.3. The KEGG Orthology (KO) identifiers were affiliated to the annotated transcripts of *R. babjevae* CBS 7808 and *R. babjevae* DBVPG 8058 using KofamKOALA [26] with an e-value cut-off of 0.01.

2.5. Genome Divergence Analysis

Synteny relationship analysis between *R. babjevae* CBS 7808 and *R. babjevae* DBVPG 8058 was performed using NUCmer (MUMmer, version 3.23). The maximum gap between adjacent matches in a cluster were set to 500 and the minimum cluster length to 100. Visualization of NUCmer alignments and other genomic features was performed with Circa (http://omgenomics.com/circa; accessed on 23 June 2021).

The level of sequence divergence between both *R. babjevae* strains as well as with other closely related *Rhodotorula* species, including *R. glutinis* ZHK (JAAGPT010000000.1), *R. graminis* WP1 (JTAO00000000.1) and *R. toruloides* strains CBS 14 (PRJEB40807), CGMCC 2.1609 (LKER00000000.1), VN1 (SJTE00000000.1) and NBRC 0880 (LCTV00000000.2), was evaluated using the alignment-free distance measure K_r [27]. We calculated Average Nucleotide Identity (ANI) values using the web-based calculator available at Kostas Lab [28]. DNA–DNA homology (DDH) was estimated with the Genome-to-Genome Distance Calculator (GGDC) 2.1 (http://ggdc.dsmz.de/distcalc2.php; accessed on 26 June 2021) using the program GBDP2_MUMMER [29].

Whole genome alignments of *R. babjevae* strains were performed using LASTZ (version 7.0.2) implemented in Geneious prime, version 2021.0.1 (Biomatters Ltd., Auckland, New Zealand) [30]. Nucleotide alignment and phylogenetic tree construction using MAFFT v7.450 [31] and PhyML 3.3.20180621 [32] with 100 bootstraps, respectively, were performed on the Geneious prime platform.

Whole genome comparison and identification of orthologous gene clusters and paralogous genes were performed on the web-based OrthoVenn2 platform (https://orthovenn2.bioinfotoolkits.net; accessed on 20 October 2021) using a threshold e-value of 1×10^{-15} and an inflation of 1.5 [33]. To identify duplicated genes (paralogs) with high sequence identity, an all-against-all sequence identity search was performed on the NCBI Genome Workbench version 3.7.0 [34] using BLASTp (BLOSUM62 matrix) with a cut-off e-value of 1×10^{-15}. The output file was screened for protein sequences with at least 70% coverage and 70% sequence identity.

3. Results and Discussion

3.1. Genome Assembly, Ploidy Estimation, and Gene Annotation of R. babjevae Strains

The genome of both *R. babjevae* strains was assembled by a combined approach of long- and short-read sequencing with a coverage depth of about 2000 X. A summary of the genomic data is presented in Table 1. The CBS 7808 draft genome has an overall size of 21,862,387 bp and a GC content of 68.23%. Repetitive sequences make up 5.93% of the total length of the genome, of which 4.98% are single repeats and 0.96% are regions of low complexity. The draft genome of DBVPG 8058 has a total size of 21,522,072 bp and a GC content of 68.24%. The approach identified 6.73% as repetitive sequences, including 5.65% as single repeats and 1.09% as regions of low complexity. The similarity of genome features, such as genome size, GC content, and percentage of repetitive regions, confirms that they are closely related species. The genome size is comparable to that of other *Rhodotorula* species, but the GC content is slightly higher [3,11–13,15,18] (Table 1).

Table 1. Genomic data from *Rhodotorula* species.

Reference	This Study	This Study	[11]	[13]	[3]	[15]
Strain number	*R. babjevae* CBS 7808	*R. babjevae* DBVPG 8058	*R. graminis* WP1	*R. glutinis* ZHK	*R. toruloides* CBS 14	*R. toruloides* NP11
Genome size (Mbp)	21.9	21.5	21.0	21.8	20.5	20.2
Coverage	2058	2122	8.6	470	1514	96
GC content (%)	68.23	68.24	67.76	67.8	61.83	62.05
Bases masked (%)	5.93	6.73	6.5	NA	2.01	2.53
No. Scaffolds	3	1	26	30	3	34
No. Contigs	24	33	325	NA	23	NA
Protein-coding genes	7591	7481	7283 [a]	6774 [a]	9464	8171
Avg. no. exons per gene	4.0	3.9	6.2	NA	5.9	NA
Sequencing platform	Nanopore and Illumina	Nanopore and Illumina	Sanger	PacBio and Illumina	Nanopore and Illumina	Illumina and Sanger

NA—not available; [a]—refers to predicted genes.

Sequence assembly resulted for *R. babjevae* CBS 7808 in 24 contigs and three scaffolds with a length N50 of 1,067,634 bp (Figure 1a, Table S2). A telomeric region was predicted at one of the termini for 13 contigs and scaffolds larger than 250,000 bp. The draft genome of strain DBVPG 8058 consists of 33 contigs and one scaffold with a length N50 of 789,767 bp (Figure 1b, Table S3). From the contigs and scaffolds with sizes larger than 250,000 bp in DBVPG 8058 genome assembly, two have telomere sequences at both termini and 15 at one terminus each. The low numbers of contigs and scaffolds in the genome assemblies from both *R. babjevae* strains indicate high accuracy, contiguity, and completeness. Two putative circular sequences were identified in each strain. Among them, contig_2 in CBS 7808 and contig_79 in DBVPG 8058 contained the mitochondrial genes. Both mitochondrial genomes are similar in size with 30.876 bp and 28.432 bp, respectively, and have a GC content of 38.9% (Tables S2 and S3).

(**a**)

Figure 1. *Cont.*

(**b**)
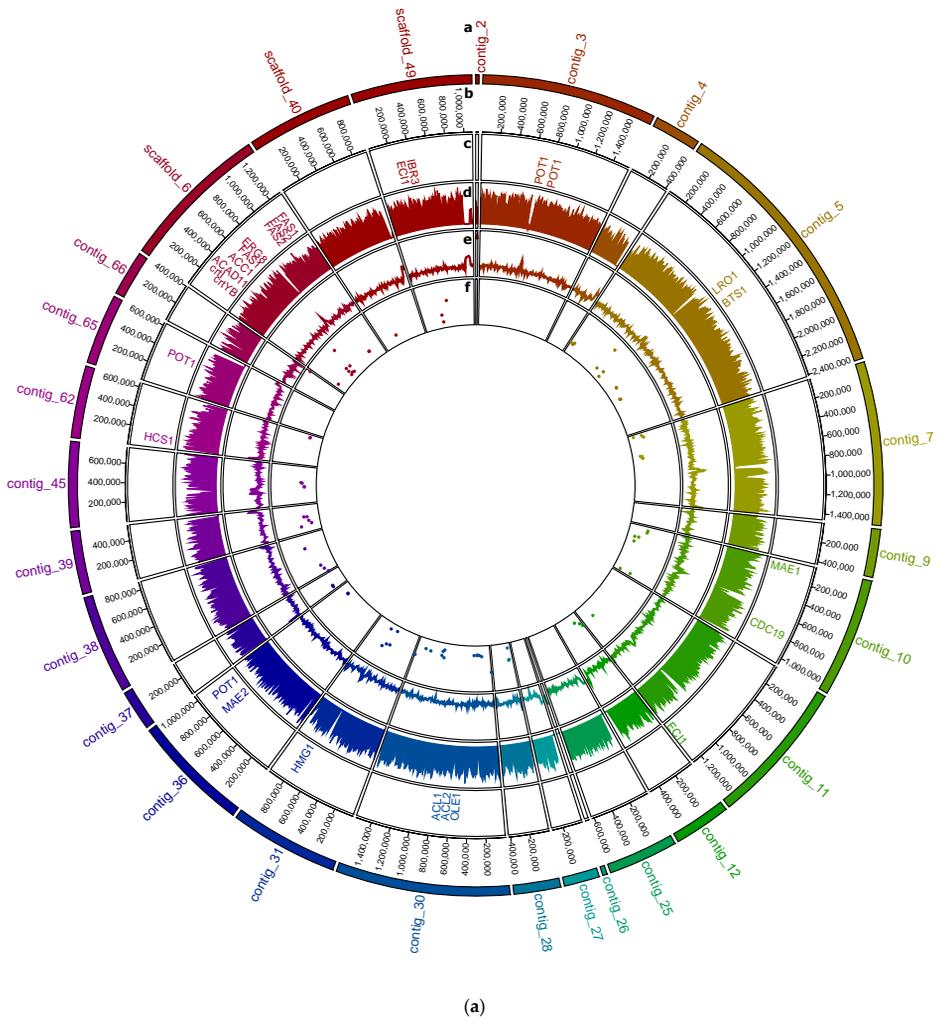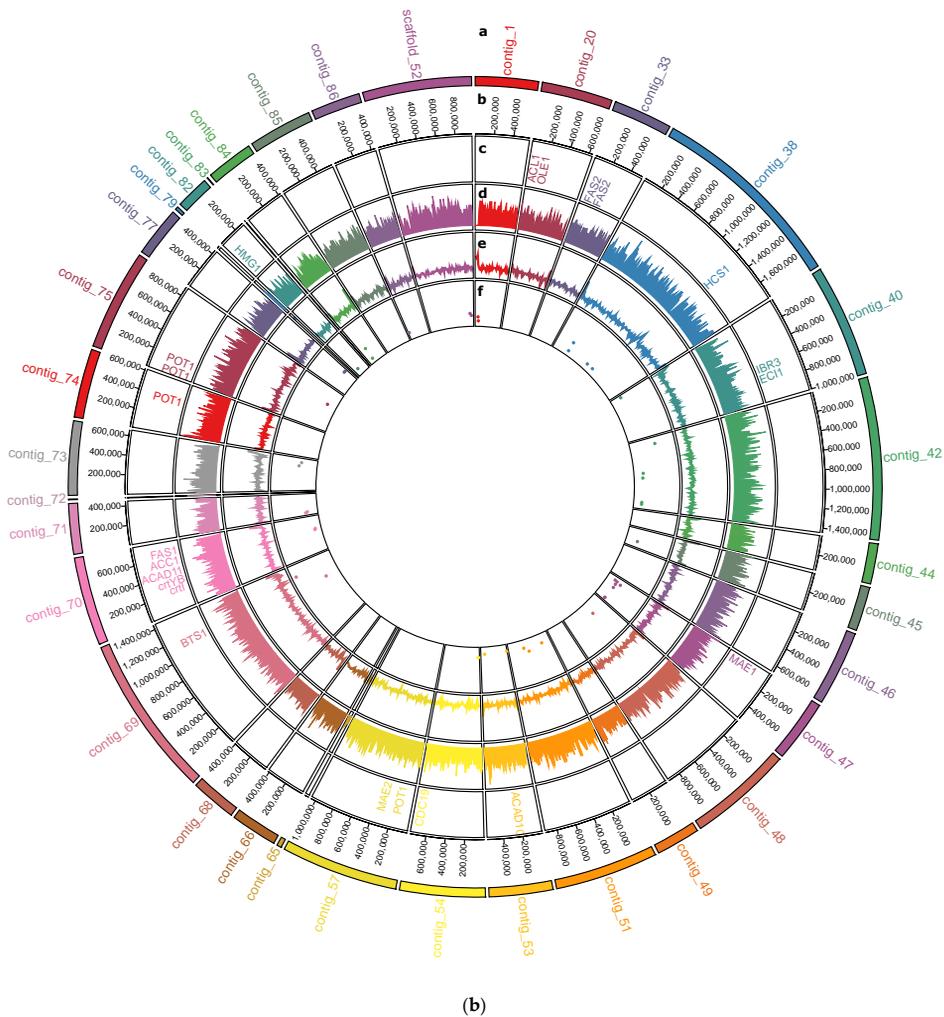
Figure 1. Overview of the genome assemblies of *Rhodotorula babjevae* strains: (**a**) CBS 7808, (**b**) DBVPG 8058. The concentric circles show from outside to inside: the contig name (a) and sizes (b), distribution of lipid and carotenoid metabolism related genes (c), and in non-overlapping 10 kb windows, the gene density (d), the deviation from the average GC content (e), and the density of duplicated genes with 70% sequence coverage (f).

To estimate the ploidy in *R. babjevae* strains, we used nQuire. nQuire quantifies the distribution of the base frequencies at variable sites, and thus differentiates between different degrees of ploidy [24]. In both strains, the alleles occurred at frequencies of about 25% and 75%, indicating that both *R. babjevae* strains are tetraploid (Figure 2). Furthermore, we also used a k-mer counting approach to estimate ploidy. Using a k-mer length of 31, as recently shown by Sen et al. [12] for *R. mucilaginosa* JGTA-S1, only one peak appears in the plots. However, when the k-mer length is reduced to 17, as recently shown by Zou et al. [35],

two distinct peaks appear for both *R. mucilaginosa* JGTA-S1 and the *R. babjevae* strains (Figure 2). The first and larger peak indicates tetraploidy while the second smaller peak indicates diploidy.
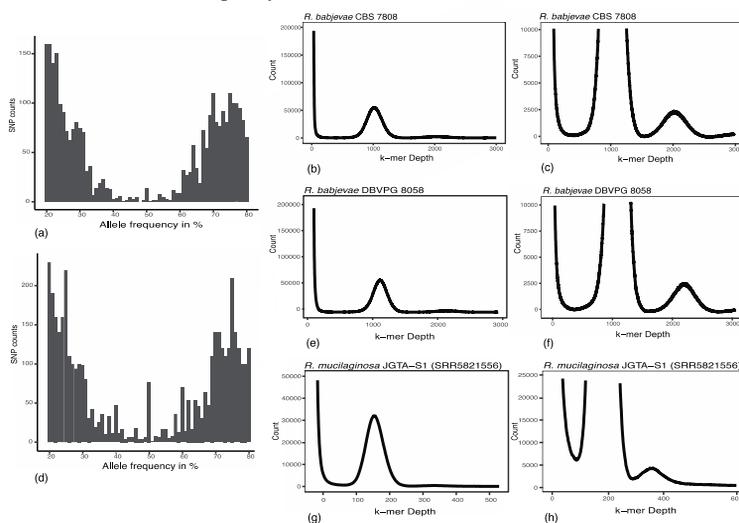
Figure 2. Ploidy estimation of *Rhodotorula babjevae* and *Rhodotorula mucilaginosa*. *R. babjevae* CBS 7808: (**a**) Allele frequency values of single nucleotide polymorphisms (SNP) obtained through nQuire calculations using minimap2, (**b**) Distribution of 17-kmer frequencies using KmerCountExact from the BBmap package, (**c**) Zoomed in peaks of the 17-kmer frequency histogram. *R. babjevae* DBVPG 8058: (**d**) Allele frequency values of SNP (**e**) Distribution of 17-kmer frequencies, (**f**) Zoomed in peaks of the 17-kmer frequency histogram. *R. mucilaginosa* JGTA-S1: (**g**) Distribution of 17-kmer frequencies, (**h**) Zoomed in peaks of the 17-kmer frequency histogram. The reproduction of *R. mucilaginosa* JGTA-S1 ploidy estimation was performed using the Illumina data SRR5821556.

The ploidy level of *R. babjevae* strains has not been studied so far. The genomes of the closely related strains *R. toruloides* NP11 and *R. mucilaginosa* JGTA-S1 are considered to be haploid [12,15]. However, our analyses indicate that both *R. mucilaginosa* JGTA-S1 and *R. babjevae* CBS 7808 and DBVPG 8058 may be tetraploid. Tetraploidy has previously been widely recognized in yeast [36–39]. Knowing the ploidy level is of great importance for genetic engineering and for the development of efficient gene manipulation protocols.

A total of 7591 protein-coding genes and 7607 associated transcripts were annotated in the CBS 7808 genome using MetaEuk (Table 1). The average number of estimated exons per gene is 3.97 (Table 1). The genome of DBVPG 8058 has 7481 protein-coding genes, 7516 associated transcripts and 3.93 estimated exons per gene (Table 1). The proportion of split genes in both genomes is correspondingly high, amounting to 6390 for CBS 7808 and DBVPG 8058, respectively. This is consistent with previous findings for *Rhodotorula* spp. [3,12,15]. The distribution of exon counts in the genomes of *R. babjevae* strains CBS 7808 and DBVPG 8058 is shown in Table S4. 315 and 309 open reading frames (ORF) complementary to annotated genes were predicted in CBS 7808 and DBVPG 8058, respectively. The presence of antisense transcripts has previously been reported for the related species *R. toruloides* [3]. In yeast, the level of antisense transcription has been anti-correlated to sense mRNA, indicating antisense-dependent gene regulation through transcription interference under certain growth conditions [40,41]. Figures S1–S3 show the assignment of genes to the Gene Ontology (GO) categories' biological processes, cellular components, and molecular functions, of which the top 10 are summarized in Figure 3a,b.

A total of 2691 and 2660 CDS from CBS 7808 and DBVPG 8058, respectively, could be assigned KO numbers (Figure 3c). The biosynthesis of saturated and unsaturated fatty acids, glycerolipid metabolism, terpenoid backbone biosynthesis, carbon metabolism, and fatty acid metabolism are depicted in detail in Figures S4 and S5. Some examples of annotated genes that encode crucial enzymes for lipid and carotenoid metabolism are *CDC19*, *MAE1*, *MAE2*, *ACL1*, *ACL2*, *ACC1*, *FAS1*, *FAS2*, *OLE1*, *ACAD10*, *ACAD11*, *IBR3*, *D6C81_05617*, *POT1*, *LRO1*, *HMG1*, *HCS1*, *ERG8*, *crtYB*, *crtI*, and *BTS1* (Tables S5 and S6). A difference in this respect is the absence of *ACL2*, and the presence of *ACAD10* in DBVPG 8058.

Benchmarking of universal single-copy orthologs (BUSCOs, using fungi_odb9) identified that 95.5% and 96.9% of the assessed genes in CBS 7808 and DBVPG 8058, respectively, were complete and single-copy (Figure S6). This supports the high quality of the draft genome assemblies reported here. Furthermore, 0.7% and 0.3% of the assessed genes in CBS 7808 and DBVPG 8058, respectively, were fragmented and the rest were missing (Figure S6). A small percentage of BUSCO genes might still be undetectable due to sequence regions with low coverage, repetitive elements, or assembly problems that cannot be solved even with the hybrid approach and would require additional sequencing and manual analysis. In addition, when a BUSCO gene was missing, there were either no significant matches or the BUSCO matches were below the range of values for the selected BUSCO profile. Finally, some marker genes that are part of the BUSCO "fungi" profile that we used as reference may not be part of the two *R. babjevae* strains.

(**a**)　　　　　　　　(**b**)

Figure 3. *Cont.*

(**c**)
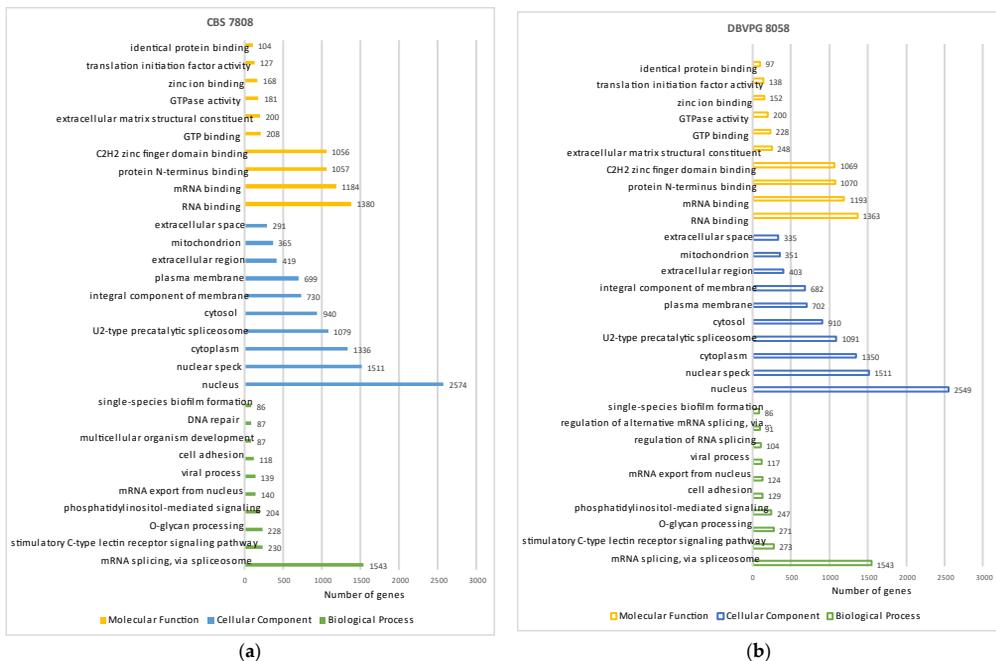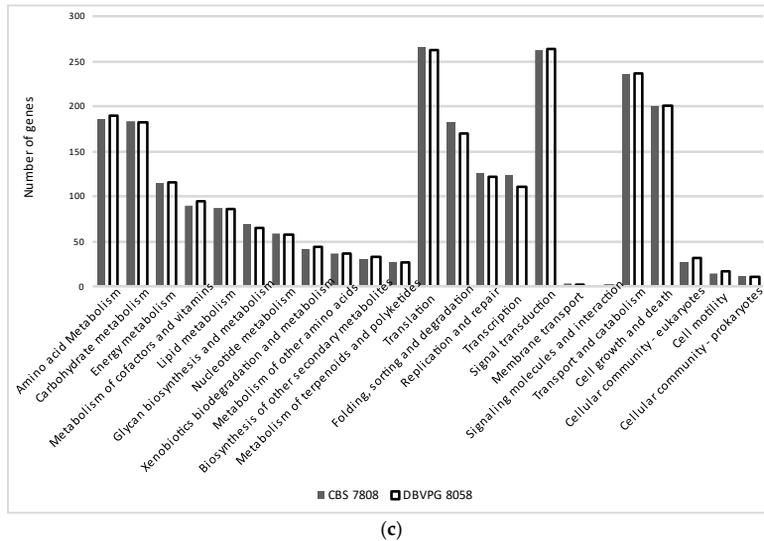
Figure 3. Assigned numbers of genes to the top 10 of the GO categories: biological processes, cellular components, and molecular functions in CBS 7808 (**a**) and DBVPG 8058 (**b**). (**c**) Assigned number of genes to the five KEGG top categories: metabolism, genetic information processing, environmental information processing, cellular processes, and organismal systems. KEGG (Kyoto Encyclopedia of Genes and Genomes), GO (Gene Ontology).

3.2. Chromosome Organization

The *R. babjevae* genome assemblies were aligned for comparison using NUCmer. Out of a total of 27 contigs and scaffolds in CBS 7808, 24 matched 30 of the 34 assembled sequences in DBVPG 8058 (Figure 4). In general, the number of undisturbed segments is high. However, there are also major chromosomal rearrangements (Figure 4). LASTZ alignments of each contig from one *R. babjevae* strain with the whole genome of the other strain confirmed the results of the synteny analysis (Table S7, Figures S7 and S8). Based on these alignments we deduce that *R. babjevae* has a maximum of 21 chromosomes with sizes ranging from 0.4 to 2.4 Mbp (Table 2, Figure 4). The molecular karyotype of several *Saccharomyces* yeast strains has been identified as 16 [42]. Karyotyping studies in *Rhodotorula* species have identified at least 10 chromosomes in isolates of *R. mucilaginosa* and 11 in *R. toruloides* while Martín-Hernández et al. proposed that *R. toruloides* CBS 14 has at least 18 chromosomes [3,43,44]. The pairwise identity between chromosomes ranges from 82% to 87%. The mitochondrial genomes have 97% pairwise identity (Table S7, Figure S7). Four of the putative chromosomes are affected by large translocation events. This affects chromosomes 3 and 6, and chromosomes 9 and 14 (Table 2). Minor inversions were noticed in other chromosomes (Table S7, Figure S8). Each *R. babjevae* strain contains two contigs that are strain-specific (Tables S7 and S8). These are small linear contigs with higher read depths than the chromosomes, except for circular contig_26 in CBS 7808, which has a lower read depth than the chromosomes. These variations in read depth may indicate relaxed replication regulation. The linear DNA sequence from CBS 7808 contig_46 has two annotated genes, one of which encodes the Retrovirus-related Pol polyprotein from transposon 17.6. DNA plasmids have previously been found in filamentous fungi, including the close relative *R. toruloides*, with sizes ranging from 2.5 to 11 kb and typically encoding enzymes involved in plasmid replication [3,45,46]. This might indicate the presence of strain-specific extrachromosomal endogenous DNA.
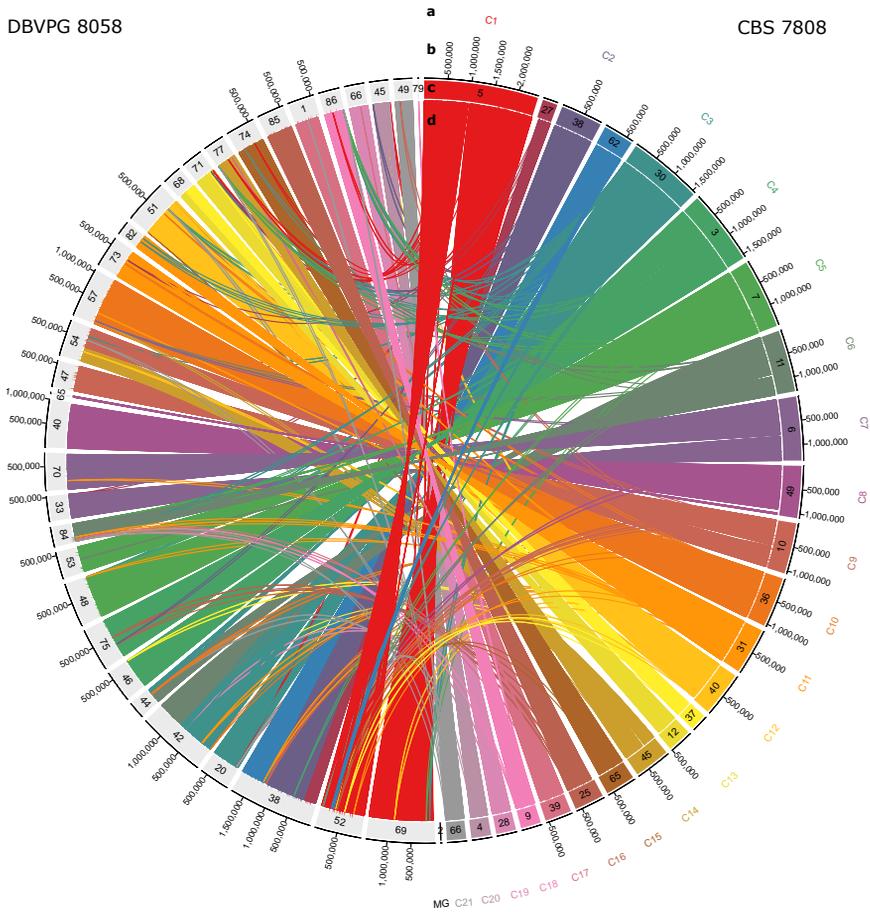
Figure 4. Genome alignment of *Rhodotorula babjevae* strains CBS 7808 and DBVPG 8058. Maximal unique matches between CBS 7808 and DBVPG 8058 were obtained using NUCmer 3.0 and visualized with Circa. The concentric circles show from outside to inside: putative chromosome names or mitochondrial genome, MG, in the reference strain CBS 7808 (a); contig and scaffolds' sizes (b); and names (c). Ribbons are showing the unique and repetitive alignments using CBS 7808 contigs and scaffolds as the reference (d). Contigs from DBVPG 8058 are colored gray.

Table 2. Putative chromosomes in *Rhodotorula babjevae* deduced from whole genome LASTZ alignments (Table S7, Figure S8).

R. bajevae CBS 7808	*R. bajevae* DBVPG 8058	Genetic Structure	GC Content	Comments	Size (Mbp)
Contig_5 (2,415,752 bp)	Contig_69 (1,447,990 bp) Scaffold_52 (977,625 bp)	Putative chromosome 1	67–69%	Figure S8A	2.4
Contig_27 (320,063 bp) Contig_38 (881,966 bp) Contig_62 (644,441 bp)	Contig_38 (1,780,658 bp)	Putative chromosome 2	67–69%	Figure S8B	1.8
Contig_30 (1,569,459 bp)	Contig_20 (637,402 bp) Contig_42 (1,446,680 bp) Contig_44 (357,974 bp)	Putative chromosome 3	67–69%	Figure S8C Large translocation event between Chr. 3 and Chr.6	1.6
Contig_3 (1,574,520 bp)	Contig_46 (670,828 bp) Contig_75 (900,917 bp)	Putative chromosome 4	67–69%	Figure S8D	1.6
Contig_7 (1,460,653 bp)	Contig_48 (931,129 bp) Contig_53 (571,073 bp)	Putative chromosome 5	67–69%	Figure S8E	1.5
Contig_11 (1,300,441 bp)	Contig_42 (1,446,680 bp) Contig_44 (357,974 bp) Contig_84 (425,340 bp)	Putative chromosome 6	67–69%	Figure S8F Large translocation event between Chr. 3 and Chr.6	1.3
Scaffold_6 (1,337,997 bp)	Contig_33 (529,001 bp) Contig_70 (789,767 bp)	Putative chromosome 7	67–69%	Figure S8G	1.3
Scaffold_49 (1,089,446 bp)	Contig_40 (1,004,683 bp) Contig_65 (41,334 bp)	Putative chromosome 8	67–69%	Figure S8H	1.1
Contig_10 (1,067,634 bp)	Contig_47 (557,103 bp) Contig_54 (766,724 bp)	Putative chromosome 9	67–69%	Large translocation event between Chr. 9 and Chr.14 Figure S8I	1.1
Contig_36 (1,056,323 bp)	Contig_57 (1,049,892 bp)	Putative chromosome 10	67–69%	Figure S8J	1.1

Table 2. *Cont.*

R. bajevae CBS 7808	*R. bajevae* DBVPG 8058	Genetic Structure	GC Content	Comments	Size (Mbp)
Contig_31 (979,228 bp)	Contig_73 (659,761 bp) Contig_82 (299,180 bp)	Putative chromosome 11	67–69%	Figure S8K	1.0
Scaffold_40 (948,604 bp)	Contig_51 (924,743 bp)	Putative chromosome 12	67–69%	Figure S8L	0.9
Contig 37 (362,520 bp) Contig_12 (511,897 bp)	Contig_68 (408,627 bp) Contig_71 (449,691 bp)	Putative chromosome 13	67–69%	Figure S8M	0.9
Contig_45 (762,860 bp)	Contig_54 (766,724 bp) Contig_77 (446,828 bp)	Putative chromosome 14	67–69%	Figure S8N Large translocation event between Chr. 9 and Chr.14	0.8
Contig_65 (630,535 bp)	Contig_74 (614,034 bp)	Putative chromosome 15	67–69%	Figure S8O	0.6
Contig_25 (627,118 bp)	Contig_85 (573,802 bp)	Putative chromosome 16	67–69%	Figure S8P	0.6
Contig_39 (564,129 bp)	Contig_1 (565,532 bp)	Putative chromosome 17	67–69%	Figure S8Q	0.6
Contig_9 (429,397 bp)	Contig_86 (443,617 bp)	Putative chromosome 18	67–69%	Figure S8R	0.4
Contig_28 (422,133 bp)	Contig_66 (419,035 bp)	Putative chromosome 19	67–69%	Figure S8S	0.4
Contig_4 (418,972 bp)	Contig_45 (394,205 bp)	Putative chromosome 20	67–69%	Figure S8T	0.4
Contig_66 (406,102 bp)	Contig_49 (396,114 bp)	Putative chromosome 21	67–69%	Figure S8U	0.4

Chr., chromosome.

3.3. Genome Divergence Analysis

The genomes of the *R. babjevae* strains were compared to each other and to genomes of closely related *Rhodotorula* species in terms of DDH, ANI and K_r for tracing genome divergence (Figure 5, Table S9). The *R. babjevae* strains share 44.20% DDH estimates, 84.48% ANI and K_r values of 0.09. In general, the genetic divergence between *R. babjevae* strains was comparable to the divergence with *R. graminis* and *R. glutinis*, but higher than expected for strains of the same yeast species [47]. For instance, the divergence between strains of *R. toruloides* was much lower than that of the two *R. babjevae* strains (Table S9).

	DDH	ANI	kr
	22.40%	77.01%	0.221540
	44.20%	84.48%	0.088959
	39.50%	82.88%	0.096059
	41.40%	84.07%	0.075659

(**a**)

	DDH	ANI	kr
	22.30%	77.33%	0.223232
	39.50%	82.88%	0.096059
	44.20%	84.48%	0.088959
	40.40%	83.43%	0.095288

(**b**)
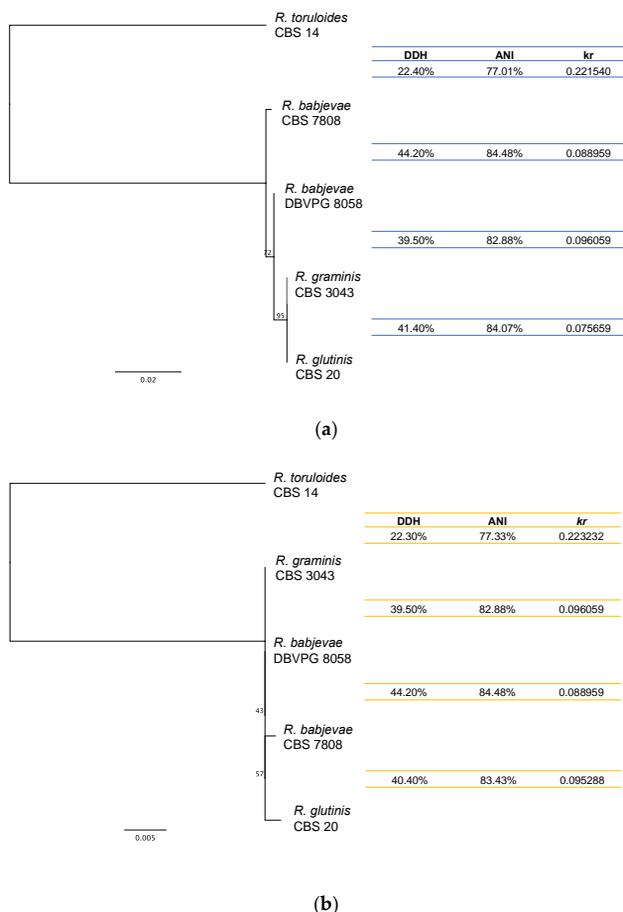
Figure 5. Phylogenetic relationship of *Rhodotorula babjevae* strains and their placement within the *Rhodotorula* genus. The phylogenetic tree was built based on: (**a**) ITS; and (**b**) D1/D2 LSU of rRNA gene sequences. It was inferred using PhyML with 100 bootstraps on Geneious prime version 2021.0.1. *Rhodotorula toruloides* was selected as outgroup. Similarities between whole genome sequences of the corresponding strains are presented in terms of the alignment-free distance measure *kr*, Average Nucleotide Identity (ANI), and DNA–DNA homology (DDH). *Rhodotorula graminis* WP1 and *R. glutinis* ZHK genome sequences were used for the calculations instead of *R. graminis* CBS 3043 and *R. glutinis* CBS 20, respectively.

Moreover, the protein-coding sequences of the *R. babjevae* strains and their closest relatives *R. graminis* and *R. glutinis* were analyzed using OrthoVenn2 web platform to identify and compare orthologous gene clusters. The *R. babjevae* species share 6598 out of a total of 7223 orthologous clusters produced by OrthoVenn2, including both single-copy gene clusters and overlapping gene clusters such as paralogs (Figure 6). Of the shared clusters, 5933 are common within the three *Rhodotorula* species assessed, representing putative shared orthologous proteins that evolved from common ancestral genes. In addition, CBS 7808 has 389 single genes and one cluster that had no orthologs in the

other genomes, while strain DBVPG 8058 has 355 single genes. These unique genes could account for the specific functional capabilities of the *R. babjevae* strains as a result of gene loss or gain events. Of the 79 orthologous clusters shared only between *R. babjevae* strains, some of the assigned GO terms are: Positive regulation of the unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter (GO:0036083), protein O-linked glycosylation (GO:0006493), glucan catabolic process (GO:0009251), cellular calcium ion homeostasis (GO:0006874), sulfate assimilation (GO:0000103), and carbohydrate transport (GO:0008643). The two *R. babjevae* strains show a high genome pairwise similarity and a high number of shared orthologous clusters, though not as high as for *R. graminis* and *R. glutinis* (Figure 6). In general, *R. babjevae*, *R. glutinis*, and *R. graminis* are very closely related species with a short evolutionary distance between them as compared to other species in the genus (i.e., *R. toruloides*). The strains CBS 7808 and DBVPG 8058 have high interstrain variability and a greater evolutionary distance to *R. graminis* than to *R. glutinis*.
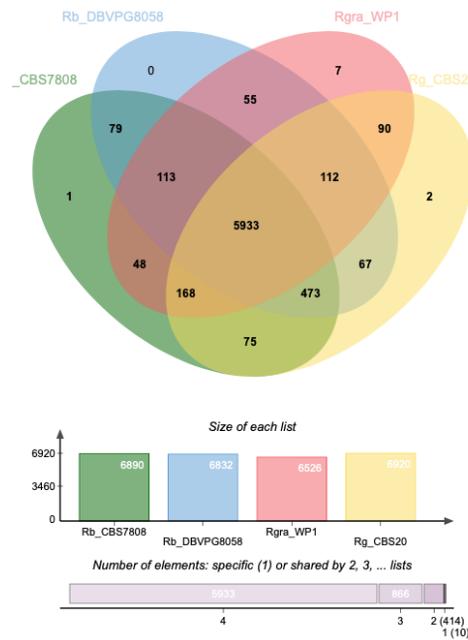
Figure 6. Distribution of shared orthologous clusters between *Rhodotorula babjevae* strains CBS 7808 and DBVPG 8058, *R. graminis* WP1 and *R. glutinis* CBS 20. The Venn diagram was generated using OrthoVenn2.

A total of 59 and 30 paralogous gene clusters were identified in CBS 7808 and DBVPG 8058, respectively, using OrthoVenn2 (Tables S10 and S11). Applying a cut-off value of 70% sequence coverage to them, we identified 29 and 19 duplicated genes, respectively, that potentially have not diverged in function. On the other hand, an all-against-all protein sequence similarity search was performed in each of the two strains using BLASTp with an e-value of 1×10^{-15}, 70% coverage, and 70% sequence identity. This resulted in a total of 34 and 21 duplicated sequences in CBS 7808 and DBVPG 8058, respectively, and a total of 41 and 29 duplicated sequences with 70% sequence coverage, respectively, that were identified by any of the tools (Figure 1, Tables S12–S14). The higher accumulation of duplicated genes in CBS 7808 could be related to a higher number of gene duplication events due

to faster evolution of the strain. The majority of these duplications lies adjacent to each other or in close proximity. Tandem duplications have been suggested as a mechanism of adaptive evolution to changing environments [48]. They may have arisen through homologous recombination between sequences on sister chromatids or homologous chromosomes [48]. It has been reported that considerable redundancy of duplicate gene pairs persists even after 100 million years of evolution in *Saccharomyces cerevisiae* [49]. Some of the predicted functions of the genes, which are duplicated only in CBS 7808, are Uncharacterized protein C17G8.02 (NAD biosynthesis), Mannose-6-phosphate isomerase and Phosphoenolpyruvate carboxykinase (ATP) (carbon metabolism), Acetyl-CoA carboxylase (fatty acid metabolism), Alpha-ketoglutarate-dependent sulfonate dioxygenase, Sulfite reductase [NADPH] hemoprotein beta-component and Sulfite reductase [NADPH] subunit beta (Sulfur metabolism), and Probable quinate permease (import of quinic acid as a carbon source). Some of the duplicated genes involved in metabolic processes identified only in DBVPG 8058 are mitochondrial Aspartate aminotransferase (intracellular NAD(H) redox balance) and Leucine-rich repeat extensin-like protein 3 (AtLRX3, cell morphogenesis). In both strains, the most frequently duplicated gene is *SRRM2*, which codes for the Ser/Arg repetitive matrix protein 2 and is involved in mRNA splicing. Cwc21p is encoded by *CWC21*, an ortholog of human *SRRM2* in *S. cerevisiae*. It has been proposed that it resides in the catalytic center of the spliceosome and possibly fulfills its role in response to changing cellular environmental conditions [50]. The predicted function Ser/Arg repetitive matrix protein 2 was annotated in 1055 genes in CBS 7808 and 1068 in DBVPG 8058. Alternative splicing is an essential driver of proteomic diversity and may potentially provide a high level of evolutionary plasticity.

The type strain CBS7808 of *R. babjevae* investigated here, was first isolated from herbaceous plants in Moscow, Russia [51]. *R. babjevae* DBVPG 8058 was isolated from wild apples in Uppland locality, Sweden. The phylogenetic placement of DBVPG 8058 to the *R. babjevae* species was performed by the Industrial Yeasts Collection DBVPG by aligning 5.8S-ITS rDNA and D1/D2 26S rDNA regions in a similar manner as illustrated in Figure 5. However, the genome divergence values (DDH, ANI and K_r) proved to be more sensitive for delineating *Rhodotorula* species. Phylogenetic placement based on the standard rDNA regions may not be sufficient to understand yeast diversity and species delineation, as shown before [47,52,53]. These *R. babjevae* strains showed different behavior during enzymatic cell wall degradation for DNA purification both in this study and in another study where xylose medium was used [54]. Highly dynamic genome structures have already been found in closely related yeast species [20,55–59]. A dynamic genome structure of *R. babjevae* could enhance the physiological capabilities and thus the species' environmental adaptability. [60–63]. However, their genetic divergence suggests that they may belong to different species. A genome comparison study using whole genome sequences from different strains of closely related *Rhodotorula* species would allow gaining a deeper knowledge about their genome structure and evolution, as well as identifying new species.

Taxonomic classification using Sourmash [64] and the GenBank reference (https://osf.io/4f8n3; accessed on 17 March 2022) assign to both genome assemblies: Eukaryota superkingdom, Basidiomycota phylum, Microbotryomycetes class, Sporidiobolales order, Sporidiobolaceae family, *Rhodotorula* genus, and *Rhodotorula graminis* species. The taxonomic classification might indicate that *R. graminis* was the closest relative of *R. babjevae* with available genomic data. Previous studies have shown a close evolutionary relationship between *R. babjevae* and *R. graminis*, which was also demonstrated here [13,65].

4. Conclusions

The hybrid sequencing approach resulted in high-resolution genomes of *R. babjevae* DBVPG 8058 and CBS 7808[T]. Both strains are tetraploid and have a maximum of 21 chromosomes. Some of the chromosomes show large-scale translocation events. Moreover, we demonstrated a high genome divergence between the *R. babjevae* strains, as high as the

divergence to other closely related *Rhodotorula* species. This indicates that the two strains do not belong to the same species.

References

1. Poontawee, R.; Yongmanitchai, W.; Limtong, S. Efficient oleaginous yeasts for lipid production from lignocellulosic sugars and effects of lignocellulose degradation compounds on growth and lipid production. *Process Biochem.* **2017**, *53*, 44–60. [CrossRef]
2. Brandenburg, J.; Poppele, I.; Blomqvist, J.; Puke, M.; Pickova, J.; Sandgren, M.; Rapoport, A.; Vedernikovs, N.; Passoth, V. Bioethanol and lipid production from the enzymatic hydrolysate of wheat straw after furfural extraction. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 6269–6277. [CrossRef] [PubMed]
3. Martín-Hernández, G.C.; Müller, B.; Chmielarz, M.; Brandt, C.; Hölzer, M.; Viehweger, A.; Passoth, V. Chromosome-level genome assembly and transcriptome—Based annotation of the oleaginous yeast *Rhodotorula toruloides* CBS 14. *Genomics* **2021**, *113*, 4022–4027. [CrossRef]
4. Chmielarz, M.; Blomqvist, J.; Sampels, S.; Sandgren, M.; Passoth, V. Microbial lipid production from crude glycerol and hemicellulosic hydrolysate with oleaginous yeasts. *Biotechnol. Biofuels* **2021**, *14*, 65. [CrossRef]
5. Ayadi, I.; Belghith, H.; Gargouri, A.; Guerfali, M. Screening of new oleaginous yeasts for single cell oil production, hydrolytic potential exploitation and agro-industrial by-products valorization. *Process Saf. Environ. Prot.* **2018**, *119*, 104–114. [CrossRef]
6. Blomqvist, J.; Pickova, J.; Tilami, S.K.; Sampels, S.; Mikkelsen, N.; Brandenburg, J.; Sandgren, M.; Passoth, V. Oleaginous yeast as a component in fish feed. *Sci. Rep.* **2018**, *8*, 15945. [CrossRef]
7. Guerfali, M.; Ayadi, I.; Mohamed, N.; Ayadi, W.; Belghith, H.; Bronze, M.R.; Ribeiro, M.H.L.; Gargouri, A. Triacylglycerols accumulation and glycolipids secretion by the oleaginous yeast *Rhodotorula babjevae* Y-SL7: Structural identification and biotechnological applications. *Bioresour. Technol.* **2019**, *273*, 326–334. [CrossRef]
8. Sen, S.; Borah, S.N.; Bora, A.; Deka, S. Production, characterization, and antifungal activity of a biosurfactant produced by *Rhodotorula babjevae* YS3. *Microb. Cell Fact.* **2017**, *16*, 95. [CrossRef]
9. Seveiri, R. Characterization and prospective applications of the exopolysaccharides produced by *Rhodosporidium babjevae*. *Adv. Pharm. Bull.* **2020**, *10*, 254–263. [CrossRef]
10. Sen, S.; Borah, S.N.; Kandimalla, R.; Bora, A.; Deka, S. Sophorolipid Biosurfactant Can Control Cutaneous Dermatophytosis Caused by *Trichophyton mentagrophytes*. *Front. Microbiol.* **2020**, *11*, 329. [CrossRef]
11. Firrincieli, A.; Otillar, R.; Salamov, A.; Schmutz, J.; Khan, Z.; Redman, R.S.; Fleck, N.D.; Lindquist, E.; Grigoriev, I.V.; Doty, S.L. Genome sequence of the plant growth promoting endophytic yeast *Rhodotorula graminis* WP1. *Front. Microbiol.* **2015**, *6*, 978. [CrossRef]
12. Sen, D.; Paul, K.; Saha, C.; Mukherjee, G.; Nag, M.; Ghosh, S.; Das, A.; Seal, A.; Tripathy, S. A unique life-strategy of an endophytic yeast *Rhodotorula mucilaginosa* JGTA-S1—A comparative genomics viewpoint. *DNA Res.* **2019**, *26*, 131–146. [CrossRef]
13. Li, C.J.; Zhao, D.; Cheng, P.; Zheng, L.; Yu, G.H. Genomics and lipidomics analysis of the biotechnologically important oleaginous red yeast *Rhodotorula glutinis* ZHK provides new insights into its lipid and carotenoid metabolism. *BMC Genomics* **2020**, *21*, 834. [CrossRef] [PubMed]
14. Gan, H.M.; Thomas, B.N.; Cavanaugh, N.T.; Morales, G.H.; Mayers, A.N.; Savka, M.A.; Hudson, A.O. Whole genome sequencing of *Rhodotorula mucilaginosa* isolated from the chewing stick (*Distemonanthus benthamianus*): Insights into *Rhodotorula* phylogeny, mitogenome dynamics and carotenoid biosynthesis. *PeerJ* **2017**, *5*, e4030. [CrossRef] [PubMed]
15. Zhu, Z.; Zhang, S.; Liu, H.; Shen, H.; Lin, X.; Yang, F.; Zhou, Y.J.; Jin, G.; Ye, M.; Zou, H.; et al. A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. *Nat. Commun.* **2012**, *3*, 1111–1112. [CrossRef] [PubMed]
16. Fakankun, I.; Fristensky, B.; Levin, D.B. Genome sequence analysis of the oleaginous yeast, *Rhodotorula diobovata*, and comparison of the carotenogenic and oleaginous pathway genes and gene products with other oleaginous yeasts. *J. Fungi* **2021**, *7*, 320. [CrossRef] [PubMed]
17. Tkavc, R.; Matrosova, V.Y.; Grichenko, O.E.; Gostincar, C.; Volpe, R.P.; Klimenkova, P.; Gaidamakova, E.K.; Zhou, C.E.; Stewart, B.J.; Lyman, M.G.; et al. Prospects for fungal bioremediation of acidic radioactive waste sites: Characterization and genome sequence of *Rhodotorula taiwanensis* MD1149. *Front. Microbiol.* **2018**, *8*, 2528. [CrossRef]
18. Goordial, J.; Raymond-Bouchard, I.; Riley, R.; Ronholm, J.; Shapiro, N.; Woyke, T.; LaButti, K.M.; Tice, H.; Amirebrahimi, M.; Grigoriev, I.V.; et al. Improved high-quality draft genome sequence of the eurypsychrophile *Rhodotorula* sp. JG1b, isolated from permafrost in the hyperarid upper-elevation McMurdo Dry Valleys, Antarctica. *Genome Announc.* **2016**, *4*, 15–17. [CrossRef]
19. Olsen, R.A.; Bunikis, I.; Tiukova, I.; Holmberg, K.; Lötstedt, B.; Pettersson, O.V.; Passoth, V.; Käller, M.; Vezzi, F. De novo assembly of *Dekkera bruxellensis*: A multi technology approach using short and long-read sequencing and optical mapping. *Gigascience* **2015**, *4*, 56. [CrossRef]
20. Tiukova, I.A.; Pettersson, M.E.; Hoeppner, M.P.; Olsen, R.A.; Käller, M.; Nielsen, J.; Dainat, J.; Lantz, H.; Söderberg, J.; Passoth, V. Chromosomal genome assembly of the ethanol production strain CBS 11270 indicates a highly dynamic genome structure in the yeast species *Brettanomyces bruxellensis*. *PLoS ONE* **2019**, *14*, e0215077. [CrossRef]
21. Chmielarz, M.; Sampels, S.; Blomqvist, J.; Brandenburg, J.; Wende, F.; Sandgren, M.; Passoth, V. FT-NIR: A tool for rapid intracellular lipid quantification in oleaginous yeasts. *Biotechnol. Biofuels* **2019**, *12*, 169. [CrossRef] [PubMed]
22. Pi, H.W.; Anandharaj, M.; Kao, Y.Y.; Lin, Y.J.; Chang, J.J.; Li, W.H. Engineering the oleaginous red yeast *Rhodotorula glutinis* for simultaneous β-carotene and cellulase production. *Sci. Rep.* **2018**, *8*, 2–11. [CrossRef] [PubMed]
23. Brandt, C.; Bongcam-Rudloff, E.; Müller, B. Abundance tracking by long-read nanopore sequencing of complex microbial communities in samples from 20 different biogas/wastewater plants. *Appl. Sci.* **2020**, *10*, 7518. [CrossRef]

24. Weiß, C.L.; Pais, M.; Cano, L.M.; Kamoun, S.; Burbano, H.A. nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinform.* **2018**, *19*, 122. [CrossRef]
25. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef]
26. Aramaki, T.; Blanc-Mathieu, R.; Endo, H.; Ohkubo, K.; Kanehisa, M.; Goto, S.; Ogata, H. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **2020**, *36*, 2251–2252. [CrossRef]
27. Gremme, G.; Steinbiss, S.; Kurtz, S. Genome tools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2013**, *10*, 645–656. [CrossRef]
28. Rodriguez-R, L.M.; Konstantinidis, K.T. The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* **2016**, *4*, e1900v1. [CrossRef]
29. Meier-Kolthoff, J.P.; Auch, A.F.; Klenk, H.P.; Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **2013**, *14*, 60. [CrossRef]
30. Harris, R.S. Improved Pairwise Alignment of Genomic DNA. Ph.D. Thesis, The Pennsylvania State University, State College, PA, USA, 2007.
31. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
32. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef]
33. Xu, L.; Dong, Z.; Fang, L.; Luo, Y.; Wei, Z.; Guo, H.; Zhang, G.; Gu, Y.Q.; Coleman-Derr, D.; Xia, Q.; et al. OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **2019**, *47*, W52–W58. [CrossRef]
34. Kuznetsov, A.; Bollin, C.J. NCBI Genome Workbench: Desktop software for comparative genomics, visualization, and GenBank data submission. *Methods Mol. Biol.* **2021**, *2231*, 261–295. [CrossRef]
35. Zou, C.; Chen, A.; Xiao, L.; Muller, H.M.; Ache, P.; Haberer, G.; Zhang, M.; Jia, W.; Deng, P.; Huang, R.; et al. A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Res.* **2017**, *27*, 1327–1340. [CrossRef]
36. Krahulec, J.; Lišková, V.; Boňková, H.; Lichvariková, A.; Šafranek, M.; Turňa, J. The ploidy determination of the biotechnologically important yeast *Candida utilis*. *J. Appl. Genet.* **2020**, *61*, 275–286. [CrossRef] [PubMed]
37. Fijarczyk, A.; Hénault, M.; Marsit, S.; Charron, G.; Fischborn, T.; Nicole-Labrie, L.; Landry, C.R. The genome sequence of the Jean-Talon strain, an archeological beer yeast from Québec, reveals traces of adaptation to specific brewing conditions. *G3 Genes Genomes Genet.* **2020**, *10*, 3087–3097. [CrossRef]
38. Gallone, B.; Steensels, J.; Prahl, T.; Soriaga, L.; Saels, V.; Herrera-Malaver, B.; Merlevede, A.; Roncoroni, M.; Voordeckers, K.; Miraglia, L.; et al. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **2016**, *166*, 1397–1410.e16. [CrossRef] [PubMed]
39. Peter, J.; Chiara, M.D.; Friedrich, A.; Yue, J.; Pflieger, D.; Bergström, A.; Sigwalt, A.; Barre, B.; Freel, K.; Llored, A.; et al. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* **2018**, *556*, 339–344. [CrossRef] [PubMed]
40. Nevers, A.; Doyen, A.; Malabat, C.; Néron, B.; Kergrohen, T.; Jacquier, A.; Badis, G. Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. *Nucleic Acids Res.* **2018**, *46*, 6009–6025. [CrossRef]
41. Wery, M.; Gautier, C.; Descrimes, M.; Yoda, M.; Vennin-Rendos, H.; Migeot, V.; Gautheret, D.; Hermand, D.; Morillon, A. Native elongating transcript sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast. *RNA* **2018**, *24*, 196–208. [CrossRef]
42. Borovkova, A.N.; Michailova, Y.V.; Naumova, E.S. Molecular Genetic Features of Biological Species of the Genus *Saccharomyces*. *Microbiology* **2020**, *89*, 387–395. [CrossRef]
43. Białkowska, A.M.; Szulczewska, K.M.; Krysiak, J.; Florczak, T.; Gromek, E.; Kassassir, H.; Kur, J.; Turkiewicz, M. Genetic and biochemical characterization of yeasts isolated from Antarctic soil samples. *Polar Biol.* **2017**, *40*, 1787–1803. [CrossRef]
44. De Jonge, P.; De Jongh, F.C.M.; Meijers, R.; Steensma, H.Y.; Scheffers, W.A. Orthogonal-field-alternation gel electrophoresis banding patterns of DNA from yeasts. *Yeast* **1986**, *2*, 193–204. [CrossRef] [PubMed]
45. Cahan, P.; Kennell, J.C. Identification and distribution of sequences having similarity to mitochondrial plasmids in mitochondrial genomes of filamentous fungi. *Mol. Genet. Genom.* **2005**, *273*, 462–473. [CrossRef] [PubMed]
46. Wang, Y.; Zeng, F.; Hon, C.C.; Zhang, Y.; Leung, F.C.C. The mitochondrial genome of the Basidiomycete fungus *Pleurotus ostreatus* (oyster mushroom). *FEMS Microbiol. Lett.* **2008**, *280*, 34–41. [CrossRef]
47. Libkind, D.; Čadež, N.; Opulente, D.A.; Langdon, Q.K.; Rosa, C.A.; Sampaio, J.P.; Gonçalves, P.; Hittinger, C.T.; Lachance, M.A. Towards yeast taxogenomics: Lessons from novel species descriptions based on complete genome sequences. *FEMS Yeast Res.* **2020**, *20*, foaa042. [CrossRef] [PubMed]
48. Lallemand, T.; Leduc, M.; Landès, C.; Rizzon, C.; Lerat, E. An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. *Genes* **2020**, *11*, 1046. [CrossRef]
49. Dean, E.J.; Davis, J.C.; Davis, R.W.; Petrov, D.A. Pervasive and persistent redundancy among duplicated genes in yeast. *PLOS Genet.* **2008**, *4*, e1000113. [CrossRef]
50. Grainger, R.J.; Barrass, J.D.; Jacquier, A.; Rain, J.-C.; Beggs, J.D. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA* **2009**, *15*, 2161–2173. [CrossRef]

51. Golubev, W. *Rhodosporidium babjevae*, a new heterothallic yeast species (Ustilaginales). *Syst. Appl. Microbiol.* **1993**, *16*, 445–449. [CrossRef]
52. Chand Dakal, T.; Giudici, P.; Solieri, L. Contrasting patterns of rDNA homogenization within the *Zygosaccharomyces rouxii* species complex. *PLoS ONE* **2016**, *11*, e0160744. [CrossRef] [PubMed]
53. Conti, A.; Corte, L.; Pierantoni, D.C.; Robert, V.; Cardinali, G. What is the best lens? Comparing the resolution power of genome-derived markers and standard barcodes. *Microorganisms* **2021**, *9*, 299. [CrossRef]
54. Brandenburg, J.; Blomqvist, J.; Shapaval, V.; Kohler, A.; Sampels, S. Oleaginous yeasts respond differently to carbon sources present in lignocellulose hydrolysate. *Biotechnol. Biofuels* **2021**, *14*, 124. [CrossRef]
55. Wang, Q.; Sun, M.; Zhang, Y.; Song, Z.; Zhang, S.; Zhang, Q.; Xu, J.R.; Liu, H. Extensive chromosomal rearrangements and rapid evolution of novel effector superfamilies contribute to host adaptation and speciation in the basal ascomycetous fungi. *Mol. Plant Pathol.* **2020**, *21*, 330–348. [CrossRef] [PubMed]
56. Passoth, V.; Hansen, M.; Klinner, U.; Emeis, C.C. The electrophoretic banding pattern of the chromosomes of *Pichia stipitis* and *Candida shehatae*. *Curr. Genet.* **1992**, *22*, 429–431. [CrossRef] [PubMed]
57. Legrand, M.; Jaitly, P.; Feri, A.; D'Enfert, C.; Sanyal, K. *Candida albicans*: An emerging yeast model to study eukaryotic genome plasticity. *Trends Genet.* **2019**, *35*, 292–307. [CrossRef]
58. Narayanan, A.; Vadnala, R.N.; Ganguly, P.; Selvakumar, P.; Rudramurthy, S.M.; Prasad, R.; Chakrabarti, A.; Siddharthan, R.; Sanyal, K. Functional and comparative analysis of centromeres reveals clade-specific genome rearrangements in *Candida auris* and a chromosome number change in related species. *MBio* **2021**, *12*, e00905-21. [CrossRef]
59. Hagen, F.; Khayhan, K.; Theelen, B.; Kolecka, A.; Polacheck, I.; Sionov, E.; Falk, R.; Parnmen, S.; Lumbsch, H.T.; Boekhout, T. Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. *Fungal Genet. Biol.* **2015**, *78*, 16–48. [CrossRef]
60. Gordon, J.L.; Byrne, K.P.; Wolfe, K.H. Mechanisms of chromosome number evolution in yeast. *PLoS Genet.* **2011**, *7*, e1002190. [CrossRef]
61. Hellborg, L.; Piškur, J. Complex nature of the genome in a wine spoilage yeast, *Dekkera bruxellensis*. *Eukaryot. Cell* **2009**, *8*, 1739–1749. [CrossRef]
62. Chang, S.L.; Lai, H.Y.; Tung, S.Y.; Leu, J.Y. Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet.* **2013**, *9*, e1003232. [CrossRef] [PubMed]
63. Vassiliadis, D.; Wong, K.H.; Blinco, J.; Dumsday, G.; Andrianopoulos, A.; Monahan, B. Adaptation to industrial stressors through genomic and transcriptional plasticity in a bioethanol producing fission yeast isolate. *G3 Genes Genomes Genet.* **2020**, *10*, 1375–1391. [CrossRef] [PubMed]
64. Pierce, N.T.; Irber, L.; Reiter, T.; Brooks, P.; Brown, C.T. Large-scale sequence comparisons with sourmash. *F1000 Res.* **2019**, *8*, 1006. [CrossRef] [PubMed]
65. Civiero, E.; Pintus, M.; Ruggeri, C.; Tamburini, E.; Sollai, F.; Sanjust, E.; Zucca, P. Physiological and phylogenetic characterization of *Rhodotorula diobovata* DSBCA06, a nitrophilous yeast. *Biology* **2018**, *7*, 39. [CrossRef]

III

Biotechnology for Biofuels
and Bioproducts

RESEARCH

Enhanced glycerol assimilation and lipid production in *Rhodotorula toruloides* CBS14 upon addition of hemicellulose primarily correlates with early transcription of energy-metabolism-related genes

Giselle C. Martín-Hernández[1†], Mikołaj Chmielarz[1†], Bettina Müller[1], Christian Brandt[2], Adrian Viehweger[3], Martin Hölzer[4] and Volkmar Passoth[1*]

Abstract

Background Lipid formation from glycerol was previously found to be activated in *Rhodotorula toruloides* when the yeast was cultivated in a mixture of crude glycerol (CG) and hemicellulose hydrolysate (CGHH) compared to CG as the only carbon source. RNA samples from *R. toruloides* CBS14 cell cultures grown on either CG or CGHH were collected at different timepoints of cultivation, and a differential gene expression analysis was performed between cells grown at a similar physiological situation.

Results We observed enhanced transcription of genes involved in oxidative phosphorylation and enzymes localized in mitochondria in CGHH compared to CG. Genes involved in protein turnover, including those encoding ribosomal proteins, translation elongation factors, and genes involved in building the proteasome also showed an enhanced transcription in CGHH compared to CG. At 10 h cultivation, another group of activated genes in CGHH was involved in β-oxidation, handling oxidative stress and degradation of xylose and aromatic compounds. Potential bypasses of the standard *GUT1* and *GUT2*-glycerol assimilation pathway were also expressed and upregulated in CGHH 10 h. When the additional carbon sources from HH were completely consumed, at CGHH 36 h, their transcription decreased and NAD^+-dependent glycerol-3-phosphate dehydrogenase was upregulated compared to CG 60 h, generating NADH instead of NADPH with glycerol catabolism. *TPI1* was upregulated in CGHH compared to cells grown on CG in all physiological situations, potentially channeling the DHAP formed through glycerol catabolism into glycolysis. The highest number of upregulated genes encoding glycolytic enzymes was found after 36 h in CGHH, when all additional carbon sources were already consumed.

Conclusions We suspect that the physiological reason for the accelerated glycerol assimilation and faster lipid production, was primarily the activation of enzymes that provide energy.

Keywords *Rhodotorula toruloides*, Transcriptomics, Lignocellulose, Glycerol, Biofuels

†Giselle C. Martín-Hernández and Mikołaj Chmielarz have contributed equally to this work

*Correspondence:
Volkmar Passoth
Volkmar.passoth@slu.se
Full list of author information is available at the end of the article

Martín-Hernández *et al. Biotechnology for Biofuels and Bioproducts* (2023) 16:42

Page 2 of 16

Introduction

One of the world's fastest-growing food commodities is vegetable oils (VO) [1]. VO are also used as the main feedstock for biodiesel production, a renewable energy source and alternative to fossil fuels [2]. VO consumption in Sweden and the European Union (EU) is much higher than production. Thus, a significant proportion of the VO used for biodiesel production is imported [3]. Furthermore, a high greenhouse gas potential is associated with producing many VO, such as palm-, soybean- and peanut oil, whose emissions are reported to exceed 2000 kg CO_2 equivalents per ton produced [4, 5]. Substantial rainforest clearing due to land use is also associated with producing these VO [4, 6].

Microbial oils have the potential to replace VO in the production of biodiesel and feed and food additives [7, 8]. Oleaginous yeasts are known to be among the fastest lipid producers on earth [9–11]. Among these, *Rhodotorula* species, which are basidiomycetes oleaginous yeasts, can convert a variety of industrial low-value residues, such as lignocellulose hydrolysates, including hemicellulose from pulp-and-paper industry or crude glycerol, a residue from biodiesel production, into lipids and carotenoids [11–17]. In a recent study, we could show that lipid production rate was increased in *Rhodotorula* species when cultured on crude glycerol and hemicellulose hydrolysate (CGHH), compared to crude glycerol (CG) as the sole carbon source [13]. The physiological basis of this phenomenon remains to be elucidated.

The increasing availability of *R. toruloides* genome assemblies has enabled differential gene expression studies, which both contribute to the understanding of carbon metabolism and allow the construction of metabolic models at the genome level for this species [18–25]. Recently, we assembled the genome of *R. toruloides* CBS14 at the chromosomal level, with the highest number of annotated transcripts published so far for *R. toruloides* [26]. This study aimed to compare gene transcription of *R. toruloides* CBS14 cultivated in either CGHH or CG, to get an insight into the metabolic pathways that are activated due to the presence of HH and enhance glycerol turnover.

Materials and methods

Bioreactor cultivation and sampling

Cultivations were performed as described in [13]. Briefly, bioreactors (Multifors, Infors HT, Bottmingen, Switzerland) with 500 mL working volume were used for growing *R. toruloides* CBS14 containing either CGHH (50% CG, 10% HH) or only CG (50% CG) as carbon sources as well as 0.75 g/L yeast extract (BactoTM Yeast Extract, BD, France), 1 g/L $MgCl_2$ $6xH_2O$ (Merck KGaA, Germany), 2 g/L $(NH_4)_2HPO_4$ ($\geq$ 98%, Sigma-Aldrich, USA) and 1.7 g/L YNB without amino acids and ammonium sulphate (DifcoTM, Becton Dickinson, France). To inoculate the bioreactors, cells from a -80 °C stock culture were streaked on YPD-agar (glucose 20 g/L, yeast extract 10 g/L, peptone 20 g/L) and incubated for 3 days. From the plates, cells were inoculated into 100 mL YPD in 500 mL baffled Erlenmeyer flasks. After incubation on an orbital shaker at 125 rpm and 25 °C for 72 h, the cells were transferred to 50 mL-Falcon tubes, washed twice with NaCl-solution (9 g/L) and re-suspended with NaCl-solution. Cultures were inoculated with the cell suspension, to reach an initial OD of 1 in the cultivation. Cultivations were performed in triplicate at 25 °C, pH 6, and an oxygen tension of 21%.

CG was obtained from Perstorp Holding AB, Sweden. The glycerol concentration was specified as 80%; however, there were differences from batch to batch. For the cultivations described here, the same batch was used as in the bioreactor experiments in [13]. HH was generated from wheat straw, after steam explosion and enzymatic digestion. Pretreatment was performed at Lund University, Sweden, as described previously [16]. Briefly, the straw was soaked with 1% acetic acid overnight, and then steam exploded at 190 °C. The liquid fraction, representing the solubilised hemicellulose, was removed from the solid fraction by pressing and used in the experiments. HH contained 26.2 g/L xylose, 7 g/L glucose, 6.6 g/L acetic acid and small amounts of arabinose (<0.5 g/L). The nitrogen content was 0.6 g/l [27]. The pH was set to 6 by addition of appropriate amounts of 5 M NaOH [13]. The C/N-ratios were 32.5 for CGHH and 30.7 for CG.

Samples for RNA-isolation and determination of the concentrations of biomass, carbon sources and lipids were isolated from *R. toruloides* CBS14 cultures grown at different growth conditions: they were taken from CG cultures after 10 h, 30 h, and 60 h and from CGHH cultures after 10 h, 36 h, and 60 h. Cell dry weight was determined gravimetrically, and glucose, xylose, acetic acid and arabinose were determined by HPLC [17]. Lipid content was measured using FT-NIR, as described previously [28]. Cell samples for RNA isolation (3 mL) were rapidly collected in Falcon tubes and placed in ice to decrease sample temperature.

RNA extraction and sequencing

Total RNA was immediately extracted in triplicates from each sample following a previously described protocol [26]. rRNA-depleted samples were sequenced on the Illumina MiSeq system using the reagent kit v3, and paired-end RNA sequencing reads were obtained (2×75 bp). The base-calling pipeline included MiSeq Control Software v2.6.2.1, RTA v1.18.54 and bcl2fastq v2.20.0.422.

RNA-Seq data analysis

The RNAflow differential gene expression pipeline v1.1.0 [29] was used for transcriptome analysis. This includes read quality control, count normalization, reference-based mapping, gene quantification, differential expression, and visualization. Samples were quality-checked with FastQC, and raw reads were trimmed with fastp to remove low-value bases and adapter contamination [30]. After quality-trimming, the transcriptomics data had an average coverage of 1.6 M (0.8–3.8 M) reads per sample, with an expected read depth of 28X. For read depth estimation, the whole transcriptome size was calculated to be 8,484,192 bp, based on 5958 average number of transcribed genes (TPM > 0) per sample, and an average transcript length of 1424 bp. The remaining rRNA was detected and removed using SortMeRNA [31]. Then, reads were splice-aware aligned with HISAT2 [32] to the *R. toruloides* assembly as previously reported [26]. FeatureCounts was used for gene-level expression quantification [33] only considering uniquely mapped reads and using the annotation file from [26]. To identify and remove low-expressed genes, the TPM values (transcripts per million kilobases) were determined using RNAflow [29]. The TPM value was determined for each sample and each gene. Furthermore, the mean TPM value per condition was calculated over all biological replicates. A gene had to have a particular mean TPM value above a defined threshold to be considered in the subsequent analyses. We performed all calculations using a TPM 5 as the default threshold. Finally, differential gene expression analysis was performed using DESeq2 [34] to identify significantly (adjusted p value < 0.05) differentially expressed genes. We adjusted the p values attained by the Wald test in DESeq2 [34] for multiple testing using the Benjamini and Hochberg method, implemented as a default in DESeq2's results() function. As recommended, we used these adjusted p values to identify genes with significantly different expressions. Corresponding R packages were used to conduct principal component analysis and generate expression heatmaps and box plots. Details of the tool versions, R packages, and custom scripts used can be found at https://github.com/hoelzer-lab/rnaflow.

Visualization of gene expression density in the genome of *R. toruloides* CBS14 was performed using Circa (http://omgenomics.com/circa; accessed on May 2022). The respective KEGG orthology (KO) numbers were assigned to those annotated proteins encoded by differentially expressed genes. Subsequently, metabolic pathways and cellular processes were determined using KofamKOALA [35]. Gene ontology (GO) terms from differentially expressed genes that occurred at least 10 times were visualized using REVIGO [36] in semantic similarity-based scatterplots. Blast homology search (v 2.13.0+)

was performed to identify genes and proteins belonging to central metabolic pathways annotated with a similar function in CBS14 [37, 38].

Results

RNA sampling points were selected according to the dynamics of glycerol consumption in CG or CGHH

To identify genes and metabolic pathways active during glycerol turnover in *R. toruloides* CBS14, cells were cultivated under different growth conditions. Differential gene expression analysis was performed by bulk RNA-sequencing (RNA-seq) at different timepoints as explained below: as mentioned above, *R. toruloides* CBS14 showed different growth rates in cultivation media using either CG or CGHH as the main carbon source [13]. Faster growth, faster initial glycerol consumption, and more rapid lipid formation were observed in CGHH compared to CG [13]. Thus, sampling timing was selected based on the observed dynamics of glycerol consumption. RNA isolation was done in three independent cultivations for each culture medium (sampling points are illustrated in Fig. 1). The first sampling was performed after 10 h to allow the cultures to adapt to the cultivation conditions. In CGHH, the consumption of glycerol was visible after 10 h. However, a physiologically comparable situation was reached in the CG culture after about 30 h, so a further sample was taken from the CG culture at this time. Another sampling point was chosen after 36 h in CGHH and 60 h in CG. In CGHH, about 20 g/L of glycerol was left at this timepoint and the additional carbon sources from HH were consumed entirely. This culture condition was thus similar to the CG-culture after 60 h, where about 20 g/L was also left. In the CGHH culture, another sample was taken after 60 h. At this point, glycerol was still present, but only half as much after 36 h. Thus, the expression profile of this sample may reflect physiological responses to different glycerol concentrations.

Global gene expression patterns differ clearly between timepoints and carbon sources

Prior to differential expression analysis with DESeq2, the quality of the transcriptome reads was checked. Passed reads were mapped and quantified using the annotated genome of *R. toruloides* CBS14 [26]. The number of TPM was calculated per gene and sample. The expression levels of each condition were thus normalized against gene length and sequencing depth. Weakly expressed genes were filtered out. The density of highly expressed genes within contigs and scaffolds from CBS14 genome assembly is shown in Fig. 2. An expression level of at least 105 TPM is evenly distributed throughout the genome, except for contigs 49 (length 62 kbp) and 64
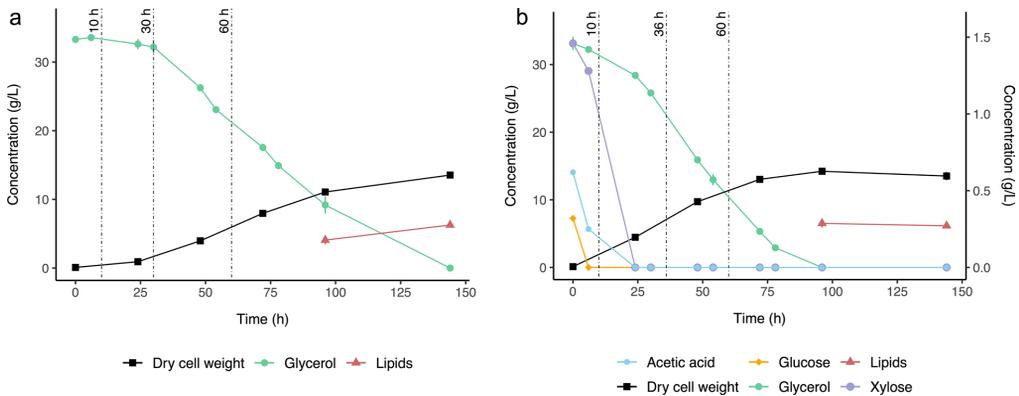
Fig. 1 Bioreactor cultivation of *Rhodotorula toruloides* CBS14 using as carbon source: **a** crude glycerol (CG), or **b** mixture of CG and hemicellulose hydrolysate (CGHH). The reconstruction of the cultivation curves was performed using results from [13]. Independent cultures, different from those whose values are the basis of the shown growth curves, were performed for RNA-isolation. At each sampling, concentrations of biomass, lipids and substrate were determined, to ensure the cells were in the same growth phase as indicated in the figure. Glycerol concentrations are given on the primary y axis. The secondary y axis indicates the concentrations of xylose, glucose and acetic acid in the hemicellulose hydrolysate. Vertical dashed lines represent sampling points for RNA extraction

(length 151 kbp), from which no transcripts were recovered. Differences in the expression profile can be spotted between different timepoints and media. In addition, gene expression density and transcription levels in the mitochondrial genome were much higher than in the rest of the genome (results not shown). We conducted PCA to analyze differences in the clustering of biological replicates and global gene expression patterns between the samples (Additional file 1: Figs. S1–S4). The sampling time (reflected by PC1) explains 86.6% of the variance, and 8% variance is explained by medium composition (reflected by PC2) (Additional file 1: Fig. S1). The genes with annotated function which contributed most to the differences between conditions were in decreasing order: RHOT147219 (encoding NADH-ubiquinone oxidoreductase chain 1), RHOT147222 (cytochrome c oxidase subunit 1), RHOT142646 (sulfated surface glycoprotein 185), RHOT149100 (putative protein TPRXL) and RHOT149239 (elongation of fatty acids protein 3).

Transcription levels of genes were first compared pairwise between the sampling points of the same growth condition (CG or CGHH). More specifically, we compared each of the two later sampling points with the first after 10 h of growth and assigned the identified differentially expressed genes to the KEGG metabolic pathways and cellular processes (Additional file 1: Fig. S5). The accounted genes summarized in Additional file 1: Fig. S5 showed a significant up- or downregulation ($p < 0.05$) with a $\log_2$Fold change > 1.5 or < − 1.5, which is also in line with the high variance in gene expression shown by

the principal component 1 (PC1 86.6%, see Additional file 1: Fig. S1). In both CGHH and CG, more genes per KEGG pathway and process were higher transcribed at the 10 h samplings compared to later samplings. In CGHH, glucose was exhausted at this timepoint (Fig. 1b), and thus, the significantly higher gene expression at 10 h compared to later timepoints is probably related to the transition to a broader spectrum of metabolic activities to assimilate other carbon sources [39]. In comparison, the number of differentially expressed genes in CG with no additional carbon sources remained close to zero between 10 and 30 h of growth for most pathways and processes. This is also indicated by the low expression variance between these samples (Additional file 1: Fig. S1).

Changes in transcript abundance were further evaluated to identify differentially expressed genes (adjusted p value < 0.05) between *R. toruloides* cell cultures grown on different carbon sources at a similar physiological situation, as illustrated in Volcano plots (Fig. 3; Additional file 1: Figs. S2–S4, and Additional file 1: Tables S1–S3). They correspond to samples after adapting to cultivation conditions in each medium (CGHH 10 h vs. CG 10 h), when glycerol consumption became visible (CGHH 10 h vs. CG 30 h), and when there is about 20 g/l of glycerol left, and the additional carbon sources from HH were completely consumed (CGHH 36 h vs. CG 60 h). GO term enrichment analysis revealed that these differentially expressed genes are involved in the biological processes illustrated in Additional file 1: Fig. S6.
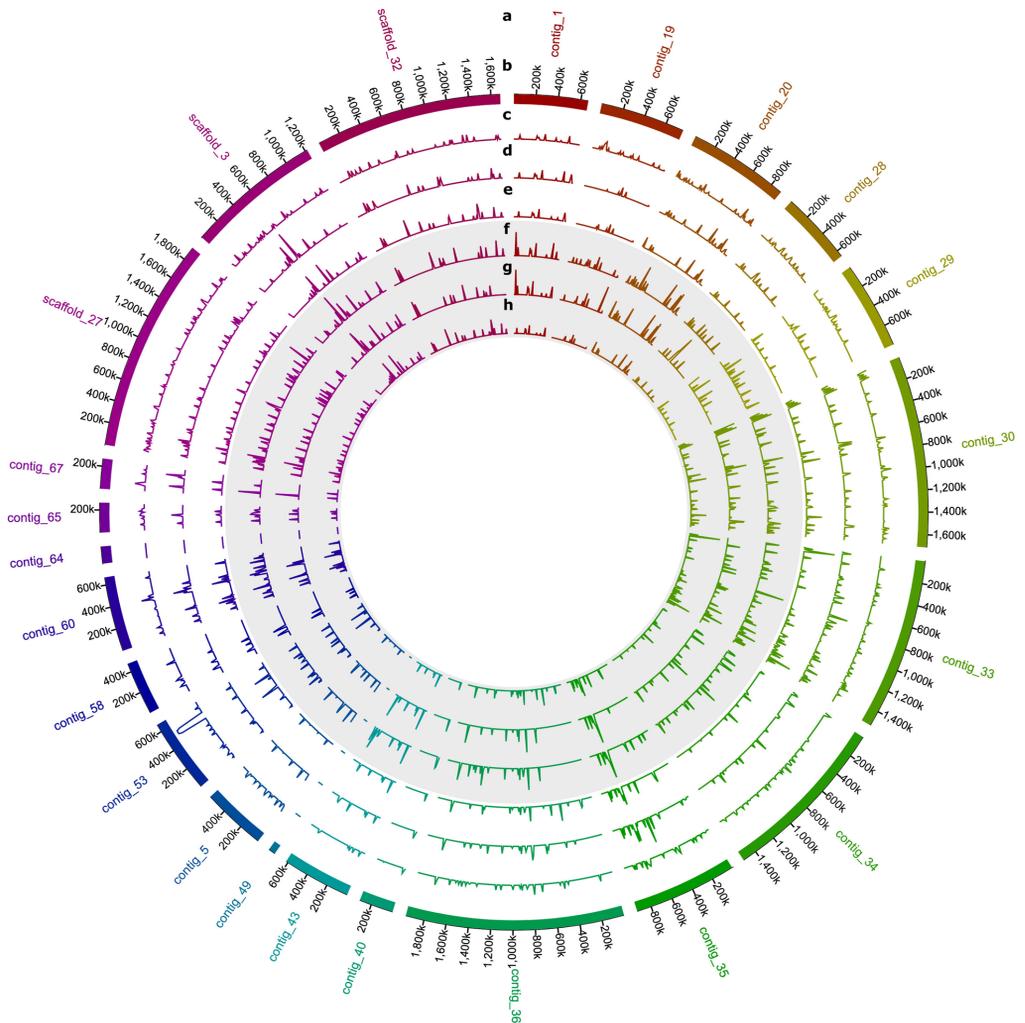
Fig. 2 Map of genes expressed over the mean level of TPM in *Rhodotorula toruloides* CBS14. Gene densities when grown on each of the two different carbon sources are indicated in concentric circles. From outside to inside: **a** *R. toruloides* CBS14 contig names; **b** sizes; and in 10 kb windows, density of genes expressed over the mean TPM level in CBS14 when grown in CGHH as main carbon source at **c** 10 h **d** 36 h and **e** 60 h; and in CG at **f** 10 h, **g** 30 h and **h** 60 h. The circles representing samples from cells grown in CG are also indicated in gray. Only nuclear encoded genes are included in this graph. CG, crude glycerol; CGHH, mixture of CG and hemicellulose hydrolysate; TPM, transcripts per million kilobases

Increased protein turnover and energy metabolism in CGHH after 10 h of cultivation

Of 634 differentially expressed genes, 396 were significantly higher transcribed ($p < 0.05$, log_2Fold change < 0) in CGHH 10 h than in CG 10 h (Fig. 3a; Additional file 1: Table S1). Many of these genes were generally higher expressed also when compared to the other sampling points, both in CGHH and CG. Genes encoding enzymes involved in metabolic pathways were most differentially expressed within the assigned KEGG orthologs. An exceptionally high proportion of these genes are involved in amino acid metabolism and were upregulated mainly in cells grown in CGHH as carbon source (Fig. 4a). While genes involved in signal transduction had the highest
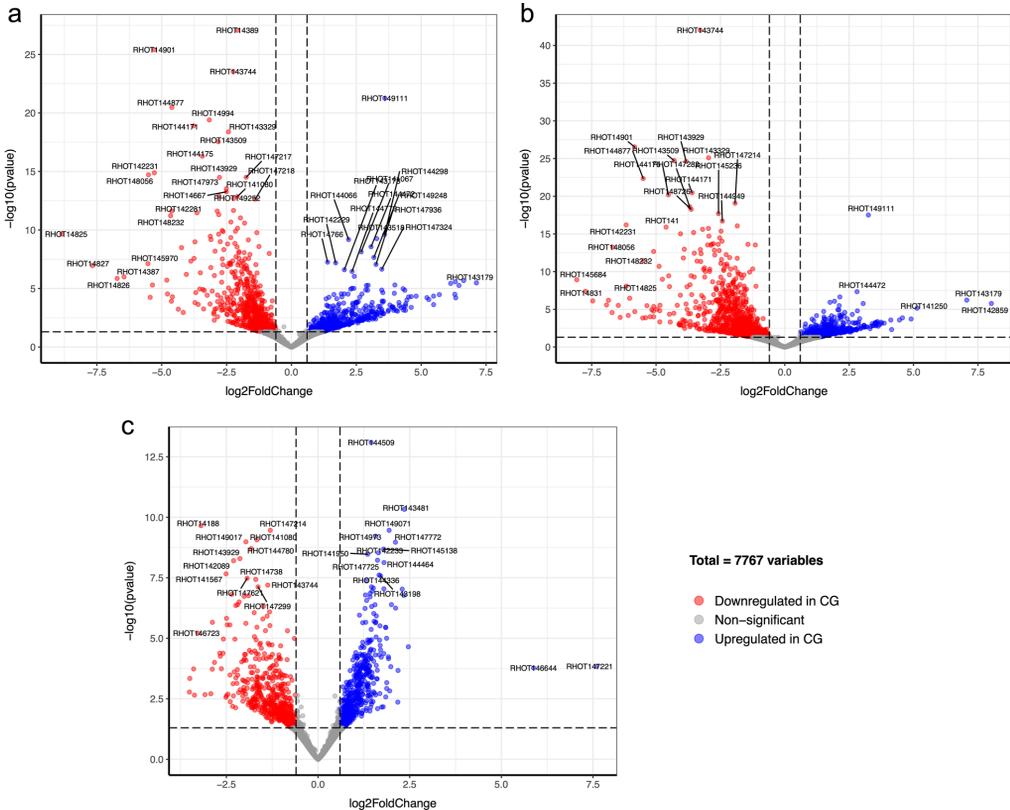
Fig. 3 Top differentially expressed genes when comparing the cultures at timepoints with similar physiological situations (for detailed explanation, see section "RNA sampling points were selected according to the dynamics of glycerol consumption in CG or CGHH"): **a** CGHH 10 h versus CG 10 h; **b** CGHH 10 h versus CG 30 h; **c** CGHH 36 h versus CG 60 h. Upregulated genes (adjusted *p* value < 0.05 and $\log_2$Fold change > 0.6) in CG are indicated in blue while downregulated genes (adjusted *p* value < 0.05 and $\log_2$Fold change < − 0.6) in red. Genes that are not significantly differentially expressed are in gray. The higher in the y axis, the more significant, and the further to the left, the more downregulated in CG. For example, the gene encoding Peroxisomal multifunctional enzyme type 2 (RHOT14901) is significantly downregulated in CG 10 h and CG 30 h compared to CGHH 10 h. CG, crude glycerol; CGHH, mixture of CG and hemicellulose hydrolysate

number of upregulated genes among cellular processes in CG 10 h, genes involved in translation was highest in CGHH 10 h (Fig. 4a).

27 ribosomal protein genes were upregulated in CGHH 10 h, including both cytoplasmic and mitochondrial ribosomes (Additional file 1: Fig. S7b). Genes involved in ribosome biogenesis and spliceosome formation, as well as translation initiation factors and components of all three DNA-dependent RNA-polymerases were also higher expressed in CGHH. Apart from genes involved in protein synthesis, protein degradation also appeared activated in CGHH 10 h. Transcription of 13 proteasome-related genes was upregulated in CGHH (Additional

file 1: Fig. S7c), while none were downregulated. Compared to all other measuring points, in both CGHH and CG, the TPM values of most of these proteasome-related genes were about 2–3-fold higher in CGHH 10 h.

Besides gene expression, protein synthesis, and protein degradation through proteasome, a high proportion of the upregulated genes in CGHH 10 h compared to CG 10 h were associated with energy metabolism (Fig. 4a). This includes especially genes encoding proteins for oxidative phosphorylation and other mitochondrial enzymes. 21 genes involved in the mitochondrial electron transport chain-complex I (NADH ubiquinone oxidoreductase), III (cytochrome c reductase), IV (cytochrome
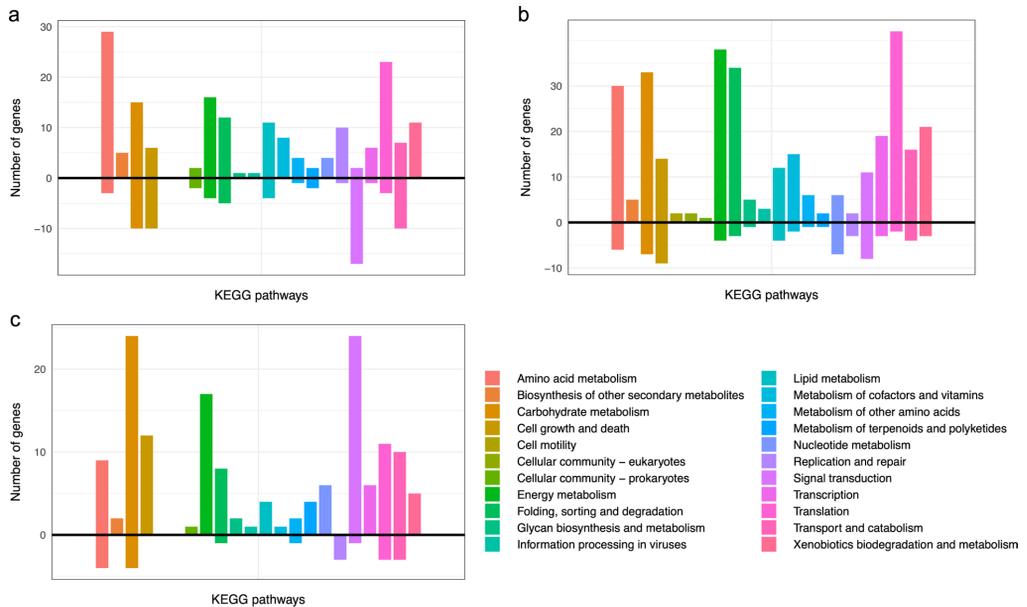
Fig. 4 Number of differentially expressed genes in CG compared to CGHH per KEGG pathway and cellular process: **a** 10 h in CGHH versus 10 h in CG, **b** 10 h in CGHH versus 30 h in CG, and **c** 36 h in CGHH versus 60 h in CG. The accounted genes showed significant ($p < 0.05$) upregulation (+y axis) or downregulation (−y axis) in CGHH with an absolute log$_2$Fold Change higher than 1.5. Bar colors indicate different KEGG pathways. CG, crude glycerol; CGHH, mixture of CG and hemicellulose hydrolysate

c oxidase), and F-type ATPase were significantly regulated ($p < 0.05$) in CGHH 10 h compared to CG 10 h. They included 20 genes that were upregulated in CGHH 10 h and one that was downregulated (Additional file 1: Fig. S7a). Two prohibitin genes (RHOT148333 and RHOT147096), involved in the formation of respiratory supercomplexes [40], were regulated similarly. A variety of genes encoding mitochondrial enzymes or accessory components followed the same pattern: they were upregulated in CGHH 10 h compared to CG 10 h, had the highest TPM values in CGHH 10 h, and were at 36 h lower, but still at a similar level as in CG 10 h (Fig. 5g and Additional file 1: Table S4). This includes genes encoding enzymes involved in the tricarboxylic acid (TCA) cycle, such as NADP$^+$-specific isocitrate dehydrogenase (ICDH), succinate-CoA ligase, and fumarate hydratase (RHOT145845, -7009 and -5604, respectively), or in the synthesis of cofactors of mitochondrial enzymes, such as riboflavin (RHOT149252, 2556 and 2045), lipoic acid (by lipoyl synthase, RHOT145711) and thiamine pyrophosphate (TPP, by thiamine pyrophosphokinase, RHOT149040). Lipoic acid and TPP are cofactors of pyruvate dehydrogenase and α-keto glutarate dehydrogenase [41].

On the other hand, transcription of most genes involved in glycolytic reactions and regulation were not significantly different between CG 10 h and CGHH 10 h, when this medium still had xylose and acetic acid as additional carbon sources (Fig. 1b). The exceptions were the genes encoding a phosphofructokinase (*PFK2*, RHOT143173), which was downregulated in CGHH 10 h, phosphoglyceromutase (*GPM1*, RHOT146772) and triosephosphate isomerase (*TPI1*, RHOT141080), which were upregulated in CGHH 10 h. Other glycolytic genes were expressed constitutively. Phosphoglyceromutase (PGM) has been identified as the only glycolytic enzyme significantly differentially expressed in *R. toruloides* cells depending on the carbon source, glucose, or xylose [42]. It was reported for a recombinant *S. cerevisiae* strain that the overexpression of a *PGM*-gene enhanced xylose fermentation [43]. Furthermore, *PFK2* was also downregulated on YP (yeast/peptone) medium supplemented with either xylose or glucose when compared to the expression on YP only [22]. In the same study, *TPI1* was overexpressed when YP medium was supplemented with acetate [22].

Some of the suggested genes that are involved in xylose assimilation in *R. toruloides* are NAD(P)H-dependent
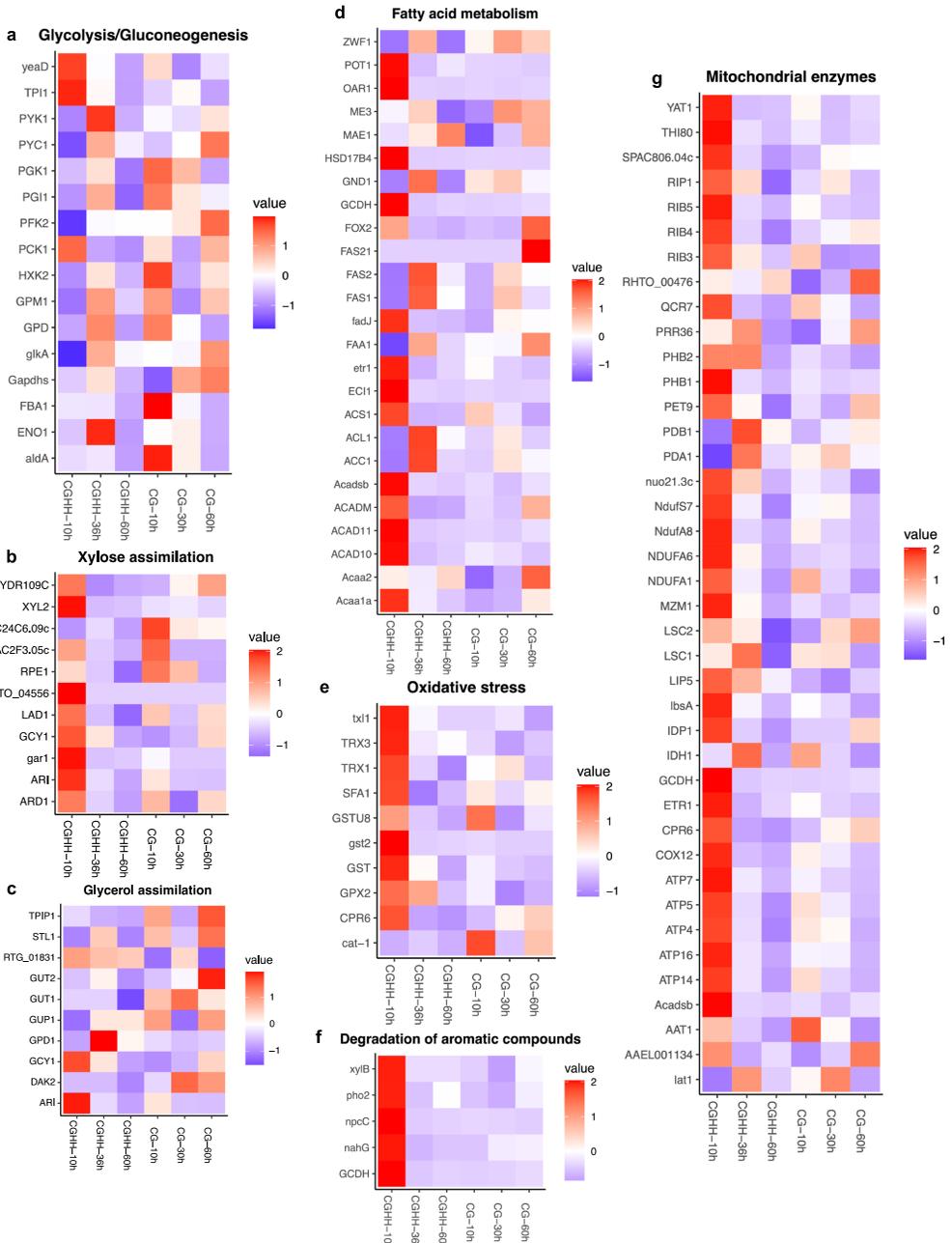
Fig. 5 Gene expression in central metabolic pathways. Scaled TPM values from genes encoding enzymes involved in **a** glycolysis and gluconeogenesis, **b** xylose assimilation, **c** glycerol assimilation, **d** fatty acid metabolism and NADPH generation, **e** handling oxidative stress, **f** degradation of aromatic compounds and **g** mitochondrial enzymes involved in respiration. Color key stands for the z-scores obtained for each gene (normalized for all cultivation conditions, separately performed for each of the metabolic pathways)

xylose reductase (XR), NADH-dependent xylitol dehydrogenase (XDH), xylulose kinase (XK), and phosphoketolase [42]. By performing homology analyses, we identified the gene RHOT147125 (annotated as *ARI* in *R. toruloides* CBS14, homolog to Rhto_03963 in *R. toruloides* NP11 [25]) as XR. Its transcription was more than doubled in CGHH (Fig. 5b and Additional file 1: Table S4), although significant upregulation ($p < 0.05$) was only found compared to CG 30 h and to the later CGHH sample points. In both media, RHOT147125 expression decreased with time. Besides XR, Tiukova et al. [42] identified three aldo/keto reductases that were potentially involved in xylose assimilation. They are annotated in CBS14 as galacturonate reductase (RHOT148868), glycerol 2-dehydrogenase (NADP$^+$) (RHOT14370), and uncharacterized oxidoreductase C2F3.05c (RHOT143299). RHOT148868 and RHOT14370 were upregulated ($p < 0.05$) in CGHH 10 h compared to CG 10 h. The gene encoding XDH (RHOT149007) was significantly higher transcribed ($p < 0.05$) at CGHH 10 h and strongly decreased between 10 and 36 h. (Fig. 5b; Additional file 1: Table S4). In CG, XDH was equally low expressed at all sample points. According to homology analysis encoded by RHOT149455, XK was poorly expressed (Additional file 1: Table S4) at all sample points in both media without any significant differential expression, as observed in a previous study [22]. Contradictory, a phosphoketolase gene (RHOT147705) was significantly higher transcribed ($p < 0.05$) in CG 10 h than in CGHH 10 h. Tiukova et al. [42] proposed that with excess carbon, the reaction catalyzed by phosphoketolase might be unnecessary, resulting in most xylulose-5-P entering the pentose phosphate pathway (PPP). However, RHOT146580, encoding ribulose-phosphate 3-epimerase, was also expressed but with no significant differences between media and sample points. Jagtap et al. [22, 44] suggested that *R. toruloides* might use an alternative route for xylose utilization, in which xylulose is converted to D-arabitol rather than to xylulose-5-phosphate by the activity of XK. An enzyme catalyzing this alternative reaction is D-arabinitol 4-dehydrogenase. The genome harbors a gene coding for arabinitol 4-dehydrogenase (*LAD1*, RHOT144154). It was transcribed highest in CGHH 10 h, and the expression decreased with time (Fig. 5b; Additional file 1: Table S4). The differences were also significant between CGHH 10 h and all the timepoints in CG. However, whether it can act on D- and L-arabinitol remains unclear. *ARD1* (RHOT146692), which encodes D-arabinitol 2-dehydrogenase, was also higher transcribed in CGHH 10 h compared to CG 10 h but without statistical significance. Its expression decreased with time in CGHH. *ARD1* could be involved in the formation of D-ribulose from D-arabinitol. These results agree with the proposition from Jagtap et al. [22, 44], though Ribulokinase (RHOT145356) was expressed at an even lower level and not significantly different.

Five genes that may be important in the degradation of aromatic compounds were clearly expressed in CGHH 10 h, while there was only weak expression in CG (Fig. 5f; Additional file 1: Table S4). Aromatic monomers could originate from the lignin and thus reside in the hemicellulose portion of CGHH.

The highest number of upregulated genes involved in lipid metabolism was found at 10 h of cultivation, with higher levels in CGHH (Fig. 4a; Additional file 1: Fig. S5). The gene encoding acetyl-CoA synthetase (ACS, RHOT148257) was upregulated in CGHH 10 h compared to CG 10 h. RHOT148257 was also upregulated in CGHH 10 h compared to all other measuring points, in both CGHH and CG. This acetate-converting enzyme is part of the acetate assimilation pathway and was likely upregulated, since cells at this point were consuming the acetic acid present in the HH (Fig. 1b) [13, 45]. At later timepoints, when acetate was also no longer detected in the medium, the transcription of this gene was downregulated compared to 10 h. Interestingly, its expression in CG 10 h was relatively high, about half that in CGHH 10 h, even though no acetate was present in the cultivation medium. Here, too, the gene was downregulated with increasing cultivation time. Acetate could originate as a secondary metabolite from other metabolic pathways associated with glycerol assimilation. ATP-dependent citrate lyase (ACL, RHOT147175), thought to be the main producer of acetyl-CoA in FA synthesis [46, 47], was downregulated in CGHH 10 h compared to CG 10 h. High levels of cytoplasmic acetyl CoA produced by ACS could affect the expression level of ACL. A previous study also showed reduced expression of *ACL1* in *R. toruloides* grown on YP supplemented with acetate compared to when supplemented with glucose [22]. At later timepoints, higher transcription levels were found in CGHH (Fig. 5d; Additional file 1: Table S4). The expression of acetyl CoA carboxylase (RHOT148968) showed no significant differences between media or timepoints. The FA synthase genes *RtFAS1* (RHOT148939) and *RtFAS2* (RHOT146383) were transcribed at low levels in both substrates, particularly in CGHH, upon 10 h of cultivation, but without significant differences between the media. In contrast, two genes involved in fatty acid (FA) biosynthesis, 3-ketoacyl-acyl carrier protein reductase (FabG, RHOT148056) and enoyl carrier protein reductase (RHOT148822), were significantly upregulated in CGHH 10 h compared to CG 10 h. Transcription of these genes declined at later timepoints in CGHH down to levels comparable to those in CG (Fig. 5d; Additional file 1: Table S4). A diglyceride acyltransferase encoding

gene (RHOT149017), involved in triacylglycerol biosynthesis, was also upregulated in CGHH 10 h compared to CG 10 h, and its expression significantly decreased with cultivation time in CGHH. Upregulation of this enzyme on acetate-containing medium has previously been observed elsewhere [22]. RHOT147182, which codes for a putative acyl-CoA desaturase, had very low TPM values in CGHH 10 h compared to the other conditions and was downregulated in CGHH compared to CG 10 h. In addition, ten genes involved in FA degradation, about half of the genes being mitochondrial and the other half being peroxisomal, were significantly upregulated in CGHH 10 h compared to CG 10 h. FA accumulation is considered higher at later timepoints when there is nitrogen or phosphate limitation but a surplus of carbon [21, 27]. FA degradation at earlier growth stages could be related to an increase in released FA through autophagy processes triggered by glucose depletion [48]. Mitochondrial $NADP^+$-specific ICDH was the only enzyme coding gene involved in NADPH-generation whose transcription differed significantly between media at 10 h.

The catabolic L-glycerol 3-phosphate (G3P) pathway, involving glycerol kinase (*GUT1*) and FAD-dependent glycerol-3-phosphate dehydrogenase (*GUT2*), is used by *Saccharomyces cerevisiae* as the main assimilation pathway for glycerol as demonstrated by deletion studies targeting *GUT1* and *GUT2* [49, 50]. Another proposed pathway in yeast is the catabolic dihydroxyacetone (DHA) pathway. It is performed by glycerol dehydrogenase (GDH) and DHA kinase (DAK) [51, 52]. A third pathway, termed the catabolic glyceraldehyde (GA) pathway, has been proposed for *Neurospora crassa*. Here, the glycerol is first oxidized by an $NADP^+$-dependent GDH to GA, which is then either phosphorylated by a GA kinase to GA-3-P or reduced by an aldehyde dehydrogenase to D-glycerate. A glycerate 3-kinase then converts the D-glycerate to 3-P-D-glycerate [51, 53, 54].

At 10 h of cultivation, transcripts of two putative glycerol transporters (*STL1*, RHOT147915, and *GPU1*, RHOT144353) were more abundant in CG than in CGHH. Enzymes belonging to the catabolic G3P pathway were transcribed under all conditions without significant differences. Enzymes belonging to the catabolic DHA pathway were also expressed, indicating the presence and expression of alternative pathways of glycerol assimilation in *R. toruloides*. The genome harbors two $NADP^+$-dependent glycerol dehydrogenase genes (*GCY1*-homologs), RHOT14370 and RHOT144361, that convert glycerol to DHA. They were highest expressed in CGHH 10 h. Transcription levels decreased with time, except for RHOT14370 in CG, where the level increased. This enzyme was previously described as involved mostly in glycerol anabolic reactions [51]. The genome also

encodes a DHA kinase 2 homolog, alternative name glycerone kinase 2 (*DAK2*, RHOT142321), which phosphorylates both DHA and GA, indicating that it may also be involved in the GA pathway in addition to the catabolic DHA pathway. *DAK2* expression decreased in the 60 h samples in both media compared to the earlier timepoints. However, there were no significant differences between the conditions and further investigations are required to confirm this tendency. The genome harbors an Alcohol dehydrogenase [NADP(+)] gene (*ARI*, RHOT147125), whose encoding protein was found to have 54% sequence identity to $NADP^+$-dependent GDH from *Trichoderma reesei* (ABD83952.1), besides 100% identity to XR from *R. toruloides* NP11. *ARI* could have mediated the conversion of glycerol to GA, which represents the first step of the catabolic GA pathway, in addition to its role in xylose metabolism [51]. *ARI* was transcribed under all conditions, and the transcription levels decreased with time (Fig. 5c). It was higher transcribed in CGHH 10 h than in CG 10 h. However, the significance of these differences could not be proven, thus, further investigations are required to confirm this tendency. A variety of aldehyde dehydrogenases were expressed under both experimental conditions; however, differences in their expression could not be shown. We identified RHOT146637 as D-glycerate 3-kinase (*RTG_01831*) by performing homology analyses. It was transcribed under all conditions but without significant differences. Summarized, this suggests that *R. toruloides* can utilize all three glycerol assimilation pathways described in fungi [51]. The expression level of enzymes belonging to the DHA catabolic pathway could account for the differences in glycerol assimilation between cells grown in CGHH and CG after 10 h.

Several genes involved in handling oxidative stress were upregulated in CGHH at 10 h, including three thioredoxin genes (RHOT143685, 7078, and 3176). However, some stress-related genes were also upregulated in CG, including a catalase gene (RHOT141031) (Fig. 5e; Additional file 1: Table S4).

Central metabolic pathways that were differentially regulated at the 10 h sampling point are represented in Fig. 6a.

The number of differentially expressed genes was highest when glycerol consumption became visible

Of the 787 differentially expressed genes in CGHH 10 h, 585 were upregulated in CG 30 h (Fig. 3b; Additional file 1: Table S2). Central metabolic pathways that were differentially regulated are shown in Fig. 6b. The pathways with a higher number of differentially expressed genes were energy, carbohydrate, and amino acid metabolism, in descending order (Fig. 4b). These mainly were
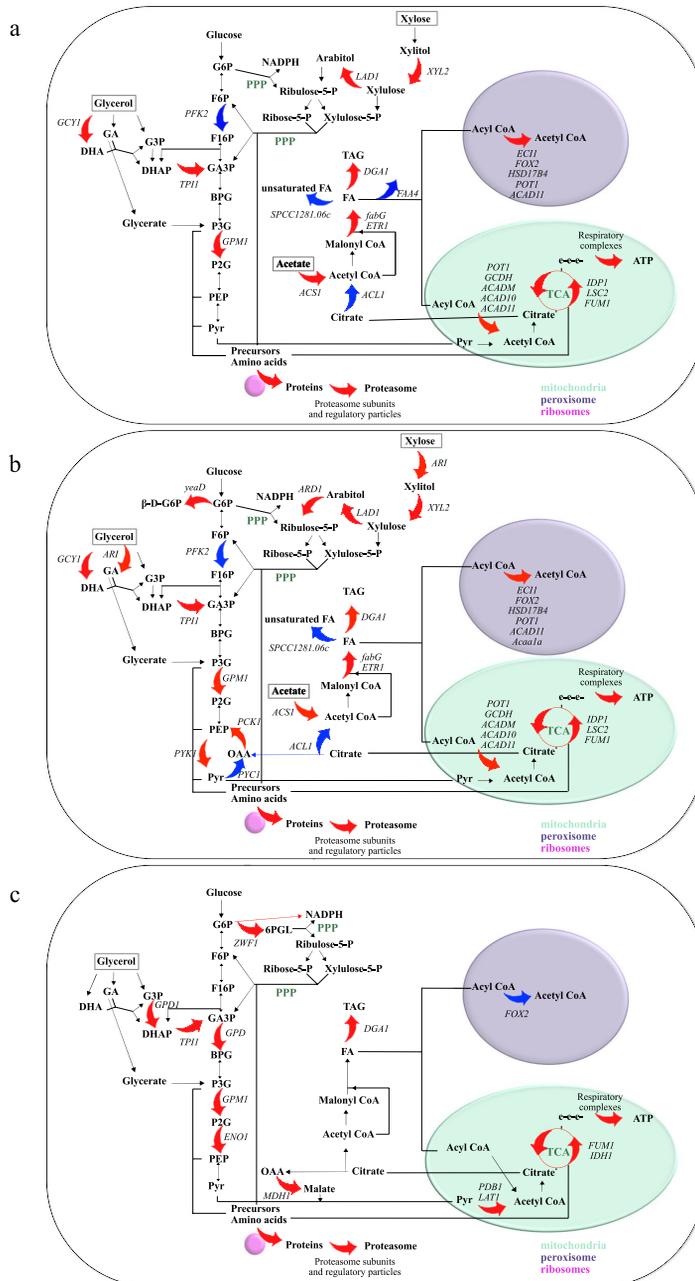
Fig. 6 Differentially regulated central metabolic pathways in *R. toruloides* CBS14. Genes expressed in cells grown on a mixture of crude glycerol (CG) and hemicellulose hydrolysate (CGHH) were compared to cells grown in CG as only carbon source: **a** 10 h versus 10 h **b** 10 versus 30 h and **c** 36 h versus 60 h, respectively. Red arrows indicate upregulated genes (adjusted *p* value < 0.05) in CGHH, while blue arrows indicate downregulated genes (adjusted *p* value < 0.05). Genes that are expressed without significant differences are represented with black arrows

upregulated in CGHH 10 h, similar as when compared to the earlier timepoint in CG.

27 genes associated with oxidative phosphorylation were upregulated in CGHH 10 h (Additional file 1: Fig. S8a) compared to CG 30 h. In addition, the upregulated genes at 10 h cultivations which encoded mitochondrial enzymes or components, such as prohibitin, NADP$^+$-specific ICDH, succinate-CoA ligase, fumarate hydratase, riboflavin synthase, lipoyl synthase, and thiamine pyrophosphokinase, were still upregulated in CGHH 10 h when compared to CG 30 h.

Differences in the transcription of glycolytic genes included the downregulation of *PFK2* and upregulation of *GPM1* and *TPI1* in CGHH, similarly as described at 10 h comparison. A gene encoding pyruvate kinase (RHOT144157) was upregulated in CGHH. Phosphoenolpyruvate carboxykinase gene (*PCK1*, RHOT144519), which can be linked with gluconeogenic reactions, was also upregulated in CGHH, while pyruvate carboxylase gene (*PYC1*, RHOT149202) was downregulated. *PCK1* was previously found to be upregulated during growth of *R. toruloides* on acetate [22]. Another upregulated gene involved in carbohydrate metabolic processes encodes a putative glucose-6-phosphate 1-epimerase (RHOT149194). Of the described differentially expressed genes involved in xylose assimilation in *R. toruloides*, *ARI*, the aldo/keto reductases RHOT148868 and RHOT14370, *XYL2*, *LAD1,* and *ARD1* were upregulated in CGHH (Fig. 6b), which suggests an increased flux via arabitol.

The processes of folding, sorting and degradation, and translation also had a high number of upregulated genes in CGHH 10 h compared to CG 30 h (Fig. 4b). Transcription of 46 ribosomal proteins (Additional file 1: Fig. S8b), five RNA polymerases, and 20 genes associated with the spliceosome was upregulated in CGHH. There were no downregulated genes whose functions are involved in transcription. In addition, the transcription of genes involved in protein degradation was activated in CGHH. Transcription of 19 genes involved in proteasome assembly was upregulated (Additional file 1: Fig. S8c), while none was downregulated in CGHH.

The highest number of upregulated genes involved in lipid metabolism was also found on the carbon source CGHH at these timepoints (Fig. 4b). In a similar pattern to that described in the comparison of the 10 h sample points, transcription of the genes encoding for ACS, FabG, enoyl carrier protein reductase, and diglyceride acyltransferase was upregulated in CGHH. ACL and probable acyl-CoA desaturase were downregulated, and there were no significant differences between media for *ACC1*, *RtFAS1*, and *RtFAS2*. Twelve genes involved in FA degradation were significantly upregulated in CGHH,

about six of which were peroxisomal. Transcription of genes encoding enzymes involved in cytosolic NADPH-generation did not differ significantly between media when glycerol consumption became visible.

The genes involved in glycerol assimilation which were differentially expressed when comparing CGHH 10 h with CG30 h were two *GCY1* genes and *ARI*. They might be associated with the catabolic pathways via DHA and GA, respectively.

Several genes related to oxidative stress management were also upregulated in CGHH at 10 h compared to CG 30 h, including the three thioredoxin genes described above and a glutathione-S transferase (RHOT149349).

The depletion of additional carbon sources induces changes in glycerol utilization and carbohydrate pathways

311 of 632 differentially expressed genes were higher transcribed in CGHH 36 h than in CG 60 h (Fig. 3c; Additional file 1: Table S3). Central metabolic pathways that were differentially expressed are shown in Fig. 6c. The cellular process with the highest number of upregulated genes was signal transduction in CGHH (Fig. 4c). Carbohydrate metabolism was the pathway with a higher number of differentially expressed genes, with a more prominent expression level on the CGHH carbon source.

Transcription of seven genes associated with Glycolysis and Gluconeogenesis was upregulated in CGHH, while none was downregulated. Besides *TPI1* and *GPM1,* the other upregulated genes were an aldehyde dehydrogenase (RHOT148569), glyceraldehyde 3-phosphate dehydrogenase (RHOT147990), pyruvate dehydrogenase E1 component subunit beta (RHOT14206), pyruvate dehydrogenase complex component E2 (RHOT146289), and enolase (RHOT142969). Another upregulated gene involved in carbohydrate metabolic processes was that for glucose-6-phosphate 1-dehydrogenase (RHOT146681), which provides NADPH and pentose phosphates for the synthesis of FA and nucleic acids.

There were few significant differences in gene expression related to lipid metabolism between CGHH 36 h and CG 60 h. The exceptions were diglyceride acyltransferase, involved in triacylglycerol biosynthesis, and the peroxisomal multifunctional beta-oxidation protein (RHOT144031), involved in FA degradation. They were upregulated in CGHH and CG, respectively. Interestingly, *RtFAS21* (RHOT146384), which forms an antisense RNA [26], was expressed in CG 60 h.

25 genes associated with oxidative phosphorylation were upregulated in CGHH 36 h (Additional file 1: Fig. S9a) compared to CG 60 h. Other genes encoding mitochondrial enzymes or associated components still upregulated in CGHH when glycerol was the sole carbon source were prohibitin-2, NAD$^+$-specific ICDH

(RHOT14435), fumarate hydratase, DHBP synthase, and lipoyl synthase. A cytoplasmic malate dehydrogenase (RHOT147988) was also upregulated in CGHH.

There were two RNA polymerase genes (DNA-directed RNA polymerase I subunit rpa1 and DNA-directed RNA polymerase II subunit rpb1) that were downregulated and none upregulated in cultivations with CGHH as carbon source. However, the transcription of 22 ribosomal proteins (Additional file 1: Fig. S9b) was upregulated. In addition, the transcription of genes involved in protein degradation was activated in CGHH, including seven genes involved in proteasome assembly (Additional file 1: Fig. S9c).

At these timepoints, *GCY1* and *ARI* transcription, possibly involved in glycerol and xylose metabolism, were no longer significantly different between cultivations. However, expression of an NAD^+-dependent glycerol-3-phosphate dehydrogenase (RHOT141674) was upregulated in CGHH, generating NADH along with glycerol catabolism. NADH could then be transported to the mitochondria by the malate dehydrogenase shuttle or *GUT2*.

Discussion

A considerable transcriptional upregulation of genes involved in oxidative phosphorylation and of several mitochondrial enzymes was observed in CGHH 10 h cultures compared to CG 10 h and CG 30 h (Fig. 4a, b). For some genes, this effect was still visible even after longer cultivation, when the additional carbon sources originating from the HH were already consumed, and the remaining glycerol concentrations were similar, as especially shown for CGHH 36 h compared to CG 60 h (Figs. 4, 5, 6). Energy metabolism could thus have been activated in response to the HH addition. High availability of ATP is of central importance for forming biomass and lipids, as well as for efficient uptake of limiting resources, such as nitrogen and for inhibitor tolerance [55].

Among the mitochondrial enzymes, lower transcription of the $NADP^+$-specific ICDH and succinate-CoA ligase genes was observed in CGHH 36 h compared to CG 60 h, although significance could not be demonstrated. Further studies are required to confirm this tendency. Nevertheless, this would be consistent with the specific increase in lipid production observed during the growth on glycerol supplemented with HH [13]: the inhibition of ICDH is a central regulatory element in lipid accumulation [56] in oleaginous yeasts and succinate-CoA ligase catalyzes the subsequent reaction in the TCA-cycle. On the other hand, we found an upregulated NAD^+-specific ICDH in CGHH 36 h compared to CG 60 h. Which cofactor, NAD^+ or $NADP^+$ is used by the ICDH may have different physiological consequences for the cell [57]. During the early growth phase in CGHH,

the generated NADPH might provide biosynthetic reducing power needed to promote cell growth and stress responses. In the later growth phase, around 36 h, the ICDH may have primarily provided an energy source in the form of NADH.

Transcription of genes involved in protein turnover was largely upregulated in the CGHH culture, especially at the beginning of growth (Fig. 4a, b). Probably, this might be due to the temporary presence of various carbon sources and other compounds present in HH, which require establishment and later inactivation of metabolic pathways.

Total glycerol consumption took about 48 h less in the culture with HH than in the cultures without HH (Fig. 1). No upregulation of genes involved in glycerol assimilation via the catabolic G3P pathway was observed in CGHH when sampled after 10 h. A potential bypass of the standard *GUT1* and *GUT2* pathway, as described by Klein et al. [51], was expressed, similar to the DHA pathway demonstrated in *S. pombe* [52]. We found two upregulated GDH genes in CGHH that were annotated as $NADP^+$-dependent. These were initially upregulated in CGHH-cultivated cells (Fig. 6). They might represent additional pathways to direct glycerol into main metabolic pathways while providing NADPH to biosynthetic pathways, including FA synthesis. Moreover, genes belonging to the catabolic GA pathway proposed in *N. crassa* [51, 53, 54] were also expressed. To our knowledge, this is the first indication that the catabolic GA pathway for glycerol assimilation was identified in a basidiomycete and a yeast. However, further investigations are required to confirm the existence of this pathway in *R. toruloides*.

In addition, *TPI1* was upregulated in CGHH compared to cells grown in CG in the same physiological situation (i.e., CGHH 10 h and CG 10 h, CGHH 10 h and CG 30 h, and CGHH 36 h and CG 60 h), possibly directing the DHAP formed by glycerol catabolism to glycolysis. At 10 h cultivations, the expression level of genes encoding glycolytic enzymes was similar under both growth conditions, i.e., with or without the addition of HH, except for *PFK2* and *GPM1*. This might be due to the fact that glucose was rapidly depleted in CGHH. After 36 h in CGHH, the number of upregulated glycolytic genes increased compared to CG 60 h.

When the cells grew in CGHH, they simultaneously assimilated acetic acid and glucose, which supplied the cell with additional amounts of acetyl-CoA for FA synthesis [58, 59]. ACS was significantly more highly transcribed after 10 h in CGHH compared to CGHH 60 h and CG 60 h. This points to a higher acetyl-CoA production required for the slightly enhanced lipid synthesis observed in CGHH [13]. Both acetate assimilation and alternative glycerol degradation pathway may provide

the additional acetyl-CoA and NADPH, respectively, that would have been required for the enhanced lipid synthesis seen in CGHH [13].

However, cells harvested after 10 h in CGHH showed higher transcription of genes involved in lipid degradation and lower transcription of genes involved in FA production compared to later timepoints in CGHH and CG. Activation of genes involved in β-oxidation has previously been observed upon cultivation on xylose and under nitrogen-limiting and lipid-accumulating conditions [42]. In the closely related species *R. glutinis*, we even observed a breakdown of the cell's intracellular lipids with simultaneous xylose assimilation [27].

During xylose assimilation, *XYL2* and *LAD1* were upregulated in CGHH 10 h compared to CG 10 h and CG 30 h, suggesting an increased flux via arabitol. These expression differences disappeared when xylose was exhausted after 36 h in CGHH. It has been shown that the growth on xylose appears to correlate with the induction of several genes involved in oxidative stress response [42]. Likewise, in this study, xylose appears to trigger stress responses, particularly in CGHH, after 10 h of cultivation, although the reason remains unclear. However, the presence of acetic acid in our cultivation might also have contributed to the activation of the oxidative stress response. The addition of acetic acid led to the production of reactive oxygen species in *S. cerevisiae* [60].

Activation of energy-yielding enzymes, meaning those enzymes involved in oxidative phosphorylation and mitochondrial function, appears to be the primary physiological cause of accelerated glycerol assimilation and lipid production.

In addition to providing additional carbon sources, the substrate itself could also have led to the induction of respiratory genes as a response to stress (for example, due to the presence of aromatic compounds in the HH). In this context, we have observed the activation of genes involved in the degradation of aromatic compounds. Aromatic compounds can be metabolized by *Rhodotorula* yeasts [61] as a source of carbon, but they are also toxic at the same time. Due to the pre-treatment of the lignocellulose, soluble aromatic compounds may have been present in small amounts in the HH fraction [62] and thus have been involved in stress response. In the presence of HH, the energy demand of the cells could be higher, which could explain the activation of the β-oxidation in CGHH 10 h compared to CG at 10 and 30 h (Fig. 6). Once the substrate has been detoxified, the cells switch to glycerol metabolism and lipid accumulation suggested by the upregulation of genes ($p < 0.05$) involved in these pathways in CGHH (Fig. 6). These transcriptional changes could have been enabled by the cells having a sufficiently high ATP level and, therefore, sufficient energy, since a high number of respiratory genes were upregulated ($p < 0.05$) in CGHH compared to CG in the three similar physiological situations. This could be further supported by the HH-induced increased protein synthesis and turnover, which enabled the cells to adapt efficiently to the changing carbon source.

Conclusions

There is no direct proportionality between transcription level and enzyme activity. The transcriptome is only one part of a cell's information chain, and such changes do not necessarily equate to metabolic pathways and enzyme activities [63, 64]. Nevertheless, the observation of the transcription of a large number of genes does allow some valuable conclusions to be drawn about metabolic and signaling pathways. In this case, it provides hypotheses about the molecular mechanisms triggered by a small amount of HH added to CG medium, which lead to the faster initial glycerol consumption and higher lipid accumulation.

In this study, we observed enhanced transcription of genes involved in oxidative phosphorylation and enzymes localised in mitochondria in CGHH compared to CG. Genes involved in protein turnover, including those encoding ribosomal proteins, translation elongation factors and genes involved in building the proteasome also showed an enhanced transcription in CGHH. In the presence of the HH, β-oxidation was activated. We suspect that the physiological reason for the accelerated glycerol assimilation and faster lipid production, was primarily the activation of enzymes involved in providing energy to the cells.

Our study paves the way for further detailed investigations of such underlying mechanisms. It also helps to identify new targets to obtain strains that can more rapidly accumulate lipids from residual, low-value substrates.

Supplementary Information

Additional file 1: Table S1. Differentially expressed genes between growth on different media at 10 h. **Table S2.** Differentially expressed genes between growth on different media when glycerol consumption became visible. **Table S3.** Differentially expressed genes between growth on different media when glycerol concentration is lower. **Table S4.** Gene expression in TPM within central metabolic pathways. **Figure S1.** Principal Component Analysis based on expression level from the 500 highest expressed genes in each of the cultivation media and RNA sampling points. **Figure S2.** Overview of differential expression analysis between Rhodotorula toruloides CBS 14 cultivated on crude glycerol and a mixture of crude glycerol and hemicellulose hydrolysate, at 10 h cultivation. **Figure S3.** Overview of differential expression analysis between Rhodotorula toruloides CBS 14 grown on different media when glycerol consumption became visible. **Figure S4.** Overview of differential expression analysis between Rhodotorula toruloides CBS 14 grown on different media when

about 20 g/l of glycerol were left. **Figure S5.** Differentially expressed genes in Rhodotorula toruloides CBS 14 within sampling time points in each of the growth media. **Figure S6.** Examples of upregulated genes in Rhodotorula toruloides CBS 14 grown on a mixture of crude glycerol (CG) and hemicellulose hydrolysate (CGHH) for 10 h of cultivation, whose expression is higher than CG 10 h. **Figure S7.** Examples of upregulated genes in Rhodotorula toruloides CBS 14 grown on a mixture of crude glycerol (CG) and hemicellulose hydrolysate (CGHH) for 10 h of cultivation, whose expression is higher than CG 30 h. **Figure S8.** Examples of upregulated genes in Rhodotorula toruloides CBS 14 grown on a mixture of crude glycerol (CG) and hemicellulose hydrolysate (CGHH) for 36 h of cultivation, whose expression is higher than CG 60 h. **Figure S9.** Distribution of shared ortholog clusters within differentially expressed genes in each of the three differential expression analyzes. The diagram was generated using OrthoVenn2.

Availability of data and materials
The genome assembly and annotation of *R. toruloides* CBS14, as well as BAM mapping files and transcript counts, are available at https://osf.io/yzqn5. Raw RNA reads are available in ENA under the project PRJEB40807.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
A.V., C.B., and M.H. are co-founders of nanozoo GmbH and hold shares in the company. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Author details
[1]Department of Molecular Sciences, BioCenter, Swedish University of Agricultural Sciences, Box 7015, 75007 Uppsala, Sweden. [2]Institute for Infectious Diseases and Infection Control, Jena University Hospital, Jena, Germany. [3]Institute of Medical Microbiology and Virology, University Hospital Leipzig, 04103 Leipzig, Germany. [4]Method Development and Research Infrastructure, Bioinformatics and Systems Biology, Robert Koch Institute, 13353 Berlin, Germany.

References
1. Khatri P, Jain S. Environmental life cycle assessment of edible oils: a review of current knowledge and future research challenges. J Clean Prod. 2017;152:63–76.
2. Kumar A, Vachan Tirkey J, Kumar Shukla S. Comparative energy and economic analysis of different vegetable oil plants for biodiesel production in India. Renew Energy. 2021;169:266–82.
3. Harnesk D, Brogaard S. Social dynamics of renewable energy—how the European union's renewable energy directive triggers land pressure in Tanzania. J Environ Dev. 2016;26(2):156–85.
4. Schmidt JH. Life cycle assessment of five vegetable oils. J Clean Prod. 2015;87:130–8.
5. Uusitalo V, Väisänen S, Havukainen J, Havukainen M, Soukka R, Luoranen M. Carbon footprint of renewable diesel from palm oil, jatropha oil and rapeseed oil. Renew Energy. 2014;69:103–13.
6. Hoang NT, Kanemoto K. Mapping the deforestation footprint of nations reveals growing threat to tropical forests. Nat Ecol Evol. 2021;5(6):845–53.
7. Ghazani SM, Marangoni AG. Microbial lipids for foods. Trends Food Sci Technol. 2022;119:593–607.
8. Carmona-Cabello M, García IL, Papadaki A, Tsouko E, Koutinas A, Dorado MP. Biodiesel production using microbial lipids derived from food waste discarded by catering services. Bioresour Technol. 2021;323: 124597.
9. Sawangkeaw R, Ngamprasertsith S. A review of lipid-based biomasses as feedstocks for biofuels production. Renew Sustain Energy Rev. 2013;25:97–108.
10. Abeln F, Chuck CJ. The history, state of the art and future prospects for oleaginous yeast research. Microb Cell Fact. 2021;20(1):221.
11. Passoth V, Brandenburg J, Chmielarz M, Martín-Hernández GC, Nagaraj Y, Müller B, Blomqvist J. Oleaginous yeasts for biochemicals, biofuels and food from lignocellulose-hydrolysate and crude glycerol. Yeast. 2023. https://doi.org/10.1002/yea.3838.
12. Papanikolaou S, Aggelis G: Microbial products from wastes and residues. FEMS Microbiol Lett. 2020. https://doi.org/10.1093/femsle/fnaa156.
13. Chmielarz M, Blomqvist J, Sampels S, Sandgren M, Passoth V. Microbial lipid production from crude glycerol and hemicellulosic hydrolysate with oleaginous yeasts. Biotechnol Biofuels. 2021;14(1):65.
14. Nagaraj YN, Burkina V, Okmane L, Blomqvist J, Rapoport A, Sandgren M, Pickova J, Sampels S, Passoth V. Identification, quantification and kinetic study of carotenoids and lipids in *Rhodotorula toruloides* CBS 14 cultivated on wheat straw hydrolysate. Fermentation. 2022;8(7):300.
15. Balan V. Microbial lipid production, 1 edn. Humana, New York, NY; 2019.
16. Blomqvist J, Pickova J, Tilami SK, Sampels S, Mikkelsen N, Brandenburg J, Sandgren M, Passoth V. Oleaginous yeast as a component in fish feed. Sci Rep. 2018;8(1):15945.
17. Brandenburg J, Poppele I, Blomqvist J, Puke M, Pickova J, Sandgren M, Rapoport A, Vedernikovs N, Passoth V. Bioethanol and lipid production from the enzymatic hydrolysate of wheat straw after furfural extraction. Appl Microbiol Biotechnol. 2018;102(14):6269–77.
18. Dinh HV, Suthers PF, Chan SHJ, Shen Y, Xiao T, Deewan A, Jagtap SS, Zhao H, Rao CV, Rabinowitz JD, et al. A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. Metab Eng Commun. 2019;9: e00101.
19. Coradetti ST, Pinel D, Geiselman GM, Ito M, Mondo SJ, Reilly MC, Cheng Y-F, Bauer S, Grigoriev IV, Gladden JM et al. Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. eLife. 2018;7:e32110
20. Tiukova IA, Prigent S, Nielsen J, Sandgren M, Kerkhoven EJ. Genome-scale model of *Rhodotorula toruloides* metabolism. Biotechnol Bioeng. 2019;116(12):3396–408.
21. Pinheiro MJ, Bonturi N, Belouah I, Miranda EA, Lahtvee P-J: Xylose metabolism and the effect of oxidative stress on lipid and carotenoid production in *Rhodotorula toruloides*: insights for Future Biorefinery. Front Bioeng Biotechnol. 2020. https://doi.org/10.3389/fbioe.2020.01008.
22. Jagtap SS, Deewan A, Liu JJ, Walukiewicz HE, Yun EJ, Jin YS, Rao CV. Integrating transcriptomic and metabolomic analysis of the oleaginous yeast *Rhodosporidium toruloides* IFO0880 during growth under different carbon sources. Appl Microbiol Biotechnol. 2021;105(19):7411–25.
23. Kim J, Coradetti ST, Kim YM, Gao Y, Yaegashi J, Zucker JD, Munoz N, Zink EM, Burnum-Johnson KE, Baker SE, et al. Multi-omics driven metabolic

network reconstruction and analysis of lignocellulosic carbon utilization in *Rhodosporidium toruloides*. Front Bioeng Biotechnol. 2021;8: 612832.

24. Touchette D, Altshuler I, Gostincar C, Zalar P, Raymond-Bouchard I, Zajc J, McKay CP, Gunde-Cimerman N, Whyte LG. Novel Antarctic yeast adapts to cold by switching energy metabolism and increasing small RNA synthesis. ISME J. 2022;16(1):221–32.

25. Zhu Z, Zhang S, Liu H, Shen H, Lin X, Yang F, Zhou YJ, Jin G, Ye M, Zou H, et al. A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. Nat Commun. 2012;3(1):1112.

26. Martín-Hernández GC, Müller B, Chmielarz M, Brandt C, Hölzer M, Viehweger A, Passoth V. Chromosome-level genome assembly and transcriptome- based annotation of the oleaginous yeast *Rhodotorula toruloides* CBS 14. Genomics. 2021;113(6):4022–7.

27. Brandenburg J, Blomqvist J, Shapaval V, Kohler A, Sampels S, Sandgren M, Passoth V. Oleaginous yeasts respond differently to carbon sources present in lignocellulose hydrolysate. Biotechnol Biofuels. 2021;14(1):124.

28. Chmielarz M, Sampels S, Blomqvist J, Brandenburg J, Wende F, Sandgren M, Passoth V. FT-NIR: a tool for rapid intracellular lipid quantification in oleaginous yeasts. Biotechnol Biofuels. 2019;12:169.

29. Lataretu H, Holzer M. RNAflow: an effective and simple RNA-Seq differential gene expression pipeline using Nextflow. Genes (Basel). 2020;11(12):1487.

30. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.

31. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211–7.

32. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60.

33. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.

34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

35. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2019;36(7):2251–2.

36. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE. 2011;6(7): e21800.

37. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000;7(1–2):203–14.

38. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–402.

39. Pomraning KR, Collett JR, Kim J, Panisko EA, Culley DE, Dai Z, Deng S, Hofstad BA, Butcher MG, Magnuson JK. Transcriptomic analysis of the oleaginous yeast *Lipomyces starkeyi* during lipid accumulation on enzymatically treated corn stover hydrolysate. Biotechnol Biofuels. 2019;12(1):162.

40. Azuma K, Ikeda K, Inoue S. Functional mechanisms of mitochondrial respiratory chain supercomplex assembly factors and their involvement in muscle quality. Int J Mol Sci. 2020;21(9):3182.

41. Mentel M, Chovančíková P, Zeman I, Polčic P. Learning from yeast about mitochondrial carriers. Microorganisms. 2021;9(10):2044.

42. Tiukova IA, Brandenburg J, Blomqvist J, Sampels S, Mikkelsen N, Skaugen M, Arntzen MO, Nielsen J, Sandgren M, Kerkhoven EJ. Proteome analysis of xylose metabolism in *Rhodotorula toruloides* during lipid production. Biotechnol Biofuels. 2019;12:137.

43. Garcia Sanchez R, Hahn-Hagerdal B, Gorwa-Grauslund MF. Cross-reactions between engineered xylose and galactose pathways in recombinant *Saccharomyces cerevisiae*. Biotechnol Biofuels. 2010;3:19.

44. Jagtap SS, Rao CV. Production of d-arabitol from d-xylose by the oleaginous yeast *Rhodosporidium toruloides* IFO0880. Appl Microbiol Biotechnol. 2018;102(1):143–51.

45. Brandenburg J, Blomqvist J, Pickova J, Bonturi N, Sandgren M, Passoth V. Lipid production from hemicellulose with *Lipomyces starkeyi* in a pH regulated fed-batch cultivation. Yeast. 2016;33(8):451–62.

46. Athenaki M, Gardeli C, Diamantopoulou P, Tchakouteu SS, Sarris D, Philippoussis A, Papanikolaou S. Lipids from yeasts and fungi: physiology, production and analytical considerations. J Appl Microbiol. 2018;124(2):336–67.

47. Sato R, Ara S, Yamazaki H, Ishiya K, Aburatani S, Takaku H. Citrate-mediated Acyl-CoA synthesis is required for the promotion of growth and triacylglycerol production in oleaginous yeast *Lipomyces starkeyi*. Microorganisms. 2021;9(8):1693.

48. Singh R, Kaushik S, Wang Y, Xiang Y, Novak I, Komatsu M, Tanaka K, Cuervo AM, Czaja MJ. Autophagy regulates lipid metabolism. Nature. 2009;458(7242):1131–5.

49. Sprague GF, Cronan JE. Isolation and characterization of *Saccharomyces cerevisiae* mutants defective in glycerol catabolism. J Bacteriol. 1977;129(3):1335–42.

50. Swinnen S, Klein M, Carrillo M, McInnes J, Nguyen HTT, Nevoigt E. Re-evaluation of glycerol utilization in *Saccharomyces cerevisiae*: characterization of an isolate that grows on glycerol without supporting supplements. Biotechnol Biofuels. 2013;6(1):157.

51. Klein M, Swinnen S, Thevelein JM, Nevoigt E. Glycerol metabolism and transport in yeast and fungi: established knowledge and ambiguities. Environ Microbiol. 2017;19(3):878–93.

52. Matsuzawa T, Ohashi T, Hosomi A, Tanaka N, Tohda H, Takegawa K. The gld1+ gene encoding glycerol dehydrogenase is required for glycerol metabolism in *Schizosaccharomyces pombe*. Appl Microbiol Biotechnol. 2010;87(2):715–27.

53. Viswanath-Reddy M, Bennett SN, Howe HB Jr. Characterization of glycerol nonutilizing and protoperithecial mutants of *Neurospora*. Mol Gen Genet. 1977;153(1):29–38.

54. Tom GD, Viswanath-Reddy M, Howe HB Jr. Effect of carbon source on enzymes involved in glycerol metabolism in *Neurospora crassa*. Arch Microbiol. 1978;117(3):259–63.

55. Hara KY, Kondo A. ATP regulation in bioproduction. Microb Cell Fact. 2015;14:198.

56. Garay LA, Boundy-Mills KL, German JB. Accumulation of high-value lipids in single-cell microorganisms: a mechanistic approach and future perspectives. J Agric Food Chem. 2014;62(13):2709–27.

57. Wang X, Tang X, Chen H, Zhang H, Chen YQ, Zhao J, Chen W. Purification and characterization of isocitrate dehydrogenase from *Mortierella alpina*. Process Biochem. 2022;121:575–83.

58. Passoth V. Lipids of yeasts and filamentous fungi and their importance for biotechnology. In: Biotechnology of yeasts and filamentous fungi. Springer. 2017;149–204.

59. Tehlivets O, Scheuringer K, Kohlwein SD. Fatty acid synthesis and elongation in yeast. Biochim Biophys Mol Cell Biol Lipids. 2007;1771(3):255–70.

60. Ludovico P, Rodrigues F, Almeida A, Silva MT, Barrientos A, Corte-Real M. Cytochrome c release and mitochondria involvement in programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. Mol Biol Cell. 2002;13(8):2598–606.

61. Middelhoven WJ. Catabolism of benzene compounds by ascomycetous and basidiomycetous yeasts and yeastlike fungi. A literature review and an experimental approach. Antonie Van Leeuwenhoek. 1993;63(2):125–44.

62. Jonsson LJ, Martin C. Pretreatment of lignocellulose: formation of inhibitory by-products and strategies for minimizing their effects. Bioresour Technol. 2016;199:103–12.

63. Yu R, Vorontsov E, Sihlbom C, Nielsen J. Quantifying absolute gene expression profiles reveals distinct regulation of central carbon metabolism genes in yeast. eLife. 2021;10:e65722.

64. Fredlund E, Beerlage C, Melin P, Schnürer J, Passoth V. Oxygen and carbon source-regulated expression of PDC and ADH genes in the respiratory yeast Pichia anomala. Yeast. 2006;23(16):1137–49.

Publisher's Note

Rhodotorula yeasts are oleaginous species with high biotechnological potential. We produced high-quality genome assemblies and annotations through genomics and transcriptomics. Mitogenomes, putative extrachromosomal elements and antisense transcripts were identified. Ploidy, karyotype and evolutive relationships were predicted. We then investigated *R. toruloides* transcriptional regulation on different carbon sources. The results presented in this thesis expand our understanding of the genome organization and gene expression in *Rhodotorula* yeasts, with potential applications for establishing them as versatile cell factories through genetic engineering.

Giselle de la Caridad Martín Hernández received her graduate education at the Department of Molecular Sciences, SLU Uppsala, Sweden. She holds a *Licenciatura* degree (equivalent to M.Sc.) in Biochemistry and Molecular Biology from Havana University, Cuba.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.