



The
University
Of
Sheffield.

Exploring the metabolic and genetic potential of oleaginous yeasts for crude glycerol valorization

By:
Valériane Keita

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Engineering
Department of Chemical and Biological engineering

October 2021

(This page has been left blank intentionally)

Acknowledgements

For their vision and supervision, I would like to thank Dr. Tuck Seng Wong, Pr. Dong-Yup Lee and Dr. Meiyappan Lakshmanan.

This project is tripartite collaboration between the Singaporean Agency for Science, Technology and Research (A*STAR), the University of Sheffield and the oleochemical company CRODA. So, I would like to thank those three bodies for their support and especially Sheffield Engineering Grad school, the A*STAR Graduate academy, the Departement of Chemical and Biological Engineering, the Bioprocessing Technology Institute and CRODA Hull and Ditton.

My gratitude also extends to Dr. Kang Lan Tee for her mentorship and solicitude. To Dr. Amir Zak B. Abdullah Zubir, Dr. Hossam Omar Ali for their guidance and recommendations that facilitated my transition to Singapore. And to Dr. Inas Al-nuaem, Dr. Abdulrahman Alessa, Dr. José Avalos Diaz, Dr. Miriam Gonzalez-Villanueva, Dr. Pawel Jajesniak, for their sympathy and benevolence.

My profound gratitude goes out to my international crew – Robert Berchtold, Yoonmi Choi, ‘Demi’ Huang, Erez Israeli, Abayomi Johnson, Melvin Mikuzi, Ariane Noula Kamkeng, Karen Shih, Hyunsu Shin and Anthony Torres – for keeping me grounded while reminding me that there is more to life than my research. I would also like to salute Kevin Lestcher, Noëlla Mukobo, Laura Pan, Jeremy Wagner for being my core French anchor abroad. I will cherish the memories of the good moments spent together.

To the friends that despite the distance (and a pandemic) stayed by my side on this winding path – Lyly Lim, Fanny Mermet-Meillon, Sophie Morin, Aurelie Pilato, Maïline Sey, Michel Song, Sina Zebarjadian, – I would like to express my most sincere appreciation for their trust, support, patience, and love. A special mention goes out to my “compagnionne” for our much-anticipated Monday exchanges that punctuated my weeks and so much more.

Finally, to my loving family, I would like to express my deepest gratitude. Especially to my mother Françoise and sister Sarah for being my biggest supporters and inspirations.

Abstract

The world is facing unprecedented challenges that will force us as a society to change the way we consume and produce probably by transitioning from a fossil-based economy toward a bio-based one. Microbial cell factories will be at the center of such transition bearing the promises of more sustainable manufacturing processes. Among the promising wastes explored as feedstock for biomanufacturing, crude glycerol is particularly popular. As a by-product of the oleochemical industry, crude glycerol could first be used to provide a replacement to the current oleochemicals, especially the ones that are difficult to access. Oleaginous yeasts have been described to be able to convert glycerol into diverse oleochemicals, but their current performance is not competitive and enhancement strategies are so far limited by a lack of knowledge of lipid accumulation mechanisms from glycerol. A better understanding of the genetic and metabolic landscape of oleaginous yeasts could help improve their glycerol valorization abilities. Since yeasts exhibit a wide range of performances, the study aimed at comparing the genetic and phenotypic characteristics of a panel of 17 oleaginous species while investigating the molecular basis of lipid accumulation from one type of crude glycerol named sweetwater.

The relevance and scope of the study are thoroughly discussed in **Chapters 1** and **2** which form the introduction of this manuscript (**Part I**). The introduction depicts the high industrial potential of crude glycerol and oleaginous yeasts. **Part I** also contains an assessment of current state-of-the-art highlighting the remaining bottlenecks limiting the advent of oleaginous yeast oil production from crude glycerol. The following part of the manuscript (**Part II**) is composed of 3 chapters describing the methods used in this study. While **Chapter 3** is dedicated to general methods, **Chapters 4** and **5** respectively describe an indirect Nile Red-based lipid determination method and a pipeline for the functional annotation of yeast genomes both optimized for comparison

of the studied unconventional yeasts. Applications of the methods described in **Part II** can be found in **Part III** and **IV**. In **Part III**, new insights into oleaginous yeasts metabolism and genomics were discussed. While the suitability of sweetwater as feedstock for oleaginous yeasts was demonstrated in **Part IV**. **Part IV** also looks into the molecular basis of lipid accumulation from sweetwater in *R. toruloides* (a species with high industrial potential for crude glycerol valorization). **Parts III** and **IV** respectively cover **Chapters 6** and **7**. Overall this work highlights the power of combining microbiology and comparative omic technologies to identify opportunities for applied research (here yeast oil production from crude glycerol). First, combining biochemical characterization and pan-genomics information in **Chapter 6** revealed that species with the most singular genes of lipid metabolism are also the species with the highest number of species-specific genes (i.e. *C. curvatus* and *Y. lipolytica*). But unfortunately, the level of functional annotation of species-specific genes was found critical and limited the depth of the exploration. Continuing efforts to improve the level of functional characterization of unconventional yeasts proteins should alleviate this limitation in the future. The growth and lipid production in sweetwater of strains from species highlighted in **Chapter 6** were good but yet outperformed by the strain *R. toruloides* NRRL Y-6987. The transcriptomics response to sweetwater of this strain was then studied in order to identify noteworthy features. The analysis pointed out that the sugar and nitrogen contained of sweetwater were supporting the increased lipid accumulation of the strain. Differences in performances across the studied panel could thus be partially imputed to differences in nitrogen and carbon preferences. The analysis also showed that lipid accumulation and growth lie in a very delicate balance. Balance that should be taken into account in any further engineering attempt. As detailed in the conclusion and final part (**Part V**) of this manuscript, the knowledge generated could lead to improved enhancement strategies of oleaginous yeasts for industrial applications

through advanced predictive methods like genome-scale modelling which hopefully will enable the production of competitive yeast oil from crude glycerol.

Table of content

<i>Acknowledgements</i>	<i>ii</i>
<i>Abstract</i>	<i>iii</i>
<i>Table of content</i>	<i>vi</i>
<i>List of tables</i>	<i>xi</i>
<i>List of figures</i>	<i>xii</i>
<i>List of Appendices</i>	<i>xiv</i>
<i>Abbreviations</i>	<i>xv</i>
<i>Declaration</i>	<i>xxi</i>
Part I – Introduction	1
Chapter 1 - On the relevance of crude glycerol as feedstock for biomanufacturing and the potential of oleaginous yeasts as cell factories: a literature review	2
1.1 - Foreword	2
1.2 - Crude glycerol as feedstock for biomanufacturing	4
1.2.1 - Market, accessibility, and availability	5
1.2.1.a - Sources	5
1.2.1.b - Evolution of glycerol market over the last two decades	6
1.2.1.c - Glycerol applications and availability	9
1.2.2 - Composition and usability	11
1.2.2.a - Glycerol-consuming cell factories and their products	11
1.2.2.b - Crude glycerol or crude glycerols?	13
1.2.2.c - Crude glycerol compatibility with biological applications	15
1.2.2.d - Improving crude glycerol usability	18
1.2.1 - Remaining challenges and prospects	22
1.3 - Oleaginous microorganisms: rising workhorses for the production of microbial oils	26
1.3.1 - The oleaginous realm	26
1.3.2 - Industrial uses of oleaginous microorganisms: past, present and outlook	27
1.3.3 - Lipid metabolism in yeasts: from basic concepts to oleaginous metabolism	32
1.3.3.a - Getting started with lipid metabolism in yeasts	32
1.3.3.b - Well-documented specificities of oleaginous yeasts	38
1.3.4 - Advances, trends, and remaining gaps in oleaginous yeast research	41
1.3.4.a - Advances in metabolic engineering of oleaginous yeasts	41
1.3.4.b - Systems biology and ‘omic’ technologies to the rescue	46

1.3.4.c - Prospects	48
Chapter 2 - Project description	50
2.1 - Context, aims and objectives.....	50
2.2 - Organization of the study	53
Part II – Methodologies for unconventional yeasts	54
Chapter 3 - General materials and methods	55
3.1 - Materials	55
3.1.1 - Chemicals	55
3.1.2 - Equipments	56
3.1.3 - Strains and genomes	57
3.2 - Cell culture.....	57
3.2.1 - Media and plates.....	57
3.2.2 - Cell maintenance and cultivation.....	59
3.3 - Lipid analysis.....	60
3.3.1 - Extraction procedure.....	60
3.3.2 - TLC.....	61
3.3.1 - General procedure for Nile Red staining.....	61
3.4 - Pan-genomics: principles and definitions	62
Chapter 4 - Pan-genomic data can improve the functional annotation of unconventional yeasts genomes	64
4.1 - Introduction.....	64
4.2 - Pipeline development	66
4.2.1 - Comparing tools for functional annotation.....	66
4.2.2 - Combination-validation strategy	72
4.2.3 - Method description	74
4.3 - Results and discussion	76
4.3.1 - Coverage against representation in the databases.....	76
4.3.2 - Functional characteristics of the pan-genome	82
4.4 - Conclusion: remaining challenges and possible improvements.....	84
Chapter 5 - Reconciling Nile red staining protocols and investigating of the method scalability	87
5.1 - Introduction.....	87
5.2 - Material and methods	92
5.2.1 - Cell culture.....	92
5.2.2 - Plate testing and reading position selection.....	92
5.2.2.a - Modification of the general procedure.....	92
5.2.2.b - Effect of lid presence and pipette comparison	93
5.2.2.c - Comparison of row and column sample loading.....	93

5.2.2.d - Selection of a suitable blank and cultivation scale	94
5.2.2.e - Linearity against OD ₆₀₀ for small cultivation scale	94
5.2.2.f - Exploring hydrolysis as a permeation method	94
5.2.2.g - Wavelength comparisons: procedure with preheating	95
5.2.2.h - Final adjustments to the optimized method	95
5.2.3 - Statistical analysis	96
5.3 - Results	96
5.3.1 - Parameters related to data acquisition	96
5.3.1.a - Plate and lid	97
5.3.1.b - Sample relative position on the plate	98
5.3.1 - Parameters related to assay preparation	100
5.3.1.a - Pipettes comparison	100
5.3.1.b - Blank composition	101
5.3.1.c - Cultivation scale	103
5.3.2 - Addressing remaining bottlenecks	107
5.3.2.a - Permeabilization	108
5.3.2.b - Emission wavelength selection	111
5.3.3 - Method summary	114
5.3.3.a - General recommendations and data normalization	114
5.3.3.b - Rapid screening of a panel of unconventional yeasts	116
5.3.3.c - Application of the method to the monitoring of lipid accumulation	117
5.4 - Discussion and concluding remarks	117
Part III – New insights into oleaginous yeasts metabolism	120
Chapter 6 - Combined genomic and phenotypic approach to explore the diversity of oleaginous yeast with high potential for lipid production	121
6.1 - Introduction	121
6.2 - Materials and methods	122
6.2.1 - Cell culture	122
6.2.2 - Lipid extraction and thin layer chromatography	123
6.2.3 - Principal component analysis	123
6.3 - Results and discussion	123
6.3.1 - The pan-genomic analysis reveals conserved functional categories	123
6.3.2 - The phenotypic characterization nuances pan-genomic predictions	125
6.3.2.a - Carbon utilization	125
6.3.2.b - Thermotolerance and halotolerance	128
6.3.2.c - Biomass production	131
6.3.3 - Diving deeper into genes related to lipid metabolism	134
6.3.3.a - Overview of current knowledge	134

6.3.3.a - Comparative analysis of genes related to lipid metabolism: principal component 1 (PC1)	136
6.3.3.b - Comparative analysis of genes related to lipid metabolism: principal component 2 (PC2)	139
6.3.3.c - Comparative analysis of genes related to lipid metabolism: principal component 3 (PC3)	143
6.3.3.d - Comparative analysis of genes related to lipid metabolism: PC4 and beyond	146
6.4 - Concluding remarks	148
Part IV – Toward the enhancement of sweetwater utilization	152
Chapter 7 - Sweetwater: underrated crude glycerol as feedstock for sustainable lipid production in unconventional yeasts	153
7.1 - Introduction	153
7.2 - Materials and methods	155
7.2.1 - Media and sweetwater-based preparations	155
7.2.2 - Cell culture	155
7.2.3 - Nile red staining	156
7.2.4 - Lipid extraction and TLC	156
7.2.5 - RNA extraction and sequencing	156
7.2.6 - Transcriptomic analysis	158
7.2.7 - Genomic analysis	159
7.2.8 - Phenol-sulfuric sugar assay	159
7.3 - Results and discussion	159
7.3.1 - Investigated strains and their glycerol tolerance	159
7.3.2 - Evaluating strain affinity with the crudest form of the waste	161
7.3.3 - Comparing the lipid production of 21 strains grown in SW ₁₅ and SW _{1.5} SC	164
7.3.4 - Investigating the transcriptomic changes prompted by the use sweetwater in <i>R. toruloides</i> (12)	169
7.3.4.a - The global comparison reveals a predominance of nutrient-related responses	170
7.3.4.b - Focus on the comparison between SW ₁₅ and Gly ₁₅	174
7.4 - Concluding remarks	185
Part V – Conclusion and prospects	188
Chapter 8 - Conclusion	189
8.1 - Highlights	189
8.2 - Perspectives and future work	192
8.2.1 - Methodologies for unconventional yeasts (Part II)	192
8.2.2 - New insights into oleaginous yeast metabolism (Part III)	194
8.2.3 - Toward a enhanced utilization of sweetwater (Part IV)	195
8.3 - Concluding thoughts	197

<i>Appendices</i>	198
<i>References</i>	223

List of tables

Table 1-1 Value-added products accessible by bioconversion of glycerol and their principal use.....	12
Table 1-2 Improvements of yeast native lipid production through metabolic engineering.....	43
Table 3-1 Strains and genomes investigated in this study.....	58
Table 4-1 General features of investigated genomes.....	69
Table 4-2 Comparative table of candidate tools for functional annotation	70
Table 5-1 Summary of Nile red staining protocols published between 2004 and 2021 for the estimation of yeast lipid contents	89
Table 5-2 Summary of Nile red staining protocols published between 2004 and 2021 for the estimation of yeast lipid contents	89
Table 5-3 Summary of modifications made to the general procedure	93
Table 6-1 Summary of carbon utilization abilities	127
Table 6-2 Overview of halotolerance and thermotolerance	130
Table 7-1 Tolerance to refined glycerol.	160
Table 7-2 Investigating sweetwater as a feedstock for yeast growth.	162
Table 7-3 Mining for putative acetyl-CoA synthetases (ACS) and ATP:citrate lyases (ACL).	167

List of figures

Figure 1-1 Manufacturing processes leading to bio-glycerol as a by-product.	5
Figure 1-2 Nomenclature and classification of biotic and synthetic glycerol.	7
Figure 1-3 Distribution of biodiesel global production in 2020.	8
Figure 1-4 Overview of simple strategies used to overcome crude glycerol composition.	19
Figure 1-5 Overview of microbial glycerol catabolism.	23
Figure 1-6 Biochemistry of neutral lipid accumulation in oleaginous yeasts.	34
Figure 2-1 Organisation of this study.	53
Figure 3-1 Illustration of the pan-genomic categories.	63
Figure 4-1 Graphical illustration of the annotation pipeline principle.	68
Figure 4-2 Scheme illustrating the combination-validation developed and examples.	71
Figure 4-3 Influence of the number of annotation tools on the KO coverage.	73
Figure 4-4 Method illustration.	76
Figure 4-5 Summary of functional annotation quality and phylogenetic representation of unconventional yeast in phylogenomic databases.	77
Figure 4-6 Analysis of the pan-genome annotation.	83
Figure 5-1 - Plate and reading position selection.	97
Figure 5-2 - Influence of lid presence.	98
Figure 5-3 - Influence of the relative position of samples on the plate.	99
Figure 5-4 - Comparison of multichannel and monochannel pipette for assay preparation.	101
Figure 5-5 - Choice of blank composition.	102
Figure 5-6 - Investigating the effect of cultivation scale on the fluorescence of NR-stained <i>R.</i> <i>toruloides</i> (12) and <i>Y. lipolytica</i> (20).	104
Figure 5-7 - Behaviour during Nile red assay of 5 yeasts grown in 50-mL centrifuge tubes.	105
Figure 5-8 - Exploring the use of acid hydrolysis as a permeation method to improve the efficiency of NR staining.	109
Figure 5-9 - Linearity of fluorescence signal (RFUc) against sample concentrations of strains with presumably low permeability to Nile Red.	111
Figure 5-10 - Emission spectra of 5 NR-strained strains.	112
Figure 5-11 - Comparing the fluorescence of NR-strained strains at two emission wavelengths (625 nm and 535 nm).	113
Figure 5-12 - Using Nile Red straining to monitor <i>Y. lipolytica</i> (21) and <i>L. elongisporus</i> lipid accumulation over time at two emission wavelengths (625 nm and 535 nm).	114
Figure 6-1 Functional distribution of pan-genomic genes.	124
Figure 6-2 Biomass production in glucose-containing media.	133

Figure 6-3 Oleaginous yeasts lipid metabolism at glance	137
Figure 6-4 Principal component analysis of genes associated with lipid metabolism for 16 unconventional yeast species.....	138
Figure 7-1 Lipid production in sweetwater based media.	165
Figure 7-2 <i>R. toruloides</i> (12) in SW15, SW1.5SC, Gly15 and Gly1.5SC.	170
Figure 7-3 Comparison of <i>rto</i> (12) transcriptome in 4 glycerol-based media.	171
Figure 7-4 Comparison of <i>R. toruloides</i> (12) transcriptome in SW15 and Gly15 (Ctrl).	175
Figure 7-5 Scheme summarizing the main differences between <i>rto</i> (12) growth in SW15 compared to the growth in Gly15 (Ctrl).	179
Figure 8-1 Anticipated future of functional annotations.	194

List of Appendices

Appendix 1 Phylogenetic tree showing the relationship between studied yeasts and reference yeasts.....	195
Appendix 2 Crude glycerol compositions reported in the literature.....	196
Appendix 3 Illustrations of Chapter 4.....	199
Appendix 4 Contributions of the principal components to the PCA presented in Figure 6-4.....	200
Appendix 5 Proportion of core, accessory, and unique genes per strain for genes associated with metabolic functions.....	201
Appendix 6 Aspect and properties of the three types of sweetwater used in this study.....	202
Appendix 7 Results of ANOVA and Tuckey's post hoc multiple comparison for parameters described in Table 7-1.....	203
Appendix 8 Results of ANOVA and Tuckey's post hoc multiple comparison for parameters described in Table 7-2.....	204
Appendix 9 (a)Raw data plotted in Figure 7-1a and associated standard deviations (b) Results of the ANOVA and Tuckey's post hoc multiple comparison.....	205
Appendix 10 Fluorescence microscopic observation of Nile red stained <i>R. toruloides</i> (12) cells and macroscopic cell clump	206
Appendix 11 Annotation of genes represented in Figure 7-3.....	207
Appendix 12 Top50 of annotated genes contributing the most to PC1 in the differential expression analysis of SW15 vs Gly15.....	209
Appendix 13 Differential expression results of SW15 vs Gly15 of relevant genes associated to lipid metabolism.....	212

Abbreviations

%	percentage
°C	degree Celsius
(-)	arbitrary unit
(-)-	dextrorotary
(+)-	levorotary
xg	multiple of g-force
∅	diameter
±	plus or minus
~	approximately
#	number
1,2-PDO	1,2-propanediol
1,3-PDO	1,3-propanediol
a.u.	arbitrary Unit
aa	amino acid
ACAT	acyl-CoA cholesterol acyltransferase
ACC	acetyl-CoA carboxylase
ACL	ATP:citrate lyase
ACP	acyl carrier protein
ADO	aldehyde deformilating oxygenase
AGPAT	1-acylglycerol-3-phosphate O-acyltransferase
AH-FAR	aldehyde-forming fatty acyl-CoA reductase
AL-FAR	alcohol-forming fatty acyl-CoA reductase
AMP	adenosine monophosphate
antiSMASH	antibiotic and secondary metabolites analysis shell
ARA	arachidonic acid
ATP	adenosine triphosphate
BBH	bi-directional best hit
BFA	branched fatty acid
BHT	butylated hydroxytoluene
BiGG	biochemical Genetic and Genomic knowledgebase
BLAST	basic local alignment search tool
Blk	blank
CAGR	compound annual growth rate

CBE	cocoa butter equivalent
CBI-PKU	center for Bioinformatics Peking University
CC	carrying capacities
CDP	cytidine diphosphate
cm	centimeter(s)
COIII	coproporphyrinogen III oxidase
CoA	coenzyme A
COG	cluster of orthologous groups
CPS	cellular processes and signaling
CSM	complete supplement mixture
D-	<i>dextro configuration</i>
DAG	diacylglycerol
DCA	α,ω -dicarboxylic acid
DCW	dry cell weight
DE	differential gene expression
DGAT	diacylglycerol acyltransferase
dH	dehydrogenase
DH-A	docosahexaenoic acid
DHA	dihydroxyacetone
DHAP	dihydroxyacetone phosphate
DHAK	dihydroxyacetone kinase
DMSO	dimethyl sulfoxide
DOE	design of experiment
DTU	technical University of Denmark
e.g.	<i>exempli gratia (for example)</i>
e-value	expected value (sorting metric)
EC	enzyme commission
EMBL	european molecular biology laboratory
EggNOG	evolutionary genealogy of genes: non-supervised orthologous groups
EPA	eicosapentaenoic acid
ER	endoplasmic reticulum
F₂₅₄	fluorescence factor absorbing light at 254 nm
FA	fatty acid
FAA	fatty acyl-CoA synthetase

FAEE	fatty ethyl esters
FAL	fatty alcohol
FAO	fatty alcohol oxidase
FAME	fatty methyl ester
FAS	fatty acid synthetase
FFA	free fatty acid
FU	function unknown
g	gram(s)
g/L	gram(s) per liter
G3P	glycerol-3-phosphate
G3PdH	glycerol-3-phosphate dehydrogenase
GC content	Guanine/Cytosine content
GdH	glycerol dehydrogenase
GEM	genome-scale metabolic models
GENRE	genome-scale metabolic models
GK	glycerol kinase
GLA	γ -linolenic acid
GlcSC₆₀	glucose synthetic complete with a C/N ratio of 60
Gly DHt	glycerol dehydratase
GO	gene ontology
GPAT	glycerol-3-phosphate acyltransferase
GSMM	genome-scale metabolic models
GTE	glycerine-to-epichlorohydrin
H	hydrogen
HEAR	high Eruci Acid Rapeseed
HMMer	hidden Markov marker and profiles
HR	homologous recombination
HTML	hyperText Markup Language
i.e.	<i>id est (that is)</i>
ICDH	isocitrate dehydrogenase
iD	identifier
iDg	gene identifier
ISP	information storage and processing
JSON	JavaScript Object Notation

KAAS	KEGG automatic annotation server
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology number
KOALA	KEGG orthology and link annotation
KOBAS	KEGG orthology based annotation system
λ_{ex}	Excitation wavelength
λ_{em}	Emission wavelength
L-	<i>Laevo</i> configuration
LA	lactic acid
LD	lipid droplet
LP	lysophosphatidic acid
μ_{max}	maximum specific growth rates
μL	microliter(s)
M	molar [mol(s) per liter]
MAG	monoacylglycerol
ME	malic enzyme
MFS	major facilitator superfamily
mg	milligram(s)
min	minute(s)
mL	milliliter(s)
MMseqs2	many-against-Many sequence searching
Mr.	mister
MONG	non-glycerol organic matter
n.d.	not determined
NAD(P)	nicotinamide adenine dinucleotide (phosphate)
NCBI	national center for biotechnology information
NHEJ	non-homologous end joining
nm	nanometer
NR	nile Red
OD₆₀₀ or OD	Optical density at 600 nm
OG	orthologous group
OFAT	one factor at a time
p-	para
PA	phosphatic acid

PAP	phosphatidate phosphatase
PC	principal component
PCA	principal component analysis
PHA	polyhydroxyalkanoates
PLC	phospholipase C
POP	persistent organic pollutants
POX	peroxisomal acyl-CoA oxidases
PUFA	polyunsaturated fatty acid
R²	coefficient of determination
RFU	relative Fluorescence Unit
RFUc	background corrected RFU
RFUco	background corrected RFU normalized by OD ₆₀₀
RFUcd	Background corrected RFU normalized by dilution factor (volume)
RFUcx	rFUc or RFUcd or RFUco
rpm	revolutions per minute
RT	room temperature
SBH	single-directional best hit
SCD	stearoyl-CoA desaturase
SCO	single cell oil
SD	standard deviation
SE	steryl esters
sec	second(s)
sp.	species
SW	sweetwater
t₀	initial time
TAG	triacylglycerol
TE	thioesterase
TGL	triacylglycerol lipase
TLC	thin layer chromatography
tmid	time required to reach half of the carrying capacity
UK	United Kingdom
USA	United States of America
v/v	volume per volume
VLCFA	very long-chain fatty acids

w/v	weight per volume
w/w or wt%	weight per weight
WS	wax ester synthase
YM	yeast extract malt
YNB	yeast nitrogen base without amino acids and without ammonium
YPD	yeast peptone dextrose

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university. All sources of information herein have been duly referenced. Some parts of this thesis have been adapted from the book chapter: Keita, V. M., Gonzalez-Villanueva, M., Wong, T. S. & Tee, K. L. Microbial Utilization of Glycerol for Biomanufacturing. in Engineering of Microbial Biosynthetic Pathways (eds. Singh, V., Singh, A. K., Bhargava, P., Joshi, M. & Joshi, C. G.) 245–302 (Springer, 2020).

Part I – Introduction

Chapter 1 - On the relevance of crude glycerol as feedstock for biomanufacturing and the potential of oleaginous yeasts as cell factories: a literature review

1.1 - Foreword

Considering biomanufacturing as the use of biological systems and/or extracts to manufacture goods with commercial potential, the beginnings of biomanufacturing can be traced back to prehistoric times when it provided fermented goods to the populations. Even though cheese, bread and wine productions are still relevant today, biomanufacturing greatly evolved over the years to become a field expected to provide solutions to some of the greatest challenges of the century. Challenges like food, health and energy security as well as problematics related to climate change and the depletion of fossil resources (Zhang et al., 2017). A contribution of biomanufacturing is anticipated because it could provide supplements or replacements to current sources of food, medicines, energy and/or petrochemicals. Not only biomanufacturing could deliver more sustainable ways to produce essential commodities but it also comes with great recycling and depolluting opportunities.

The progressive awareness around the environmental potential of biomanufacturing is flagrant when looking at the evolution of feedstocks used for bioethanol production. The transformation of ethanol biomanufacturing is particularly informative since it ranks among the most mature bioconversion processes (Clomburg et al., 2017). The feedstocks used for bioethanol production went from edible sugars to agricultural, forest or municipal wastes/residues (Lee & Lavoie, 2013). And the latest developments in bioethanol research are looking into production processes based on photosynthetic microorganisms that would use carbon dioxide as feedstock (Li-Beisson & Peltier, 2013).

This type of development hints at an eventual contribution of biomanufacturing in the mitigation of the greenhouse effect. The use of waste as feedstock for microbial cells factories could not only have an environmental impact but also economic fallouts since the choice of feedstock is crucial for the economic viability of certain processes (Navarrete et al., 2020).

Hence the exploration of waste as feedstock for biomanufacturing expanded far beyond bioethanol production. Lignocellulosic wastes are widely investigated for that purpose since they constitute the most abundant untapped reserve of assimilable sugars (Menon & Rao, 2012). Lignocellulosic wastes are available in numerous forms and compositions, but they are all blends of cellulose (polymer of glucose) and hemicellulose (heteropolysaccharide of glucose, mannose, galactose, xylose, and arabinose) entangled with lignin (aromatic polymer). The main impediment to their utilization is the expensive pre-treatment required for the release of assimilable monomers (i.e. glucose and xylose) from the polysaccharides. Pre-treatments that also usually lead to lignin degradation into phenolic compounds and small organic acids that can inhibit cell growth (Navarrete et al., 2020; Shahzadi et al., 2014). The need for pre-treatments, the adverse effect of lignin degradation by-products and the complexity of mixed carbon sources utilization prompted the exploration of non-lignocellulosic wastes. The most prominent non-lignocellulosic candidate feedstock for biomanufacturing is crude glycerol that despite its content in potentially inhibitory impurities can be used without heavy pre-treatments. Furthermore, high volumes of crude glycerol are available at low-cost without seasonal dependency (Coronado, 2013) which only adds up to the waste attractivity.

Yet, the number of microorganisms able to efficiently utilize glycerol as sole carbon source might limit its reach compared to glucose-containing lignocellulosic wastes since glucose is a more commonly accepted carbon source. But at least, microorganisms able

to utilize glycerol are scattered across the whole microbial realm (Chen & Liu, 2016) unlike microorganisms able to utilize carbon dioxide or other gases that are less represented (Navarrete et al., 2020). So, albeit appealing, the use of gases as feedstock will require further developments not only to harness the natural abilities of latter organisms in cost-efficient processes but also to extend the range of products that can be biomanufactured based on this technology. Regarding lignocellulosic wastes and crude glycerol, huge progress have been achieved regarding microorganisms' abilities to utilize the wastes to produce biomolecules of industrial interest (Chilakamarry et al., 2021; Iqbal et al., 2013; Ning et al., 2021). Yet, when crude glycerol has already been tested multiple times as feedstock at pilot-scale (Cofré et al., 2016; Sudiyani et al., 2013; Szymanowska-Powałowska & Białas, 2014; Yen et al., 2015) lignocellulosic wastes used as feedstock rarely exceeded the stage of laboratory demonstration (Saini et al., 2020). The implementation of lignocellulosic wastes as feedstock heavily relies on the existence of robust pre-treatment methods that are still mainly in development although some technologies recently reached the pilot-scale or early industrial stages (Kumar et al., 2020).

Hence, the work described in this manuscript focused on the use of crude glycerol as feedstock for biomanufacturing purposes. In this first chapter, elements supporting the choice of crude glycerol as study subject will be discussed, as well as the rationale behind the type of cell factories selected for this study. This chapter will also provide an overview of the current knowledge and gaps surrounding the use of crude glycerol as feedstock for the selected cell factories.

1.2 - Crude glycerol as feedstock for biomanufacturing

A good candidate feedstock for biomanufacturing must (1) be able to support microbial growth and product accumulation, (2) available in sufficient quantities (3) easily accessible and (4) inexpensive.

1.2.1 - Market, accessibility, and availability

1.2.1.a - Sources

The bulk of the glycerol accessible today is a bio-glycerol derived from vegetable oils or animal fats. Although oil and fat are both composed of triglycerides (TAGs), they differ in their physical forms as liquid (oil) or solid (fat) at ambient temperature. The oleochemical industry mainly uses refined soybean oil, palm oil, rapeseed oil, sunflower oil, coconut oil, colza oil, palm kernel oil, tallow, and lard (Hill, 2001). Among these, colza, soybean, and palm oils are most suitable for biodiesel production (Pagliaro and Rossi, 2008). In any case, the use of edible oil for biodiesel production has been under scrutiny due to its competition with food security. To improve sustainability, cost-efficiency and reduce their environmental impact, these edible oils (i.e., refined oils), are being supplanted by second-generation feedstocks including some unrefined oils (e.g., jatropha oil, waste cooking oil and tallow oil) (Tan et al., 2013).

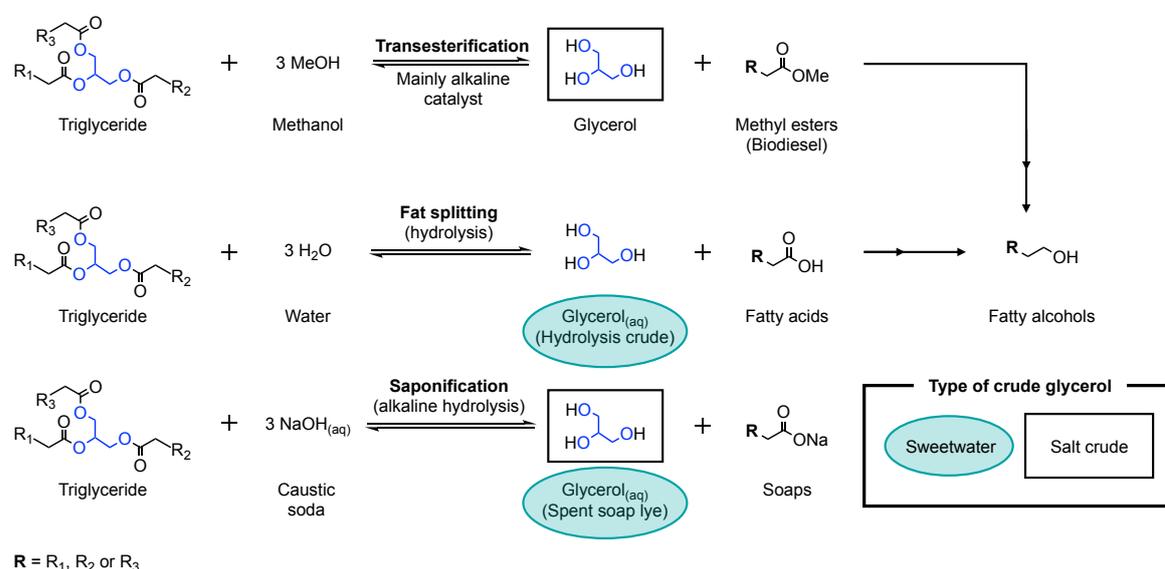


Figure 1-1 Manufacturing processes leading to bio-glycerol as a by-product. The chemical reactions for industrial processes involving transesterification (top), fat splitting (middle) and saponification (bottom) of triglyceride (*i.e.*, oils and fats) are shown. All three processes produce glycerol as a side-product. Derivatives of fatty acids and methyl esters, like fatty alcohols, can be obtained *via* further processing.

Food-grade oil refining includes steps such as degumming (lecithin and phosphorus removal), deacidification (free fatty acid removal), neutralization, bleaching and deodorization (Pagliaro and Rossi, 2008; Patel *et al.*, 2016). Consequently, impurities like colloidal matters, phospholipids, FFAs, and coloring agents are removed from food-grade edible oils, resulting in crude glycerol with a low level of contaminants (Jungermann and Sonntag, 1991). These steps increase the shelf lives of refined oils and are especially suitable for fat splitting (Angers *et al.*, 2003). On the contrary, unrefined oils used in biodiesel production (including soap stock and rendered oils) contain all the above impurities, which eventually end up in the discharged crude glycerol.

Glycerol is discarded from biodiesel production (60–70%), followed by fatty acids and fatty alcohols production (30%), with a minor fraction from soap making (Ciriminna *et al.*, 2014; Pagliaro, 2017). The chemical reactions in these three main processes are (1) transesterification utilizing methanol to produce fatty acid methyl ester (otherwise called FAME or biodiesel) that releases 100 kg of glycerol for every ton of biodiesel produced, (2) fat splitting or hydrolysis into fatty acids, and (3) saponification or hydrolysis under alkaline conditions to produce salts of fatty acid (also known as soap) as appeared in **Figure 1-1**.

1.2.1.b - Evolution of glycerol market over the last two decades

Before the explosion of biodiesel in 2004, glycerol production was equally divided between the oleochemical industry (generating bio-glycerol) and petrochemical industry (generating synthetic glycerol from C3 derivatives such as epichlorohydrin, allyl chloride or allyl alcohol derived from propene)(**Figure 1-2**)(Bauer & Hulteberg, 2013; Ciriminna *et al.*, 2014) and the global production was relatively stable at 0.8 million tonnes per annum (Gunstone & Heming, 2004). Then, the top three global glycerol suppliers were Procter & Gamble, Cognis, and Uniqema (presently Croda).

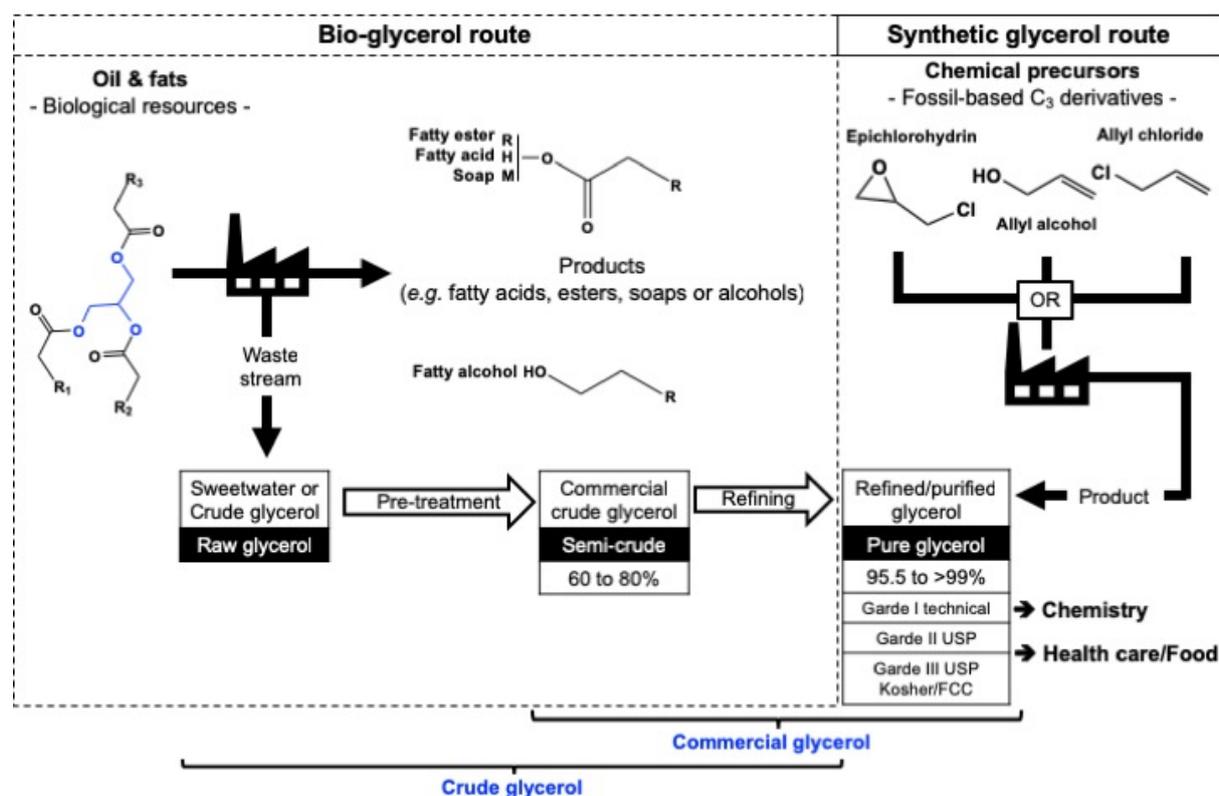


Figure 1-2 Nomenclature and classification of biotic and synthetic glycerol. Designations of bio-glycerol during processing from raw glycerol to pure glycerol are shown in black boxes (typical glycerol concentrations are specified below the black boxes in wt%). This figure illustrates the ambivalence of the term “crude glycerol”.

By 2010, the leading glycerol suppliers were biodiesel and oleochemical companies, mostly situated in Southeast Asia (Malaysia, Philippines, Thailand, and Indonesia) (Pagliaro, 2017). In 2015, Malaysia, Colombia, Argentina, and Brazil were among the biggest biodiesel producers, and glycerol supply was consolidated with four major companies (IOI Group, Wilmar International, KL Kepong and Emery Oleochemicals) representing more than 65% of the overall market (Pagliaro, 2017). And today, biodiesel production is dominated by Europe but more evenly distributed across continents (Figure 1-3)(OCDE & FAO, 2020).

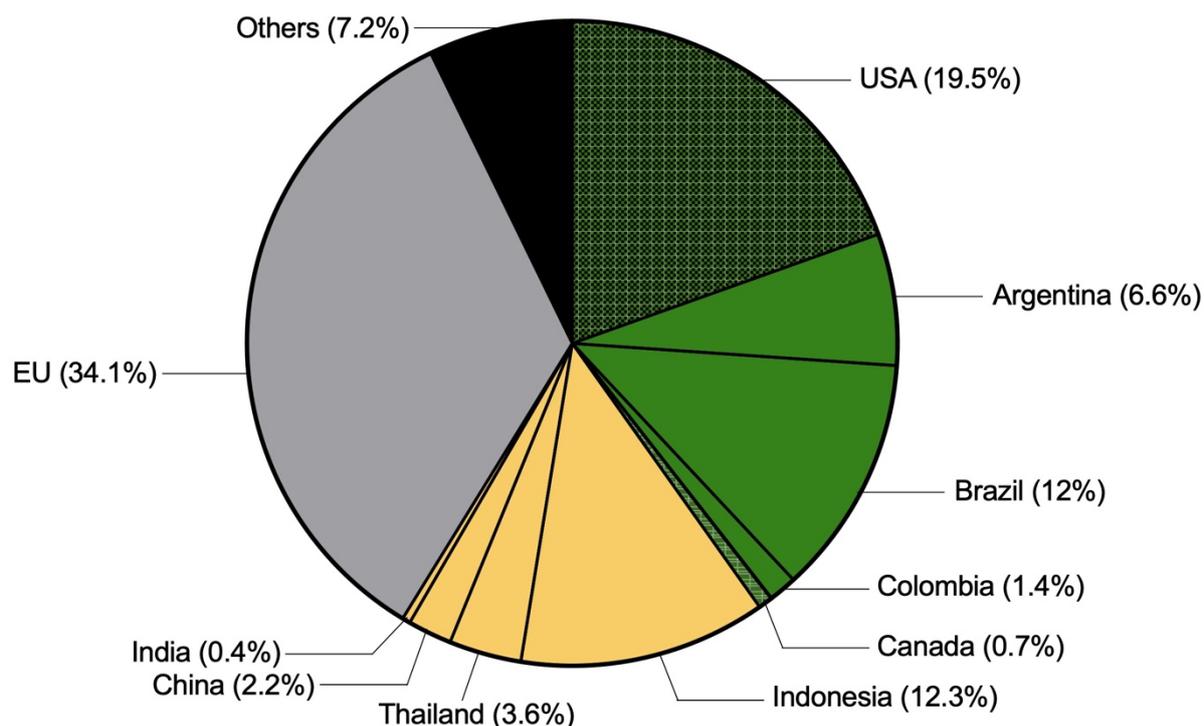


Figure 1-3 Distribution of biodiesel global production in 2020.

This diagram was plotted based on data presented in the OECD-FAO Agricultural Outlook 2020-2029.

Glycerol being a by-product of the biodiesel industry, the rapid growth of this sector generated a huge and expanding excess of glycerol that resulted in a progressive drop in glycerol prices (88.6% decrease recorded between 2000 and 2010)(Ciriminna et al., 2014). The prices of glycerol reached a critical low in 2006 such that plants producing synthetic glycerol were either shut down or retrofitted. For instance, Solvay group in France officialised their plant retrofitting in 2007 after a year of shut down (Ciriminna et al., 2014). The retrofitting was implemented with a process called Epicerol[®] that produces epichlorohydrin from glycerol (Pagliaro & Rossi, 2008) knowing that glycerol used to be synthesized from epichlorohydrin. The development of new applications and the expansion of existing markets set an upward trend in refined glycerol prices up to 2008 then prices plummeted again to reach a value between \$892 and 1069 per ton (depending on the raw material and final grade) in 2012 (Coronado, 2013). On the other hand, crude glycerol prices remained fairly stable after the price drop of 2006 and were estimated in 2011 between \$40 and 110 per ton (Coronado, 2013).

In 2019, refined glycerol was quoted at \$170/ton and refined glycerol at \$895/ton (Ruy et al., 2020). Hence refined glycerol prices stabilized in recent years when crude glycerol prices slightly recovered but remained 10 times lower than they were at the beginning of the century. As a by-product, glycerol supply is disconnected from its demand and mainly driven by the performance of the biodiesel and the oleochemical industries which causes market volatility. Given the market complexity, it is not unusual for prices to oscillate during a year. For example, in 2010, US crude glycerine price tripled from January to December rising from \$110 per ton to \$330 per ton (Coronado, 2013). The relative stability of crude glycerol price is due to its current limited applications, but the onset of new demand could result in a predictable price increase. Despite foreseeable fluctuations, the overall glycerol market is still expanding with a 4% compound annual growth rate (CAGR) forecasted between 2020 to 2027. The market was estimated at \$2.6 billion in 2019 and is anticipated to exceed \$3.5 billion by 2027 (Ruy et al., 2020).

1.2.1.c - Glycerol applications and availability

Crude glycerol is often used in low value-added applications such as animal feed, feedstock in biotechnological applications, de-icing agents, concrete additives, and road anti-dust (Ayoub and Abdullah, 2012; Pagliaro, 2017). Refined glycerol production is sometimes perceived as the major application of crude glycerol. Once refined three applications dominate glycerol demand: personal and oral care products (30%), food and beverage (>13%) and pharmaceuticals (>10%) (IHS Markit, 2018). But the use of glycerol for these applications strongly depends on glycerol price (Pagliaro, 2017). Unfavourable prices will result in glycerol replacement with cheaper alternatives such as sorbitol. When low prices favour the development of new applications for glycerol or expands existing markets (Coronado, 2013). Recent usage of glycerol includes the production of propylene glycol, epichlorohydrin, and polyether polyol, used in

polyurethane foam production (Ruy et al., 2020). Additional advances in glycerol valorisation were reviewed elsewhere (Kaur et al., 2020).

The applications of glycerol are however subjected to regional variation. For instance, in 2006, 40% of US refined glycerol was dedicated to personal & oral care products while this application only represented 15% of glycerol used in Japan. The proportion of glycerol used for pharmaceuticals was 25% and 8% for Japan and Western Europe, respectively (Greenea, 2015). These regional variations are the consequence of glycerol regional availability. The expansion of a given application from countries with a big glycerol production to countries with a lesser glycerol production is only advantageous when the “free on board” prices¹ are at the lowest so that, even with additional importations cost, glycerol remains competitive. Glycerol is a commodity that travels to the point that 2.2 million tons of crude glycerol were exported in 2019 (Ruy et al., 2020). This highlights the existence of networks of distribution which only adds to the waste accessibility even though import and transport costs could become a drawback if too high. The supply of crude glycerol was forecasted to reach 6 million tonnes by 2025. If the proportion of biodiesel-derived glycerol is predicted to increase over the coming years, the proportion of crude glycerol from other sources should remain stable (even if a small growth in the fatty acid market might be recorded) (Ciriminna et al., 2014)(for sources of crude glycerol see **section 1.2.1**). These previsions confirm that the waste is available in large quantities and will continue to be as long as the biodiesel industry remains healthy.

Hence, crude glycerol is economically attractive as feedstock for biomanufacturing due to its cost, availability, and accessibility.

¹ “Free on board” (FOB) price designates the price without insurance and freight cost as opposed to cost insurance and freight (CIF) price

1.2.2 - Composition and usability

A low-cost waste available in large volumes will only be a suitable candidate for biomanufacturing if it can be efficiently metabolized into value-added products of industrial interest.

1.2.2.a - Glycerol-consuming cell factories and their products

The natural capacity to use glycerol as a carbon source for growth, energy and metabolites production can be found in bacteria, archaea, fungi (including yeasts and filamentous fungi) and microalgae. Although some archaea have been shown to utilize glycerol (Williams et al., 2017), their industrial application remains low and is not discussed here. The efficiency of glycerol utilization varies between species and strains. This was clearly illustrated in a study where only 8 out of 126 species tested in the *Enterobacteriaceae* family were able to ferment glycerol (Holm, 2013).

Many other bacteria also utilize glycerol (e.g. *Staphylococcus aureus*, *Mycobacterium tuberculosis*, *Nocardia asteroides*, *Salmonella typhimurium*, *Mycoplasma mycoides*, *Mycoplasma pneumoniae* and *Listeria monocytogenes*) and their mechanisms for glycerol transport and dissimilation are well studied (Lin, 1976; Blötz and Stulke, 2017). Despite their low potential for biomanufacturing due to pathogenicity, their well-described glycerol metabolism could support rational engineering endeavours in strains more suitable for industrial applications. Yet, pathogenic strains will be excluded from this chapter. The maximum specific growth rates (μ_{\max}) reported in the literature vary between 0.65 h⁻¹ for *Bacillus subtilis subsp. niger*. (Kruyssen et al., 1980) to 0.012 h⁻¹ for a *Chlorella vulgaris* strain (Silva & Fonseca, 2018). If the natural microbial rates of glycerol utilization are sometimes lower than desired for industrial applications, the array of value-added chemicals accessible from glycerol is quite appealing. Most of these

molecules have multiple uses but can be sorted into 5 categories based on their major applications: for the polymer and plastic industry (LA/PHA/1,3PDO/1,2PDO), for the

Table 1-1 Value-added products accessible by bioconversion of glycerol and their principal use.

Chemicals	Type of microorganisms	Main application(s)
For the polymer and plastic industry		
Lactic acid (LA)	Bacteria (Komesu et al., 2017) Fungi (Werpy & Petersen, 2004)	Precursor of polylactic acids (plastic)
Polyhydroxyalkanoates (PHA)	Bacteria (Zhu et al., 2013)	Bioplastic could replace petroleum derived plastics
1,3-propanediol (1,3-PDO)	Bacteria (Colin et al., 2001; Jiang et al., 2014)	Starting material for the synthesis of polyester, polyethers and polyurethanes
1,2-propanediol (1,3-PDO)		Starting material for the synthesis of polyester resins
For the energy sector		
Ethanol	Bacteria (Yazdani & Gonzalez, 2007)	Biofuel
Butanol	Bacteria (Yazdani & Gonzalez, 2007)	Biofuel
Hydrogen	Bacteria (Kaur et al., 2020)	Energy carrier in fuel cells
Biogas (methane + CO ₂)	Bacteria (Kaur et al., 2020)	Alternative to fossil fuels
For human consumption (food and beverage + pharmaceutical applications)		
Propionic acid	Bacteria (Yazdani & Gonzalez, 2007)	Antifungal
Citric acid	Yeasts (Nicol et al., 2012)	Preservative
Trehalose	Bacteria (Garlapati et al., 2015)	Functional ingredient
Erythritol	Yeasts (Monteiro et al., 2018)	Functional ingredient
Mannitol	Yeasts (Monteiro et al., 2018)	Functional ingredient
For the fine chemicals industry		
Dihydroxyacetone	Bacteria (Black, 2013)	Building block
Glyceric acid	Bacteria (Yang et al., 2012)	Building block
Itaconic acid	Fungi ¹ (Zambanini et al., 2017)	Building block
Succinic acid	Bacteria (Becker et al., 2015)	Building block
3-hydroxypropionic acid	Bacteria (Becker et al., 2015)	Acrylic acid precursor
Miscellaneous		
Lipids (polyunsaturated fatty acids, carotenoids, sphorolipids)	Fungi ¹ (Nicol et al., 2012) Microalgae (Kumar et al., 2019)	Ingredient in food, pharmaceuticals, cosmetics, nutraceuticals Biofuels
Enzymes	Yeasts (Magdouli et al., 2017)	Catalysts

¹ The term “fungi” refers to both yeasts and filamentous fungi

energy sector (ethanol/butanol/biogas/hydrogen), for human consumption (propionic acid/citric acid/trehalose/erythritol/mannitol), for the fine chemicals industry (dihydroxyacetone/glyceric acid/itaconic acid/succinic acid/3-hydroxypropionic acid)

and miscellaneous applications (lipids and enzymes) (**Table 1-1**). But none of these products is currently industrially produced by bioconversion of glycerol. Not even itaconic acid despite the Rhodia patent describing its production by fermentation of glycerol (Jarry & Seraudie, 1994). Itaconic acid production by fermentation is now a well-established process but glucose supplanted glycerol as feedstock (Du, 2014). Even when traces of glycerol usage as feedstock for industrial fermentation are found, it is often used in mixture with some sugar. A good illustration of that is the use of glucose and glycerol mixtures, for succinic acid production operated by Succinity (Becker et al., 2015). The overlap between products accessible from glucose and glycerol is not helping the establishment of refined glycerol as biological feedstock, let alone crude glycerol, the low purity of which tends to further limit its applications (Sivasankaran et al., 2019).

1.2.2.b - Crude glycerol or crude glycerols?

To understand the limitations surrounding crude glycerol we need to look into its composition. This section will mainly focus on the composition of crude glycerol from transesterification and fat splitting, as these two processes discard about 90% of the glycerol in the market. The composition and quality of crude glycerol resulting from these industrial processes depend on raw material (i.e. the oil or fat) (**Figure 1-1**), catalytic conditions, and treatments (Hu et al., 2012). Defining the composition of crude glycerol is difficult for various reasons: (1) the oil and fat raw materials used in industrial processes are highly diverse, (2) these raw materials are processed via different chemical reactions, where different catalysts and reaction conditions are applied, (3) crude glycerol characterization is not standardized and data reported is often incomplete, (4) oil refining steps (i.e., degumming) before its transformation are usually not specified, and (5) crude glycerol concentrating and/or refining steps are not reported in most cases. Some studies conclude that crude glycerol composition is consistent for a given source (Hu et al., 2012), while others found significant differences between crude glycerols generated by similar processes (Hansen et al., 2009).

Hence, a survey of crude glycerol compositions reported in the literature is given in **Appendix 2**. This table illustrates the huge heterogeneity of composition observed across crude glycerols and highlights the challenge that compositional variability represents in the establishment of reproducible biomanufacturing processes based on wastes as feedstocks. Considering crude glycerol as a single entity would be misguided. It can even be argued that there are as many crude glycerols as processes are discarding them! In the same way lignocellulosic wastes are all blends of cellulose, hemicellulose and lignin (see **section 1.1**), crude glycerols can be seen as mixtures of glycerol and non-glycerol organic matter (MONG). In crude glycerol, two types of impurities can be found: raw material-specific and process-specific impurities. MONG is the major raw material-specific impurities and is mainly made of fatty matter. This fatty matter can contain unprocessed oil or fat, diglycerides (DAGs), monoglycerides (MAGs), free fatty acids (FFAs) for the fat-splitting derived waste or fatty methyl ester (FAMES) for biodiesel-derived waste. An examination of the total fatty acid composition revealed that the fatty acid profile of crude glycerol is consistent with that found in the initial raw material (Hu et al., 2012) explaining why the MONG is a raw material-specific impurity. Other raw material-specific impurities include elements that depend on the type and origin of raw-material (see **section 1.2.1.a**). These contaminants may include sulfur compounds, proteins, nitrogenous compounds, aldehydes, ketones, oxidized fatty matters, fermentation side-products (Jungermann & Sontag, 1991), sodium chloride especially (when waste cooking oils are used) (Gao et al., 2016), carbohydrates (Ayoub & Abdullah, 2012) and polyphenols (Pagliaro, 2017).

While process-specific impurities are glycerol degradation or polymerization products for the fat splitting process or alcohol and traces of catalyst for the biodiesel production process. The thermal degradation of glycerol produces acetaldehyde, acrolein, allyl alcohol and other products (Qadariyah et al., 2011), while glycerol polymerization results

in polyglycerol formation (Ardi et al., 2015). Methanol is the preferred alcohol for biodiesel-producing transesterifications but alcohol containing 1 to 4 carbons can be used for this process (Knothe *et al.*, 2010). So, the alcohol found in crude glycerol depends on the alcohol used as reactant but the most common is methanol. Biodiesel formation is either base-catalysed (e.g. with sodium hydroxide, potassium hydroxide or sodium alkoxide) or acid-catalysed (e.g., sulfuric, hydrochloric or phosphoric acid). A small amount of elemental impurities such as Na, Ca, K, Mg, P, S, and N are found in crude glycerol derived from biodiesel production using base catalysts (Thompson and He, 2006). Metal impurities, except sodium, are present at concentrations between 4–163 ppm (Ayoub and Abdullah, 2012). After a base-catalysed process, the sodium content of crude glycerol can exceed 1% (Ayoub and Abdullah, 2012) which can be problematic because the cation tends to form soaps (i.e. metallic salt of fatty acids) with free fatty acids. The soaps are then found in the crude glycerol. The use of acid catalysis prevents soap formation.

1.2.2.c - Crude glycerol compatibility with biological applications

Aside from impurities glycerol itself has been described to have adverse effects on microbial growth. With concentrations reaching up to 50–60% of biodiesel-crude glycerol, the effects of glycerol are consequences of its high concentrations. High glycerol concentrations result in substrate inhibition during microbial growth (Muniraj et al., 2015), which is often attributed to osmotic shock (Raimondi et al., 2014) and redox imbalance in cells (Mota et al., 2017). Although it should be noted that microbial cells are not equal in the face of these harmful events. For instance, osmotolerant organisms will respond better to high osmotic pressure. While the redox imbalance caused by glycerol utilization is such that certain yeast species (i.e. *Saccharomyces*) cannot even ferment anaerobically glycerol whereas most bacteria do not suffer from the same limitations (Claret da Cunha et al., 2019). Nevertheless, glycerol is a metabolically interesting carbon source because (1) unlike high glucose concentration it doesn't trigger

Crabtree effect in yeasts (Klein et al., 2016) (2) its high degree of reduction (greater than sugars) is advantageous for the production of reduced metabolites by bacterial anaerobic fermentation (Garlapati et al., 2015; Yazdani & Gonzalez, 2007) (3) as a small neutral molecule it might be able to passively diffuse through certain membranes (Sutherland *et al.*, 1997; Blotz and Stulke, 2017; Klein *et al.*, 2017; Westbrook *et al.*, 2018) (4) if not channeled into glycolysis, it can be directly diverted toward primary and secondary metabolites synthesis (

Figure 1-5).

As for the effect of impurities, both detrimental and beneficial effects have been described. The effects are often concentration-dependent, strain-specific, and poorly characterized at a molecular level. Even though crude glycerol components are discussed individually here, microbial growth behavior and productivity in crude glycerol are likely manifestations of combined effects due to all components present. Methanol at a concentration above 1% has been described to affect the fluidity of bacterial membranes (Venkataramanan et al., 2012). Differences in membrane compositions might explain the

differences in tolerance to methanol. Strains' different response to methanol are not likely limited to morphological differences and might also arise from physiological characteristics such as adaptation mechanisms or metabolic abilities. For instance, some bacteria can adjust the composition of their plasma membrane in response to alcoholic stress (Venkataramanan et al., 2012). And methylotrophic yeasts can utilize methanol as a carbon source (Yurimoto & Sakai, 2019).

Given their lipophilic nature, MONG can also interact with the plasma membrane (Venkataramanan et al., 2012). Their other reported or presumed effects include (1) their microbial utilization as carbon source (Papanikolaou et al., 2006) (2) their action as surfactants (Gao et al., 2016) (3) their incorporation into the cellular fatty acid pool (Samul et al., 2014; Valerio et al., 2015). The outcome depends on the nature of the fatty matter. Some strains can consume FFAs but cannot process glycerides because they lack the suitable lipase(s) while others can utilize both FFAs and glycerides (Solaiman et al., 2006). Non-assimilated fatty matter can positively or negatively impact cell membranes. Fatty acids interfere with the membrane structure by tail/tail interactions with the membrane fatty acids (Muranushi et al., 1981). And it was hypothesized that unsaturated fatty acids could hinder the diffusion of nutrients (Venkataramanan et al., 2012). FAME, soap, MAG, and DAG can function as surfactants to emulsify the medium and improve nutrient intake by increasing cell membrane permeability (Gao *et al.*, 2016). But soap presence can also favour bubble formation reducing thus oxygen transfer rate in a growth-limiting way (Sivasankaran et al., 2018).

The behaviour of the fatty matter contained in crude glycerol can be affected by the pH and so does microbial growth. Depending on the producing process, crude glycerol can have pHs as high as 10 or as low as 4. Fortunately, natural microbial diversity is such

that bacteria able to grow at pH 4 when other species can grow at pH 11 while the average bacteria will prefer a neutral to slightly alkaline pH unlike the typical fungi that will thrive at slightly acidic pHs (Vieira, 2013). Finally, most minerals are found only as traces in biodiesel crude glycerol (Ayoub & Abdullah, 2012) but their impact should not be underestimated. Indeed, the effects of minerals on some organisms can make the use of waste as feedstock for the bioproduction of certain metabolite impossible (Cavallo et al., 2017; Willke & Vorlop, 2001).

Given the incredible microbial diversity and crude glycerol compositional variability, there might be as many studies describing the absence of effect or positive impacts as there are studies highlighting side-effects of crude glycerols. This observation makes it hard to evaluate to which extent impurities are limiting the development of bioprocess based on crude glycerol. Especially since, as exposed in **Figure 1-4**, a few simple tricks such as careful strain selection can help researchers bypass the nuisance sometimes caused by impurities.

1.2.2.d - Improving crude glycerol usability

A common strategy is to dilute the crude glycerol with synthetic medium (Ito *et al.*, 2005). Not only this strategy eliminates glycerol substrate inhibition, but it also provides additional nutrients to support microbial growth or pH adjustment of the feedstock. Dilution is the only treatment that modulates all components of crude glycerol. For crude glycerol containing methanol, media sterilization *via* autoclaving before cultivation would evaporate the bulk of the methanol present (Athalye et al., 2009). While pH adjustment may seem like a straightforward answer to pH-related limitations, it can lead to other undesirable issues. For example, lowering pH is one way to split soap and extract fatty acids from the glycerol phase, but this pH adjustment also increases total salt content (Moon *et al.*, 2010). In addition to pH adjustment, fatty matters in MONG can be reduced via solvent extraction. But the extraction solvent must be carefully

selected to avoid decreasing the usability of the waste (Moon *et al.*, 2010). The use of fed-batch cultivation has been reported as an efficient solution to circumvent glycerol substrate inhibition as it allows better control of glycerol and nutrients concentrations (Samul *et al.*, 2014; Muniraj *et al.*, 2015). Fed-batch typically enables higher cell density and metabolite production than in batch cultivation (Beopoulos *et al.*, 2009).

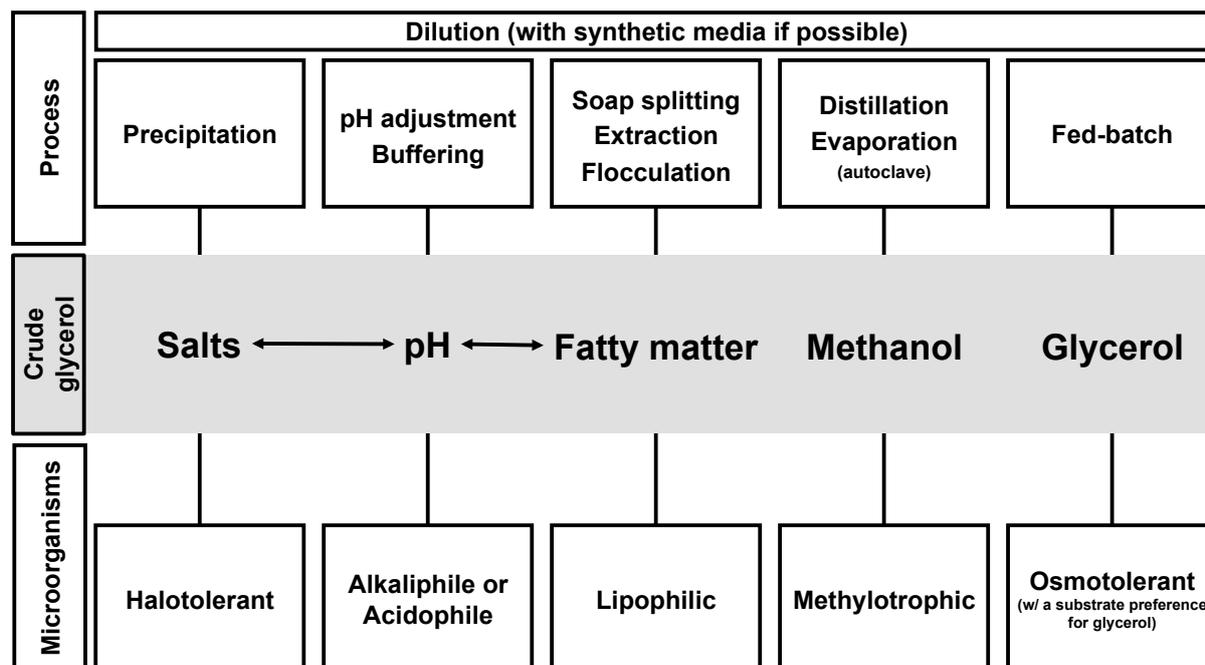


Figure 1-4 Overview of simple strategies used to overcome crude glycerol composition.

The grey section identifies crude glycerol properties that potentially affect microbial growth. The top part of the diagram shows physical modifications that can improve the compatibility of crude glycerol to microbial growth. The bottom part of the diagram shows the class of microorganisms with specific tolerance against the identified property.

Overcoming substrate inhibition by adding another carbon source or waste stream is also frequently adopted. There have been reports in the literature of crude glycerol being successfully blended with glucose (Mantzouridou *et al.*, 2008; Chen *et al.*, 2012) and other wastes such as whey (Ilic *et al.*, 2016), tomato waste (Fakas *et al.*, 2008), corn stover hydrolysate (Xin *et al.*, 2016) and dairy cattle manure (Simm *et al.*, 2017). However, the selection of a second carbon source is crucial as some carbon sources can repress glycerol assimilation (Wang *et al.*, 2019). Another advantage of combining crude

glycerol with other wastes is that the latter could provide additional nutrients. For instance, nutrients like vitamin B12 stimulate central metabolism and glycerol utilization as some enzymes, for example, glycerol dehydratase (Gly DHt) involved in 3-hydroxypropanal (3-HPA) formation (

Figure 1-5), are B12-dependent (Westbrook et al., 2018). That being said, nutrient supplementation is not always the best strategy. Sometimes nutrient starvation is preferable to stimulate the accumulation of the metabolite of interest (Możejko-Ciesielska & Pokoj, 2018).

If none of the above strategies is sufficient, one can also resort to strain engineering to increase the host's productivity and tolerance to the waste (Chen & Liu, 2016). Before that, a careful strain selection might alleviate the need for engineering (Boundy-Mills, 2012). But selecting a host with a good affinity for the waste might limit the array of products accessible if relying only on the natural abilities of said host. Which takes us back to strain enhancement to tailor the host metabolite production. If strain engineering

is not conceivable, given the current prices, the use of refined glycerol although less economically attractive is still an option. Indeed, unlike lignocellulosic wastes, glycerol refining facilities already exist and are operational.

The various purification strategies used to convert raw glycerol to pure glycerol usually combine several methods and are resource-intensive (Tan *et al.*, 2013). New developments involving membrane-based technology (ultrafiltration and reverse osmosis) are currently under evaluation as less energy-consuming alternatives to deionization, evaporation, and distillation (Jungermann and Sonntag, 1991; Pagliaro and Rossi, 2008). However, the cost and the lifetime of membranes are limiting factors (Mota *et al.*, 2017). In 2011, the lowest cost of glycerol purification achievable for purity of up to 98% (by the combination of neutralization, centrifugation, evaporation, and column distillation) was estimated at \$0.15 per kg (Posada *et al.*, 2011). Distillation generates a solid waste called glycerol pitch (Hazimah *et al.*, 2003) or glycerol residue (Yong *et al.*, 2001) that contains glycerides, salts, soaps, fatty acids, polyglycerols, ashes and glycerol (Jungermann and Sonntag, 1991; Mota *et al.*, 2017). This glycerol pitch is classified as hazardous and poisonous (Irvan *et al.*, 2018) and must be carefully disposed of. The cost of which increases the purification cost beyond \$0.15 per kg. On the other hand, some isolated contaminants can be recycled to reduce the process cost. For instance, methanol recycling is well established in the biodiesel industry (Isahak *et al.*, 2015) and unprocessed oils and some catalytic salts can also be recovered (Jungermann and Sonntag, 1991). These opportunities for recycling contribute to cost mitigation and make refined glycerol more affordable.

Unfortunately, refined glycerol is still under-represented as feedstock for industrial biomanufacturing (Garlapati *et al.*, 2015). Yazdani and Gonzalez (2007) even compared facilities needed to produce ethanol from corn (current raw material) or refined glycerol.

They conclude that not only the facilities required to work with glycerol are less complex but operational costs would be almost 40% lower than the ones currently observed with corn. Yet corn usage is so established that it leaves no room for other alternatives. This is enough to conclude that albeit inconvenient, crude glycerol composition is not the main impediment to the spread of crude glycerol as feedstock for biomanufacturing.

1.2.1 - Remaining challenges and prospects

While the prices of glycerol remain volatile, glycerol supply will continue to exceed its demand. This trend poses significant pressure on glycerol disposal, resulting in additional waste management costs and environmental concerns. There is a need for innovations in glycerol purification and valorization. Chemical conversions often require high purity glycerol whereas microorganisms can uptake glycerol from its crude form potentially saving glycerol purification cost and time. Microbial utilization of glycerol is therefore a promising course to sustainable glycerol valorization. Yet, the advent of glycerol-based

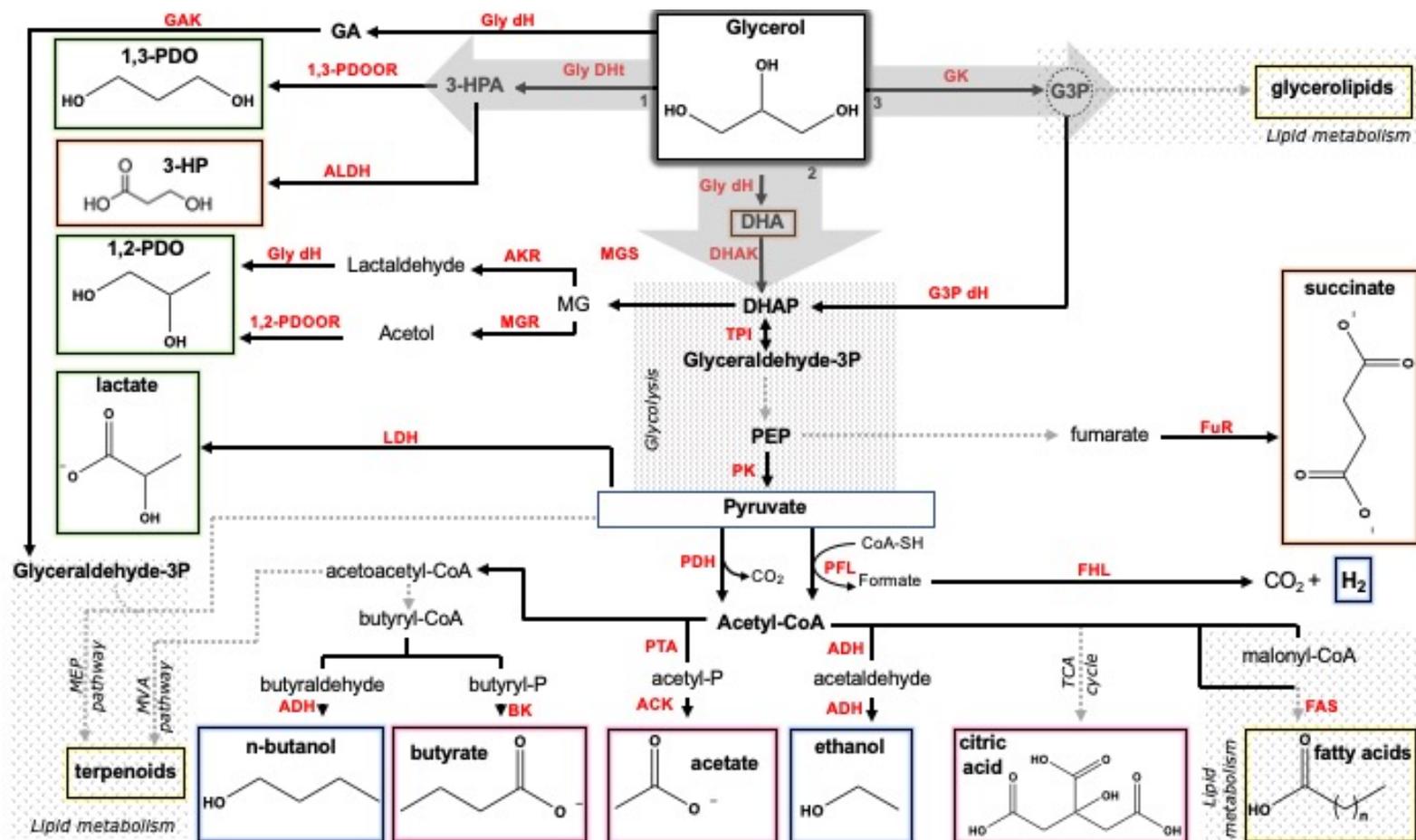


Figure 1-5 Overview of microbial glycerol catabolism

The major glycerol catabolism pathways in microbes are highlighted by thick grey arrows numbered 1 to 3. Two oxidative pathways are possible and start either *via* glycerol phosphorylation (Grey arrow n°3) or dehydrogenation (Grey arrow n°2) to produce DHAP. Alternatively, some microbes can also convert glycerol *via* a reductive pathway (Grey arrow n°1) to produce 3-HPA. Grey dashed arrows are standing for unrepresented steps. Products are color-coded by main applications (see **Table 1-1**). List of metabolite abbreviations: **1,2-PDO**, 1,2-propanediol; **1,3-PDO**, 1,3-propanediol; **3-HP**, 3-hydroxypropionic acid; **3-HPA**, 3-hydroxypropanal; **CoA-SH**, coenzyme A; **DHA**, dihydroxyacetone; **DHAP**, dihydroxyacetone phosphate; **G3P**, glycerol-3-phosphate; **GA**, glyceraldehyde; **MG**, methylglyoxal; **P**, phosphate (inorganic); **PEP**, phosphoenolpyruvate. List of enzymes abbreviations: **1,2-PDOOR**, 1,2-propanediol oxidoreductase; **1,3-PDOOR**, 1,3-propanediol oxidoreductase; **ADH**, bifunctional alcohol/aldehyde dehydrogenase; **ACK**, acetate kinase; **AKR**, aldo-keto reductase; **ALDH**, α -aldehyde dehydrogenase; **BK**, butyrate kinase; **DHAK**, dihydroxyacetone kinase; **FAS**, fatty acid synthase; **FHL**, formate hydrogen lyase; **FuR**, fumarate reductase; **G3P dH**, glycerol-3-phosphate dehydrogenase; **GAK**, glyceraldehyde kinase; **GK**, glycerol kinase; **Gly dH**, glycerol dehydrogenase; **Gly Dht**, glycerol dehydratase; **LDH**, lactate dehydrogenase; **MGR**, methylglyoxal reductase; **PDH**, pyruvate dehydrogenase; **PFL**, pyruvate formate lyase; **PK**, pyruvate kinase; **PTA**, phosphotransacetylase; **TPI**, triose phosphate isomerase.

biomanufacturing technologies will depend on (1) the efficiency of optimization endeavors (i.e. enhancement of microbial tolerance to stresses, nutritional requirements, productivity and extraction & purification techniques) (2) the creation of new opportunities for glycerol. At the moment, the optimization efforts are progressing slowly. Because despite the considerable research focusing on the biological valorization of crude glycerol, optimization strategies are often strain, glycerol and product-specific. Meaning that almost each host, glycerol type and target products combination will require specific developments. Developments sometimes hindered by the lack of appropriate tools or knowledge of the host metabolism (Yazdani & Gonzalez, 2007). To circumvent that limitation, part of the research force is trying to habilitate well-documented and amenable organisms with long industrial histories such as *E. coli* or *S. cerevisiae*. This type of approach has the advantage of refocusing the effort but will be limited by the physiological characteristics of the host because for now there is only a limited number of molecules accessible from these traditional hosts. Of course, the race to expand the range of molecules accessible from these hosts' glycerol utilization is raging. But in this race, one should take time to identify the best opportunities because even if sustainability has gained increasing importance, profitability remains a priority when developing new applications. As stated by Bauer and Hulteberg (2013): “according to standard economic theory, the process able to add the most value to the feedstock at the lowest cost will dominate the market”. And currently, glycerol valorization market is dominated by chemical conversion to epichlorohydrin, and propylene glycol despite the need for prior glycerol refining (Pagliaro, 2017). Not only the current portfolio of biomolecules accessible from glycerol overlaps with the products accessible from sugars but those products are also competing with synthetic products from the chemical industry. Hence, product selection will be determinant for the future of crude glycerol as feedstock for biomanufacturing. Among the type of biomolecules accessible from glycerol, lipids could offer good opportunities for crude glycerol valorization. Given their

diverse chemical nature lipids can be used as building blocks, biofuel, ingredients in food, pharmaceuticals, nutraceuticals, cosmetics, personal and home care products (Tao, 2007). The lipid market already valued at \$6.5 billion (Beroeinc, 2018) is expected to continue to grow. The contemporary quest for health, food, energy security and the concern for sustainability will open new opportunities. Plant oils are increasingly criticized for their competition with food crops and their contribution to land-use changes, deforestation, and the greenhouse effect (Wicke et al., 2011). Animal fats use can be offensive and is incompatible with certain lifestyles. Furthermore, both vegetable and animal lipids are associated with contamination risks (Jacobs et al., 2004; Niu et al., 2021). The opportunities will not be limited to replacement or supplementation of current sources (i.e. plant oils and animal fats), there will also be room for the production of unusual lipids difficultly accessible from conventional sources. In the same way, these lipid sourcing problematics have already opened doors to current lipid biomanufacturing processes. Industrial microbial lipid productions already exist but the technology is not so widespread that it would prevent the establishment of glycerol as feedstock like what is observed for ethanol production. Moreover, glycerol is a particularly advantageous feedstock for lipid production because it can directly be channeled into lipid synthesis (

Figure 1-5).

1.3 - Oleaginous microorganisms: rising workhorses for the production of microbial oils

1.3.1 - The oleaginous realm

Given their ability to accumulate up to 80 % of their dry weight as lipids (Beopoulos et al., 2008), oleaginous microorganisms offer a promising renewable alternative to current lipid sources. But any microorganism accumulating at least 20% of its dry weight as lipids will be considered oleaginous and this very definition leads to important disparities within the oleaginous realm turning microorganism selection into a crucial stake. Oleaginous microorganisms are found among cyanobacteria, bacteria, filamentous fungi, microalgae, and yeasts. Yet nowadays mainly fungi (i.e. yeast and filamentous fungi) and microalgae are under investigation for lipid production from crude glycerol (**Table 1-1**). Due to lower lipid contents, oleaginous bacteria were always less investigated than the other oleaginous microorganisms albeit amenable, characterized by especially short duplication times and versatile carbon utilization. Filamentous fungi's main drawbacks are their low growth rates and their current limited genetic tractability. Two characteristics shared with microalgae and cyanobacteria (Yu et al., 2014). Microalgae, despite a few demonstrations of waste feedstock use (Zhang et al., 2013), seem to have

a more limited known carbon utilization than yeasts. Yeasts are moreover described to exhibit short duplication times, good robustness as well as relatively good amenability even though the improvement of certain non-conventional yeasts has sometimes been hampered by the lack of appropriate tools for genetic manipulation (Adrio, 2017).

Several approaches combining species, so they can counterbalance one another weaknesses, have also been described (Magdouli et al., 2016). Such approaches are of course contingent on the identification of a suitable consortium and on the availability of appropriate methods to study the interactions between microorganisms. Still, microalgae were successfully combined with bacteria (Kazamia et al., 2012), yeasts (Cai et al., 2007; Levering et al., 2015; Santos et al., 2011, 2013) and even filamentous fungi (Wrede et al., 2014). The latter association is especially interesting because it was designed to improve microalgae harvesting expensive process by trapping them into fungi “self-pellets” (Wrede et al., 2014). When on the other side, yeasts are easy to harvest due to their sizes and flocculation abilities (Bekatorou et al., 2006). They are also able to accumulate up to 70% of their dry cell weight as lipid (Beopoulos et al., 2008) when grown in simple facilities using cheap feedstocks (Adrio, 2017). And as yeasts have a long industrial history in the food sector some strains are even already generally recognized as safe and commonly approved by the regulatory authorities (Beopoulos et al., 2012). Despite that long industrial tradition, a better understanding of transport and regulation mechanisms in yeast especially the one involved in lipid accumulation is still needed to harness the full potential of those oleaginous microorganisms (Yu et al., 2014).

1.3.2 - Industrial uses of oleaginous microorganisms: past, present and outlook

The first industrial production process of microbial oil was described as early as 1985. It was the production of a γ -linolenic acid (GLA) rich oil commercialized as “Oil of Javanicus” and produced by the filamentous fungus *Mucor circinelloides* industrially exploited from 1985 to 1990 (J. & E. Sturge, UK). The “Oil of Javanicus” was

unfortunately in direct competition with a plant-derived alternative. The context was initially favorable to the “Oil of Javanicus” as only one other source was known at the time but, the commercialization of the “Oil of Javanicus” triggered a price escalation and the microbial product struggled to remain profitable especially with the arrival of a third much cheaper plant-derived alternative (Cohen & Ratledge, 2015).

The production of cocoa butter equivalent (CBE) by the oleaginous yeast *Cutaneotrichosporon curvatus* (formerly *Apiotrichum curvatum*), set in 1992 (Davies, New-Zealand) was also deemed uneconomic (uncertain market uptake) even though lactose, a cheap by-product of the cheese creamery industry, was used as feedstock. The two programs were run respectively in 220 m³ (“Oil of Javanicus”) and 250 m³ (CBE) fermenters and despite their deplorable termination, they demonstrated that microbial oils are safe for human consumption and led to a better management of cultivation parameters and extraction methods. The production of the “Oil of Javanicus” proved for instance, that existing fermentation techniques could be transferred to oleaginous microorganisms for oil production. Indeed, the fermenters used for that purpose were initially designed for citric acid production using *Aspergillus niger* (Ratledge, 2013). As for the yeast CBE, it underlined the importance of culture parameters as levels of stearic acid in *C. curvatus* were increased by maintaining very low aeration rates during lipid accumulation to minimize desaturases activities (Cohen & Ratledge, 2015). These first attempts albeit unsuccessful paved the way for the advent of microbial polyunsaturated fatty acid (PUFAs) also known as ω -3 and ω -6 fatty acids.

The first viable commercialization of PUFAs started a short while after the abortion of yeast CBE with the launch of DHASCO™ quickly followed by ARASCO (Martek bioscience now part of DSM) respectively a docosahexaenoic acid (DH-A) rich single cell oil (SCO) and an arachidonic acid (ARA) rich SCO. Both oils were proposed as alternatives to the only other known source of PUFAs: fish oil. Fish oil was a depleting resource (jeopardized species), requiring deodorization to make it more suitable for

human consumption and associated with a high risk of contamination by heavy metals and persistent organic pollutants (POPs) that are quite expensive to remove. Produced by the heterotroph microalgae *Crythecodinium cohnii*, DHASCO™ is currently sold as an adult nutritional supplement as well as an additive for infant formula when used in (1:2) combination with ARASCO produced by the filamentous fungus *Mortierella alpine* (Ratledge, 2013). Up to now, ARASCO™ and DHASCO™ remain the best commercial successes when it comes to industrial lipid production by oleaginous microorganisms and generate enough profits to sustain their production although a lot of PUFAs are now commercialized.

Other commercial PUFAs bioprocesses include but are not limited to (Ratledge, 2013): CABIO oil (an ARA-rich oil also produced by *Mortierella alpine* - Cargill), DHASCO-S (a combination of DH-A and docosapentaenoic acid used for animal nutrition and produced by algae from the *Schizochytrium sp.* - Martek/DSM), DHAAid™ (another DHA rich oil produced by algae from the *Ulkenia sp.* - Lonza), EPASCO (an eicosapentaenoic acid -EPA- rich oil produced by the yeast *Yarrowia lipolytica* - Du Pont), EicoOil (a combination of EPA and other PUFAs produce by an undisclosed microalgae – Qualitas Health).

Following the success of PUFAs production, other microbial nutraceuticals and food additives were commercialized and especially the astaxanthin (BioAstin™ - Cyanotech corporation), a carotenoid produced by *Haematococcus pluvialis* which is the highest value product that can be generated by this phototroph microalga (Panis & Carreon, 2016) or the γ -decalactone, a peach like aromatic compound derived from ricinoleic acid using yeast β -oxidation commercialized by Lesaffre.

Examples of microbial nutraceuticals and food additives production are various (Leman, 1997; Muniraj et al., 2015) but only a few microbial oils are engineered for other

applications. Mention can be made, for example, of the American company Solazyme producing oil for biodiesel with microalgae of the *Chorella sp.* in heterotroph mode using glucose as feedstock. In 2009, British Petroleum (BP) even announced a partnership with Martek Biosciences (producers of DHASCO™ and ARASCO) (Scheyder, 2009) and more than 10 years later, the outcome of this collaboration is still unclear suggesting that knowledge transfer from microbial PUFA production to viable biodiesel production might not be so straightforward or that the market is not ready yet for such breakthrough. Probably for this reason, Solazyme extended the application range of its oils to food materials and cosmetic functional ingredients.

With a different range of applications (precursors in the production of nylon, resins and adhesives) but also good biofuel candidates, microbial manufacturing of α,ω -dicarboxylic acid (DCA) was investigated in oleaginous yeasts such as *Candida cloacae*, *Candida tropicalis* and *Yarrowia lipolytica*. The exploration of more environmentally friendly solutions was prompted by the high energetical demand of DCA chemical synthesis (Huf et al., 2011). This led to a microbial DCA production process based on *C. tropicalis* developed by Nippon Mining Company and now exploited by Cathay Biotechnology of Shanghai and Cognis (Beopoulos et al., 2011). Several other companies offer bioproduced saturated DCA with chain lengths varying from eleven to eighteen carbons. If the first biomanufacturers of DCA focused on long-chain DCAs, it is because the selectivity of the chemical reaction was lower for long-chain DCA than for small-chain DCA, leaving an opportunity for the microbial process establishment (Huf et al., 2011). DCA with different properties could also gain importance in the coming years. For instance, octadecenedioic acid produced from oleic acid has attracted a lot of attention (Funk, 2020). Current industrial successes will likely inspire related productions in the coming years.

Also related to the DCA family, can be mentioned the impressive ascent of bio-succinic acid (C₄) which is now one of the top value-added products derived from biomass (Putri

et al., 2020). The success of bio-succinic acid is due to unfavourable oil market and high potential as bio-based building block. With the current depletion of fossil resources, the new bio-based building blocks are expected to progressively replace petroleum-derived building blocks which will open doors for lipid with interesting chemical functionalities. Short-chain DCA and their derivatives make great building-block and so in addition to succinic acid: itaconic acid (C₅), adipic acid (C₆) and 2,5 furandicarboxylic (C₆) productions are also considered promising ventures (Werpy & Petersen, 2004). It should be noted that short-chain DCAs hardly qualifies as lipids. Since lipids are defined as compounds relatively water-insoluble constituting the fraction of cells extractable with organic solvents (Gurr et al., 2008).

Promising building blocks more in line with that definition include epoxy & hydroxy fatty acids, halogenated fatty acids, cyanolipids, acetylenic, allenic and cyclic fatty acids were reviewed by Avato and Tava (2021). Other rare lipids with commercial potential are: odd-chain fatty acids (Park et al., 2020), branched fatty acids (BFA)(Blitzblau et al., 2021), very long-chain fatty acids (VLCFA) (Avato & Tava, 2021) and acetyl-triacylglycerols (Aznar-Moreno & Durrett, 2017). Emblematic examples of these lipid classes would be: pentadecanoic acid (C₁₅) (odd-chain FA), 10-methylpalmitic acid (BFA), erucic acid (C₂₂) (VLCFA) and glycerol triacetate (acetyl-triacylglycerol). The major success of PUFAs biomanufacturing might inspire new processes. So far, PUFAs bioproductions are limited to non-conjugated isomers such as linoleic acid, γ -linolenic acid, arachidonic acid, decahexenoic acid and eicosapentanoic acid. But the interest in conjugated isomers is growing (Avato & Tava, 2021; Ledesma-Amaro & Nicaud, 2016). As for astaxanthin and γ -decalactone they could respectively open the way for the production of complex terpenoids such as the one described by Gross & König (2006) and other lipid-derived aromas (e.g. hexanal) (Ledesma-Amaro & Nicaud, 2016). And the few years down the road, the production of artificial lipid (i.e. not found in nature)

might even become relevant. Indeed, artificial lipids might provide even more tailored answers to industrial challenges.

1.3.3 - Lipid metabolism in yeasts: from basic concepts to oleaginous metabolism

1.3.3.a - Getting started with lipid metabolism in yeasts

Yeast lipids can be derived either from geranyl pyrophosphate (like steroids and carotenoids) or from fatty acyl-CoA (fatty acid, fatty alcohols, alkanes and alkenes) albeit some are a combination of both sources like steryl esters (SE) that contains an acyl chain and a sterol moiety. As commonly used fats and oils from plants and animals belong to the fatty acyl-CoA category (Sitepu et al., 2014a), this section will focus on the formation of such derivatives. Knowing that once constituted, cell fatty acyl-pool can not only be used for the formation of complex lipids but also be degraded for energy production or stored as triglycerides. Fatty acyl-CoA can be generated through 3 paths (Beopoulos et al., 2011; Klug & Daum, 2014):

- the *de novo* synthesis from acetyl-CoA
- the *ex novo* accumulation from external sources of fatty acids or alkanes
- the lipid turnover corresponding to the degradation of complex lipids and the delipidation of proteins

The initiation of fatty acid *de novo* synthesis is a cytoplasmic process. It starts with an initial biosynthetic unit: the acetyl-CoA. Acetyl-CoA is condensed with a malonyl-CoA (the elongation unit) to form an acyl chain that will be elongated of 2 carbons (supplied by the elongation units) with the consumption of 2 NADPH at each cycle of a reaction sequence accurately orchestrated by a multi-activities cytoplasmic complex called fatty acid synthetase (FAS) (**Figure 1-6**).

Cytoplasmic acetyl-CoA originates from the pyruvate produced by central metabolism. The pyruvate decarboxylase² is responsible for the transformation of pyruvate into acetaldehyde. Acetaldehyde is then oxidized into acetate by an aldehyde dehydrogenase (ALDH). So that a peroxisomal acetyl-CoA synthetase can take over the acetate to form acetyl-CoA that is subsequently transferred back to the cytoplasm. This pathway is common to all the yeasts, but most oleaginous yeast demonstrated a complementary acetyl-CoA formation path that will be discussed in **section 1.3.3.b**. As for malonyl-CoA, it originates from the carboxylation of an acetyl-CoA by the acetyl-CoA carboxylase (ACC) using a bicarbonate anion in an ATP-dependent reaction. NADPH supply is ensured by the NADPH-dependent isocitrate dehydrogenase (in mitochondria), the pentose phosphate pathway (in the cytoplasm), and the malic enzyme that generates pyruvate from malate using NADP. It should be noted that some malic enzymes are NAD-dependent and so cannot contribute to NADPH production. The latter enzyme is cytoplasmic but has also been described in mitochondria (Hoja et al., 2004).

Fatty acid synthesis is a feature common to all living organisms. Fatty acid synthases found in plants and bacteria are called type II FAS (FAS II), those FAS catalyse the

² Must not be confused with the pyruvate dehydrogenase responsible for mitochondrial acetyl-CoA synthesis.

formation of a fatty acyl chain attached to an acyl carrier protein (ACP) rather than the formation of a fatty acyl-CoA like synthesized by type I FAS (FAS I). In both cases the fatty acyl is in an activated state due to the presence of a thioester bond between the fatty acyl chain and either the ACP or the coenzyme A (CoA). FAS II are multi-enzymatic complexes and FAS I found in mammals and fungi are dimers whose subunits are encoded by two genes (Fas1 and 2 not to be confused with the FAS types I and II) (Beopoulos et al., 2008; Fillet & Adrio, 2016). Since this major difference was limiting a lot of pathway engineering endeavours, it should be noted that some bacteria (i.e. *Mycobacteria*, *Corynebacteria* or *Nocardia*) can express both types of FAS (Kosaric & Sukan, 2010). Moreover, yeasts have in addition to the cytoplasmic pathway, a mitochondrial pathway catalysed by a FAS II (Hiltunen et al., 2009).

Fatty acyl-CoA synthesized *via* the *de novo* pathway in yeasts, will at this point contain 14 to 16 carbons depending on the FAS I specificity. The length specificity of the native FAS enzyme (and subsequent enzymes) has also been described as a determining factor in metabolic engineering attempts (Fillet & Adrio, 2016). When it comes to *ex novo* fatty acid synthesis (i.e. fatty acid synthesis from hydrophobic compounds), the length specificity is determined by fatty acid transporter and activation enzymes. Depending on the nature of the hydrophobic starting material, the *ex novo* pathway can be mediated through multistep mitochondrial and peroxisomal oxidations involving cytochromes P450-dependent alkane monooxygenase systems, fatty alcohol oxidases, alcohol dehydrogenases or fatty aldehyde dehydrogenases. The aforementioned DCAs (see **section 1.3.2**) are products of this *ex novo* path and are generated through ω -oxidation of alkanes (Mishra et al., 2018).

Fatty acyl-CoA, from the *de novo* pathway or the *ex novo* pathway can join the endoplasmic reticulum (ER). In the ER elongases will add more carbons to the chains and desaturases will generate unsaturations (**Figure 1-6**). Like FAS activity, desaturases activities are NADPH-dependent and elongases use malonyl-CoA as

elongation units. The NADPH equivalents needed for the desaturation are provided by ER malic enzyme (Garay et al., 2014). The most common desaturases in yeast are the $\Delta 9$ and $\Delta 12$ desaturases but some species also express the $\Delta 15$ and $\Delta 6$ essential to produce, the precious γ -linolenic acid (GLA), the precursor of all commercial PUFAs (mentioned in **section 1.3.2**).

The modified acyl-CoA can then leave the ER to increase the cytoplasmic pool of fatty acyl-CoA or be allocated to either polar or neutral lipid formation. The term “neutral lipid” refers mainly TAG and SE even if formally some sphingolipids (especially glycosphingolipids and ceramides) and hydrocarbons (alkanes, terpenoids, steroids and carotenoids) could be included in this chemical category. Similarly, the term “polar lipid” is mostly used to designate glycerophospholipids when it should be extended to all charged lipids: phospholipids (including cardiolipins, the remaining sphingolipids namely sphingomyelins and other phosphosphingolipids) and free fatty acids. But further nomenclature considerations would be wasted because 80 to 90% of yeast lipids are stored as SE and/or TAG (Sitepu et al., 2014a), their formation is even known as the “neutral lipid storage pathway” (Beopoulos et al., 2011). The functions of other lipids are mainly limited to signal transduction and/or structural roles.

In the ER, the modified acyl-CoAs follow the Kennedy pathway (Garay et al., 2014) in which acyl groups are sequentially transferred from acyl-CoAs to a glycerol-3-phosphate (G3P) resulting consecutively in the formation of a lysophosphatidic acid (LP). LP then transformed into phosphatic acid (PA), that undergoes a dephosphorylation catalyzed by the phosphatidate phosphatase (PAP) to yield diacylglycerol (DAG). DAGs are ultimately also acylated this time through a reaction catalyzed by an acyl-CoA:diacylglycerol acyltransferase (DGAT). This pathway release TAGs that are then stored in lipids droplets. The PA transiently formed is involved in the regulation of genes encoding ACC and FAS expression and the precursor of the glycerophospholipids as its phosphate group instead of being dephosphorylated can be replaced by a cytidine

diphosphate. Interestingly, cytidine diphosphate DAG (CDP-DAG) concentration is closely linked to DAG concentration. Indeed, CDP-DAG hydrolysis catalyzed by the phospholipase C can generate DAG that can then be re-directed towards triacylglycerol formation when DAG concentration is lower than CDP-DAG concentration (Garay et al., 2014). As for the formation of steryl esters, it is catalyzed by two acyltransferases called acyl-CoA cholesterol acyltransferase (ACAT). Despite, the enzyme name, it can be noted that yeast major sterol is ergosterol and not cholesterol (Sitepu et al., 2014).

When needed, the content of lipid droplets can be mobilized through the action of lipases to release free fatty acids (FFA). FFAs can be further activated by a thioesterification catalysed by a fatty acyl-CoA synthetase³ (FAA). Some medium-chain fatty acids can directly penetrate peroxisomes. So, depending on the chain length, the FAA taking over the FFAs will be either cytoplasmic or peroxisomal. The cytoplasmic pool of fatty acyl-CoAs can be used for the synthesis of more complex lipids or sent to peroxisomes to be degraded in acetyl-CoA for energy production under carbon starvation conditions (Adrio, 2017; Klug & Daum, 2014).

Some steps of lipid metabolism are tightly regulated and dependent on both internal and external factors. Ions contribution to lipid metabolism was for instance already mentioned throughout this section (i.e. ACL ammonium-dependent, bicarbonate as ACC substrate, PAP magnesium-dependent) but no mention of vitamins was yet done. Vitamins could really impact both production costs and *de novo* synthesis of lipids. Indeed, certain vitamins like biotin (vitamin B8) or pantothenic acid (vitamin B5) are directly involved in lipid metabolism (Sitepu et al., 2014). One thing is sure, yeast lipid production might not be affected by venue or seasonal changes but both yield and oil composition depend on a lot of factors and especially the incubation temperature, the aeration, the medium pH, salinity, mineral composition (including nitrogen, sulfur

³ Not to be confused with the fatty acid synthase (FAS)

phosphorous source) (Beltran et al., 2008; Klug & Daum, 2014; Li et al., 2008; Sitepu et al., 2014; Yoon & Rhee, 1983) so all these parameters should be considered when designing a yeast-based lipid production process.

1.3.3.b - Well-documented specificities of oleaginous yeasts

Oleaginous yeasts are found across all the clades of yeast order, in both *Ascomycota* and *Basidiomycota* divisions. Such diversity suggests that oleaginicity may have emerged multiple times independently and thus can be supported by various mechanisms. Oleaginous yeast shares common properties: most of them are for instance obligate aerobes and cannot ferment carbohydrates (Sitepu et al., 2014). But the most studied trait of oleaginous yeast is their lipid accumulation under nitrogen starvation and in presence of an excess of carbon source. The process is the same for most oleaginous yeasts except for species like *Cryptococcus terricola* that can accumulate TAGs during logarithmic growth, before nitrogen depletion (Sitepu et al., 2014) and some yeasts that also respond to sulfur or phosphate limitation (Kolouchová et al., 2016; Wu et al., 2010).

De novo lipid synthesis starts to increase during the exponential growth phase where the nutrient supply is balanced allowing, rapid cell multiplication and, increasing lipid needs for membrane synthesis. Then, nitrogen starts to exhaust and the amount of nitrogen available is no longer sufficient to sustain proteins and nucleic acids productions required for cell multiplication. So, oleaginous strains continue to assimilate carbons producing this time storage lipids anticipating the coming exhaustion of the carbon source. If the system continues to evolve without further intervention, yeast will store lipids until total depletion of the carbon source and then start degrading the stored lipids for energy production (Beopoulos et al., 2008; Ratledge, 2013). A maximum lipid content is usually reached in the stationary phase quickly followed by lipid degradation. To avoid the lipid degradation phase, maintaining a constant supply of carbon source (Schulze, 2014) is recommended. Metabolically, nitrogen starvation decreases the intracellular

concentration of adenosine monophosphate (AMP) inhibiting thus the isocitrate dehydrogenase (ICDH: a key enzyme of Krebs cycle) which triggers a citrate accumulation in mitochondria. Citrate is then exported to the cytoplasm and used for the production of acetyl-CoA catalyzed by the ATP:citrate lyase (ACL) an enzyme found in numerous oleaginous microorganisms so far. This special feature of oleaginous microorganisms is complementary to the usual acetyl-CoA production pathway aforescribed (**see section 1.3.3.a**). The additional acetyl-CoA released by the action of the ACL can then enter the *de novo* lipid synthesis.

Although ACL contribution to oleaginicity is well established (even in plants and higher eukaryotes) it might be overestimated. Indeed, ACL was also found in some non-oleaginous organisms (Ratledge, 2002). As the oleaginicity definition does not include the condition under which the organism must accumulate at least 20% of its dry cell weight as lipids, it is possible that those non-oleaginous organisms expressing an ACL have an underlying potential for oleaginicity that just need to be unravelled. It could also suggest that the ACL presence is not enough to turn a strain into an oleaginous microorganism as supported by the variable effects of ACL genes overexpression depending on the organism (Dulermo et al., 2015). Another important factor to consider is that the ACL requires ammonium to function although ammonium concentration should be low during nitrogen starvation (Beopoulos et al., 2011). As suggested by recent transcriptomic and proteomic data, oleaginous organisms might benefit from better recycling systems of nitrogenous compounds during nitrogen-limiting conditions (Sitepu et al., 2014) allowing thus the allocation of ammonium to support the ACL activity for instance. Finally, it should be noted that ACC (Adrio, 2017), citrate efflux and ACL (Sitepu et al., 2014) are inhibited by long-chain fatty acid acyl-CoA (C₁₆ or C₁₈) through feedback inhibition loops. But fatty acyl-CoA accumulation strongly depends on the rate

Part I – Introduction

of fatty acyl-CoA modification and/or utilization and so might not always be a limiting factor.

Inhibited ICDH contributes less to NADPH supply during nitrogen starvation leaving the malic enzyme (ME) and the oxidative pentose pathway as main NADPH sources. As the ME was found in most oleaginous microorganisms and is associated with lipogenesis in mammal (Sitepu et al., 2014) it was thought to be the main NADPH source and the existence of ME/FAS/ACL complexes was even suggested (Beopoulos et al., 2011) but the existence of such complexes does not seem to be a feature common to all oleaginous yeasts and recent metabolic flux studies even demonstrated that the oxidative pentose pathway was the major NADPH source in *Yarrowia lipolytica* and possibly in *Rhodospiridium toruloides* (Adrio, 2017).

Oleaginous yeasts shares some similarities that can partially explain their common lipid accumulation properties but due to the taxonomic diversity of oleaginous species, several other mechanisms could be involved in oleaginicity. Especially mechanisms involving strain-specific alternative sources of acetyl-CoA such as amino acid degradation pathways or alternative acetyl-CoA saving lysine synthesis pathway (Sitepu et al., 2014). Lipid accumulation by oleaginous yeasts in response to nitrogen depletion has been studied for more than 40 years and there are still some mechanisms and regulatory paths to unravel to fully understand why all the oleaginous yeasts albeit united by some features exhibit a wide range of lipid accumulation efficiencies. A genomic and transcriptomic comparison of closely related oleaginous yeasts exhibiting different levels of lipid accumulation might reveal some of those mechanisms (Sitepu et al., 2014).

Meanwhile, the prospect of developing or finding other strains than *C. terricola* with the ability to accumulate storage lipids during the exponential growth phase would be an

asset from an industrial perspective as it would enable shorter production time and a less stringent control of the nitrogen availability (Sitepu et al., 2014). Actually, Xu *et al.* (2016) recently partially decoupled lipogenesis from nitrogen starvation in *Yarrowia lipolytica* by taking advantage of an alternative cytosolic acetyl-CoA formation pathway, which could be an interesting direction to follow.

1.3.4 - Advances, trends, and remaining gaps in oleaginous yeast research

1.3.4.a - Advances in metabolic engineering of oleaginous yeasts

The current understanding of yeasts lipid metabolism enabled several successful genetic engineering strategies for the improvement of lipid production (examples in **Table 1-2**), carbon efficiency or lipid composition. The implementation of these metabolic engineering enhancements was also facilitated by the concomitant development of appropriate genetic tools and the expansion of available genomic data, especially for popular oleaginous yeasts. Recent metabolic engineering strategies were reviewed elsewhere (Pfleger et al., 2015; Shi & Zhao, 2017; Yan & Pfleger, 2020) and seven trends, aiming at the diversion of carbon flux toward lipid biosynthesis emerged from those strategies (**Table 1-2**): (1) the overexpression of key enzymes of the lipid biosynthesis (2) the optimization of key enzymes activities (3) the blockage of competing pathways (4) the improvement precursors supply (5) the improvement of cofactor supply and redox balance (6) the release or sequestration of final products to create an irreversible pull toward product formation and also avoid toxicity and possible retro-control (i.e. product inhibition)(7) the mitigation of negative regulation of the fatty acid biosynthesis.

To illustrate the evolution of metabolic engineering strategies over the last 20 years, one example of successful strain design was picked among the ones available for the periods spanning from 2001 to 2007, 2008 to 2014 and 2015 to 2021 and listed in **Table 1-2**. These examples were harvested from pools of publications dealing with the improvement

Part I – Introduction

of native products. Excluding thus any reports pertaining to the expansion of carbon usage or to the production of non-native lipids that constitute a large proportion of the engineering efforts. Furthermore, the query was limited to three strains: the model yeast *S. cerevisiae* and two embodiments of yeast lipid research (i.e., *Y. lipolytica* and *R. toruloides*).

While it was possible to find examples from the 3 time periods for *S. cerevisiae* and *Y. lipolytica*, no engineering reports matching the criteria were found before 2016 for *R. toruloides*. Actually, the first genetic engineering methodologies suitable for the species were published only in 2014 (Koh et al., 2014; Lin et al., 2014). It took a few additional years before the description of promoters sets for *R. toruloides* (Johns et al., 2016; Liu et al., 2015, 2016; Wang et al., 2016) that were quickly followed by the first metabolic engineering publications (Lee et al., 2016; Zhang et al., 2016a). In the absence of efficient tools, enhancements of lipid production in *R. toruloides* were achieved through process engineering (Braunwald et al., 2013; Li et al., 2007; Shen et al., 2013; Wiebe et al., 2012; Wu et al., 2010, 2011; Zhao et al., 2011). Given the on-going trend of waste valorisation (see **section 1.1**) a lot of the feedstocks investigated for *R. toruloides* were wastes (Fei et al., 2016; Hu et al., 2009; Huang et al., 2016; Wang et al., 2012; Zhao et al., 2010, 2012). To bypass the limited genetic engineering tools, some researchers also investigated *R. toruloides* enhancement by mutagenesis (Qi et al., 2014, 2014; Yamada et al., 2017). The phylogenetic distance between *R. toruloides* and the model strain *S. cerevisiae* (**Appendix 1**), can explain why tools appropriate for the latter were not suitable for the former. At molecular level, two characteristics of *Rhodospiridium/Rhodotorula* species can be incriminated. The first is their high %GC content (>60%) which introduces a strong bias in codon usage compared to *S. cerevisiae* (Bornscheuer, 2018). The second is the species preferred DNA repair mechanism.

Table 1-2 Improvements of yeast native lipid production through metabolic engineering

Wild-type	Carbon source	Product(s)	Modifications	Initial metric	Final metric	Comment	Reference
<i>S. cerevisiae</i>							
4/17	Glucose	Triacylglycerides	(G) Acc1(S ₁₁₅₇ S ₁₂₅₇ >AA) + Δtgl3 + Δckb1 + Δare2 + DGA1 + Δgsy2 (S) Activity + Competing + Regulation + Competing + Biosynthesis + Competing	22 % DCW	65 % DCW	Improvement in lipid content limited growth	(Arhar et al., 2021)
BY4742	Glucose	Triacylglycerides (Others = FAL, FFA or FAEE)	(G) ACC1 + FAS1 + FAS2 (S) All Biosynthesis	4% DCW (TAG)	17% DCW (TAG)	Single additional overexpression will release other products	(Runguphan & Keasling, 2014)
BY4741	Glucose	Triacylglycerides	(G) Δsnf2 + DGA1 + FAA3 (S) Regulation + Biosynthesis + Biosynthesis	7 % DCW	30 % DCW	Higher lipid content when exogenous FA are supplemented	(Kamisaka et al., 2007)
<i>Y. lipolytica</i>							
NS18	Glucose	Triacylglycerides	(G) rtDGA1 + cpDGA2 + Δtgl3 (S) Biosynthesis + Biosynthesis + Competing	25 % DCW	77 % DCW	Fed-batch tripled the productivity	(Friedlander et al., 2016)
Polh	Glycerol	Triacylglycerides	(G) mmACL (S) Precursors	7 % DCW	23 % DCW	In native strain most of the citrate was secreted	(Zhang et al., 2014)
W29	Methyl ricinoleate	γ-Decalactone	(G) Δpox3 + Δpox4 + Δpox5 (S) Competing	150 mg/L	300 mg/L	Improvement in productivity altered growth	(Groguenin et al., 2004)
<i>R. toruloides</i>							
0126-16 (evolved variant)	Wheat straw hydrolysate	Triacylglycerides	(G) SCD1+ DGAT1 (S) All biosynthesis	54 % DCW	59 % DCW	DGAT1 alone higher lipid content but worst growth	(Díaz et al., 2018)
IFO0880	Glucose [g] Xylose [c]	Triacylglycerides	(G) ACC1 + DGA1 (S) All biosynthesis	31% DCW [g] 36% DCW [x]	61% DCW [g] 43.4 % DCW [x]	Less cell proliferation than in the wild-type	(Zhang et al., 2016a)
GM4 (<i>R. glutinis</i> *)	Glucose	Triacylglycerides	(G) mcME (S) Cofactors	19 % DCW	94 % DCW	Fatty acid profile was not affected by the modification	(Li et al., 2013)

R. toruloides* and *R. glutinis* share an anamorph/teleomorph relationship. (G) indicate the genotype and (S) standing for strategy give a glimpse into the rationale behind the modification. Strategy corresponding to the modification: **Biosynthesis, overexpression of key enzymes of the lipid biosynthesis; **Activity**, optimization of key enzymes activities; **Competing**, blockage of competing pathways; **Precursors**, improvement precursors supply; **Cofactors**, improvement of cofactor supply and redox balance; **Product**, release or sequestration of final products to create an irreversible pull toward product; **Regulation**, the mitigation of negative regulation of the fatty acid biosynthesis. **Mutation code**: %DCW, mg of product per 100mg of dry cell weight (all the values were rounded); Δ, deletion; **GENE**, overexpression; Gene(aa_{position}>newaa), point mutations; xxGENE, heterologous gene expression (xx native organism: cp: *Claviceps purpurea*, rt: *Rhodospiridium toruloides*, mc: *Mucor circinelloides*, mm: *Mus musculus*); **Abbreviation list**: aa, amino acids; **ACC1**, acetyl-CoA carboxylase; **ACL**, ATP: citrate lyase; **ARE2**, sterol acetyltransferase; **CKB1**, casein kinase II subunit beta; **DGA1/2** or **DGAT1**, diacylglycerol O-acyltransferase; **FA**, fatty acids; **FAA1**, long chain fatty acyl-CoA synthetase; **FAEE**, fatty ethyl esters; **FAL**, Fatty alcohols; **FFA**, free fatty acids; **FAS1/2**, fatty acid synthase; **GSY2**, glycogen synthase isoform 2; **ME**, malic enzyme; **POX3/4/5**, gene encoding acyl-CoA oxidases (β-oxidation); **SNF2**, sucrose non fermenting transcription regulator 2; **SCD1**, steroyl-Δ9-desaturase; **TGL3**, triacylglycerol lipase.

Many tools developed for genetic engineering of *S. cerevisiae* rely on the dominance of homologous recombination (HR) as repair mechanism in the Baker's yeast. Unfortunately, a lot of unconventional yeasts like *R. toruloides* favor non-homologous end joining (NHEJ) as repair mechanism limiting their genetic manipulation to random integrations (Koh et al., 2014; Löbs et al., 2017). Interestingly, the gap between *R. toruloides* and *S. cerevisiae* also prompted an early and detailed investigation of *R. toruloides* with a lipidomic (Buzzini et al., 2007), proteomic (Liu et al., 2009; Zhu et al., 2015), genomic (Coradetti et al., 2018), transcriptional (Bommareddy et al., 2017), multi-omics (Qi et al., 2017; Zhu et al., 2012) studies. The accumulation of omics data enabled the reconstruction of genome-scale metabolic models (Castañeda et al., 2018; Dinh et al., 2019; Tiukova et al., 2019) that recently supported a combined transcriptomic, proteomic, genomic and metabolomic integrative analysis (Kim et al., 2021). Just to illustrate how early omics data acquisition was for *R. toruloides*, the first comparative lipidomic analysis including *R. toruloides* was published in 2007 (Buzzini et al., 2007). When *S. cerevisiae* lipid profile was compared to the four other yeasts, including *Y. lipolytica*, only five years later (Hein & Hayen, 2012). *Y. lipolytica* and *S. cerevisiae* are both part of the *Saccharomycetales* order which did not prevent the need to adjust genetic engineering tools to *Y. lipolytica*. The first appropriate genetic parts and competent strains were described at the end of the previous century as reviewed elsewhere (Madzak et al., 2004) and the effort continued (Larroude et al., 2018) leading to the first lipid-related engineering attempt of *Y. lipolytica* in 2004. In 2004, the focus was however on *Y. lipolytica*'s ability to degrade methyl ricinoleate into γ -decalactone (Groguenin et al., 2004). This is not surprising knowing that *Y. lipolytica* first appeal was its utilization of hydrophobic compounds. But slowly, the focus shifted towards lipid production with metabolic engineering strategies very similar to the one implemented in *S. cerevisiae* since 2001 (**Table 1-2**). Those strategies led to lipid contents as high as

Part I – Introduction

77% of the dry cell weight (DCW) in *Y. lipolytica* when even in the most recent developments, *S. cerevisiae* lipid content never exceed 70% DCW (**Table 1-2**).

Metabolic engineering approaches require multiple expression cassettes constructions, transformations, and screenings. Typically, the effect of each individual modification will be analysed before combining the most performant ones. This strategy is thus time-consuming, cumbersome and albeit successful, only resulted in limited new industrial applications in the past few years (**section 1.3.2**). Yet metabolic engineering attempts can be credited for contributing to knowledge expansion and highlighting key bottlenecks. The first major bottleneck is that the improvement of lipid production/accumulation often leads to impaired growth (**Table 1-2**). The second issue is that some modifications have varying effects from one strain to another even for very close relatives.

Zhang *et al.* overexpressed the acetyl-CoA carboxylase (ACC1) and the diacylglycerol acyltransferase (DGA1) (2016a) in two *R. toruloides* strains (i.e. strains IFO0559 and IFO0880). The wild-type IFO0880 had a lipid content of 31.3 % DCW (**Table 1-2**) that was doubled by ACC1 and DGA1 overexpression. When the wild type IFO0559 had an initial lipid content of 42.9 % DCW against a 58.4 % DCW for the engineered version. In general, the improvements in lipid content observed after engineering strains naturally accumulating at least 40 % DCW as lipid seem to be lesser than the one observed when the wild-type has a lipid content below 40% DCW as exemplified by the work of Díaz *et al.* (2018) (**Table 1-2**). They only achieved 59% DCW lipid content for a wild-type already accumulating 54% DCW (Díaz et al., 2018). It can be hypothesized that strains accumulating high lipid contents already harbour optimized paths for lipid accumulation and are thus less sensitive to engineering attempts. Alternative engineering strategies must be found to overcome this limitation. This also highlights the importance

of a good strain selection. Since even among species known as good lipid producers, wild-type strains can have lipid content as low as 7% DCW (**Table 1-2**).

Furthermore, not only modification effects might vary from one strain to another, but the effects can sometimes be modulated by the carbon source used as described by Zhang *et al.* (2016a) since ACC1 and DGA1 overexpression increased lipid accumulation from glucose but not from xylose. Unfortunately, the number of reports comparing the metabolic engineering effects using different carbon sources is low when this knowledge could help identify genetic engineering targets that could increase lipid content independently of the carbon source used.

1.3.4.b - Systems biology and ‘omic’ technologies to the rescue

The guidance of powerful and reliable *in silico* predictive tools could help researchers overcome the current limitations of rational design and take the understanding of unconventional yeasts metabolism to another level. A better understanding of individual specificities will benefit both strain and process engineering. Because let’s not forget that the limited availability of appropriate genetic engineering tools remains the main impediment to metabolic engineering in unconventional yeast. And that, while appropriate tools are being developed, process engineering tends to dominate the engineering attempts.

Recently, genome-scale metabolic models (GEM, GSMM, GENRE) have emerged as tools with the potential to fill the gap between genotype and phenotype. GEMs are mathematical models of cellular fluxes containing at least genomic information but that can now integrate various types of omics data. GEMs can be used to evaluate a strain’s metabolic abilities, assess how some modifications might affect its performance, optimize growth conditions and evaluate its industrial potential. They can also be used to identify targets for genetic engineering and provide some very informative insight into cellular

Part I – Introduction

metabolic fluxes (Lopes & Rocha, 2017; Madhavan et al., 2017; Patra et al., 2021; Shi & Zhao, 2017; Zhang et al., 2016b). Yeast genome-scale modelling is evolving very quickly (Lopes & Rocha, 2017). For instance, the first *Y. lipolytica* GEMs were released in 2012 and a few reconstructions later, the current models can now be used not only to increase the overall lipid production (Kim et al., 2019) but also to identify targets that will push lipid production toward the formation of commercially interesting products like dicarboxylic acid (Mishra et al., 2018). A model-driven engineering of *Y. lipolytica* remains to date the best example of enhancement of lipid accumulation (Kavšček et al., 2015) with a final lipid content of 80% DCW that represents a 4-fold improvement. Yet this achievement is not so far from the 77% DCW (representing a 3-fold improvement) achieved in *Y. lipolytica* without the assistance of modelling (Friedlander et al., 2016). This illustrates how despite some recent successes the power of modelling is still limited and will require further improvement (Shi & Zhao, 2017). The predictive power and quality of the latest models were greatly improved by the increasing availability of omics data that in turn led to an GEM-based expansion of the understanding of unconventional yeast metabolism (Patra et al., 2021). The availability of experimental biochemical strain characterization data also helped refine and curate the models.

If genomes are the basis for GEMs reconstruction, the most difficult information to access for GEMs reconstruction are estimations of the macromolecular composition of strains and their ATP requirements (Patra et al., 2021). Because data availability is key in GEMs reconstruction, the oleaginous yeast GEMs were for a long time limited to *Y. lipolytica*, *C. tropicalis*, *R. toruloides* and *S. stipitis* (Lopes & Rocha, 2017) models which represent a small proportion of the 30 species that have been described to be associated with oleaginicacy (Caporusso et al., 2021). To this list can be added recent models of *Rhodotorula graminis*, *Lipomyces starkeyi*, *Pichia kudriavzevii* and *Wickerhamyces anomalus* (Correia & Mahadevan, 2018, 2020) but even with those recent additions

oleaginous yeasts genome-scale modelling remains far from been able to precisely reflect strain-to-strain metabolic variations. Patra *et al.* (2021) evaluated the unavailability of organism-specific metabolic information as the major limitation in GEMs reconstruction. And in my opinion the incompleteness of strain-specific metabolic information is as much of a limitation since experimental data shows that the response to metabolic engineering can be strain-specific (**see section 1.3.4.a**).

Although accurate strain-specific GEMs reconstruction is achievable and has been done in bacteria (Bosi *et al.*, 2016; Seif *et al.*, 2018) it remains a challenge for oleaginous yeasts due to the lack genomes. Yet, GEMs have proven to be useful when it comes to predicting targets that respect the fragile balance between growth and product accumulation. Furthermore, using the same model to investigate a candidate engineering target using different carbon sources is also possible so GEMs can offer solutions to the shortcomings of conventional rational design. But given the current state of system biology (GEMs) and synthetic biology (genetic engineering) for unconventional yeasts, cooperative development of both disciplines would be preferable.

1.3.4.c - Prospects

Although some oleaginous yeasts are currently commercially exploited for their production of lipids or derivatives, most oleochemicals are still not microbially-sourced. The main impediment to the establishment of yeast oleochemicals is their production cost. To be competitive a careful selection of both feedstock and product(s) must be done. The natural abilities of traditional industrial strains being limited, genetic engineering has been used to expand their carbon utilization, improve their performances, or tailor their lipid composition. Unfortunately, due to a lack of knowledge and/or appropriate tools these metabolic engineering endeavors are restricted to popular species (e.g. *Y. lipolytica*, *R. toruloides*) overlooking the potential of other unconventional species. Genetic engineering is not the only enhancement strategy possible, process engineering, adaptive evolution, random mutagenesis, and sexual

Part I – Introduction

hybridization (Guirimand et al., 2021; Steensels et al., 2014) are possible alternatives but none offers the same array of possibilities as genetic engineering. The development of methods and tools appropriate for the metabolic engineering of unconventional hosts seems unavoidable. In parallel, toolkits available for popular strains will continue to grow as illustrated by the recent advanced described for *Y. lipolytica* (Larroude et al., 2018) and *R. toruloides* (Schultz et al., 2019; Wen et al., 2020). Additionally, the characterization of these unconventional strains will only help future technology transfer or toolkit expansion from popular strains. Now, although genetic engineering endeavors are only possible for a few species, understanding lipid metabolism and its regulation in unconventional strains is a crucial stake. As their abilities might inspire better engineering strategies. The knowledge expansion effort could benefit from the increasing accessibility of omics technologies. And with the buildup of omics data, a growth of genome-scale modelling can be expected because GEMs are excellent tools for integrated data analysis. GEMs predictions should in turn contribute to more performant metabolic engineering strategies. Keeping in mind that the best engineering strategies might fall short without the appropriate commercial opportunities.

Chapter 2 - Project description

2.1 - Context, aims and objectives

The driving force of the project was the need to find new sustainable routes for glycerol valorization to absorb the worldwide excess of the waste that is currently causing an environmental threat and market instability. In our quest for new valorization routes, our interest focused on microbial lipid production because it addresses the growing concerns surrounding vegetable and animal oils & fats. Even though these current sources of oils & fats still have good years ahead of them, a high demand for more ethical and sustainable alternatives is expected in the coming years. But as described in **Chapter 1** microbial oil cannot compete with current sources of oil and fats for most applications except for very high-end ones (**see section 1.3.2**). So in order to develop a competitive microbial alternative, it was imperative to first try to improve microbial oil production methods. Efforts to rectify this situation have led to successful strain and process engineering examples but current approaches are too empirical and sometimes result in unwanted effects. The development of statistical design of experiment (DOE) methods is enabling a transition from “one factor at a time” (OFAT) approaches to multi-factor investigations (Gilman et al., 2021) but the use of microbial genome-scale modelling could make the whole development process adhere to quality-by-design standards advocating for quality through knowledge-driven designs. GEMs could support stages of process development such as strain selection, strain design and process engineering by providing valuable insights such as trait prediction, identification of genetic engineering targets or favorable set of conditions (e.g. media composition, oxygen levels etc.).

The contribution of GEMs to the development of yeast oleochemicals production processes is a work in progress and the improvement of existing models would require a mechanistic understanding of oleaginous yeast metabolisms. Oleaginous yeasts have been

Part I – Introduction

considered promising biomanufacturing hosts for years so their metabolisms and biochemical features have drawn a lot of attention over the years. And yet they all remain less characterized than the traditional host *S. cerevisiae*. The advent of omics technology expanded our understanding of oleaginicinity, but the expansion was fragmented (focused on individual species or clades). And our current understanding remains insufficient to fuel new industrial breakthroughs. Indeed, although it is commonly accepted now that unusual feedstocks like wastes will enable the commercial advent of microbial oil, most of the mechanisms were deciphered based on glucose utilization data. Acquiring data based on alternative feedstocks might at first seem unrealistic when considering the huge variety of options. But analyzing the effect of a few meaningful feedstocks could help us single out the individual parameters that truly impact growth and lipid production. Several studies already tried to evaluate the effect of waste impurities on growth and lipid production (Chatzifragkou & Papanikolaou, 2012; Gao et al., 2016; Pyle et al., 2008; Samul et al., 2014; Venkataramanan et al., 2012; Xu et al., 2012). Yet, these approaches remained mainly empirical, and the mechanisms, when proposed, were rarely further investigated.

The first aim of this work is thus to get a better understanding of how oleaginous yeasts differ and compare while generating knowledge that could feed GEMs reconstructions. For that purpose, a pan-genome of 17 unconventional yeasts was reconstructed and analyzed. The functional pan-genomic analysis and biochemical characterization supported the exploration of the conservation of industrially relevant traits. The functional pan-genomic analysis also gave new insight into oleaginous yeast lipid metabolisms.

The second aim was to confirm the suitability of sweetwater (fat-splitting crude glycerol) as feedstock for the lipid production of a panel oleaginous yeasts while gaining insights

into the molecular basis of lipid accumulation from crude glycerol. For that purpose, growth and lipid accumulation from crude glycerol were compared for a panel of yeasts and lipid accumulation mechanisms of the most performant strain was transcriptionally investigated. The two aims of this study are tightly related since (1) the comparison of crude glycerol utilization in different species also contributes to the characterization of their differences and similarities (2) the understanding of crude glycerol effect on yeasts could support the reconstruction of condition-specific GEMs. One could even claim the second aim to be an applied variant of the first one.

The expected impacts of this work are:

- (a) a consolidation of the phenotype::genotype bridge for oleaginous yeasts
- (b) the initiation of a reunification work of phylogenetically fragmented the knowledge already available for oleaginous yeasts. It should be emphasized that understanding how species differ and compare, will help rationalize data extrapolation from one species to another and should ultimately speed up knowledge expansion by reducing the number of strains to investigate. It will also have an impact on the development of GEMs for oleaginous yeasts since it is common practice during GEMs reconstruction to infer missing data from phylogenetically related organisms (Patra et al., 2021; Thiele & Palsson, 2010)
- (c) the improvement of the understanding of oleaginicinity by taking a step away from a glucose-centered knowledge of lipid metabolism and a step towards application-oriented mechanisms.

The main outcomes of this work would be (1) the biochemical characterization of 21 strains with industrial potential (2) the release of datasets ready to be used in GEMs reconstructions such as the functional annotation of 16 yeasts genomes including pan-

genomics information and transcriptomic data that supported the analysis of sweetwater effect on the most performing strain.

2.2 - Organization of the study

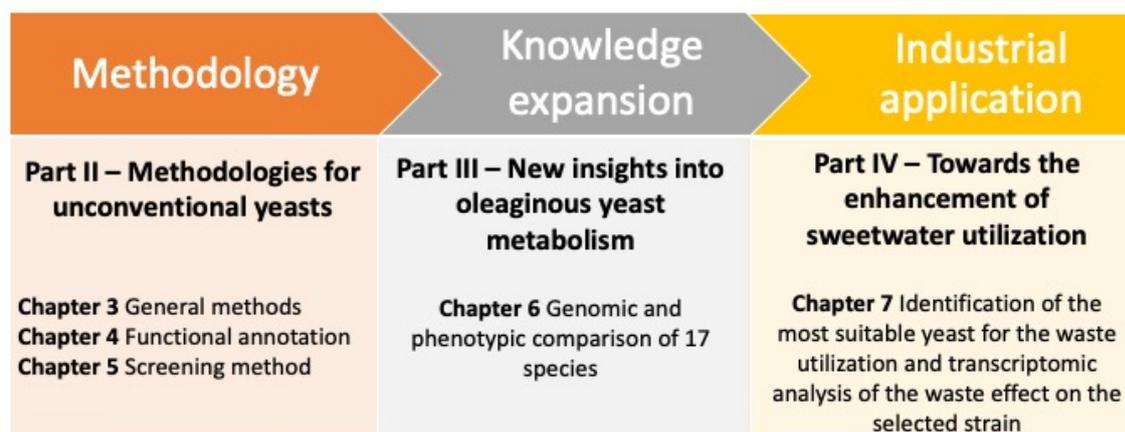


Figure 2-1 Organisation of this study

Aside from the introduction and conclusion that respectively correspond to Parts I and V, this manuscript contains 3 additional parts: (II) Methodologies for unconventional yeasts (III) New insights into oleaginous yeast metabolism and (IV) Towards the enhancement of sweetwater utilization (**Figure 2-1**). Part II describes the methodologies used and refined in Part III and IV that respectively cover the first and second aims.

Part II – Methodologies for unconventional yeasts

Chapter 3 - General materials and methods

3.1 - Materials

3.1.1 - Chemicals

Strains and their sources are listed in **Table 3-1**. Monopotassium phosphate and potassium chloride were respectively provided by Fluka (Gillingham, UK) and BDH (Brighouse, UK). D-(+)-glucose and yeast extract (granulated) were from Merck (Nottingham, UK). Peptone, agar, yeast nitrogen base without amino acids and ammonium (YNB) and complete supplement mixture (CSM) were from Formedium (Hunstanton, UK). Dimethyl sulfoxide (molecular biology grade), Disodium phosphate, sodium hydroxide pellets used to prepare 1 M solution in water were from AppliChem (Darmstadt, Germany). Malt extract was from VWR chemicals (Lutterworth, UK). And the sweetwater (SW: from splitters 8/01/18), and the HEAR oil were kindly provided by Croda (Hull, UK). All the media and solutions were prepared using MilliQ water prepared with a Millipore equipment (Molsheim, France).

Absolute ethanol, hydrochloric acid 37% (analytical grade), methanol, chloroform, glacial acetic acid were from Fisher scientific (Loughborough, UK). Sodium chloride was supplied by Formedium (Hunstanton, UK). Hexane was from Merck (Nottingham, UK). Nile red (9-diethylamino-5-benzo[α]-phenoxazinone), diethyl ether (with 1 ppm BHT as inhibitor), erucic acid, p-anisaldehyde were from Sigma-Aldrich (Saint Louis, USA). High erucic acid rapeseed (HEAR) oil splitting samples used as TLC standards were kindly provided by Mr. Robert Berchtold (University of Sheffield, UK).

Sweetwater batches (SW) were kindly provided by Croda (Hull, UK). The crudest and cheapest version of sweetwater has a glycerol content of 15 % (w/v) (pH 3.9 ± 0.1 @ 20°C) due to the presence of free fatty acids. It also contains a substantial amount of brown fatty residues observable as retentate on the filters used to further clarify the

waste (**Appendix 6**). A more concentrated version of the crudest sweetwater is available and exhibits similar properties except for its glycerol content that reaches 35% (w/v). If taken, further down the refining process, sweetwater will be cleaner but also more alkaline and more concentrated with a glycerol content of 85% (w/v). Since the solid residue can interfere with optical density readings, it was removed by a simple filtration step before any use of sweetwater. Furthermore, microbial contamination risk was addressed through steam sterilization of the waste. It will thereafter be implied that SWs have been filtered with 595 filter paper (Whatman, Dassel, Germany) and autoclaved before use. In the following manuscript, the various versions of sweetwater (SW) will be named after their glycerol content hence the crudest form will be noted SW15, the intermediate SW35 and the cleanest SW85. The same batches of SW were used throughout this study.

3.1.2 - Equipments

Optical densities at 600 nm (OD_{600}) were determined using the BioPhotometer Plus from Eppendorf AG (Hamburg, Germany) or the SpectraMax M2e microplate/cuvette reader of Molecular Devices (Wokingham, UK) that also has a fluorimeter function. Centrifugations were performed with either a Heraeus Megafuge 8R centrifuge (for 50-mL tubes) or a Fischer Scientific AccuSpin Micro 17 centrifuge from Thermo Fischer Scientific (Loughborough, UK) (for 1.5/2-mL tubes). Black polystyrene microtiter plate with clear and flat bottom from Grenier Bio-One (Frickenhouse, Germany) or black solid bottom microplates by Grenier Bio-One (Frickenhouse, Germany) were used for Nile Red assays. Clear 9-mm polystyrene lids with condensation rings were used with those microplates. Tubes and vials were heated using a dry bath from Starlab (Hamburg, Germany). Thin-layer chromatography was performed on TLC silica gel 60 F₂₅₄ plates from Merck (Nottingham, UK) and dried with a heat gun HL 1820 from Steinel (Peterborough, UK).

3.1.3 - Strains and genomes

Seventeen species were selected based on their described versatility in carbon utilization and dispositions for lipid accumulation. Were included both popular oleaginous yeast species and less investigated ones. Strains were obtained either from the Agricultural Research Service (ARS) culture collection (strain code starting by NRRL, Beltsville, USA), the “Deutsche Sammlung von Mikroorganismen und Zellkulturen” (DMSZ, Braunschweig, Germany) culture collection (strain code starting by DSM) or kindly provided by Dr. P. Unrean (Unrean & Champreda, 2017) (King Mongkut’s University of Technology, Thailand) but also available at the Thailand Bioresource Research Center (TBRC) culture collection (strain code starting by BCC) (**Table 3-1**).

At least one genome was publicly available for all the species except for *S. occidentalis* confirming the increasing availability of unconventional yeast genomes. Actually, more than one assembly was even listed for very popular species. When available, the genome of the strain experimentally studied was preferred otherwise a genome sequenced on a close relative was used as listed in **Table 3-1**. The evolutionary relationships of the studied species and how they are related to 6 reference species. (i.e. the model strains: *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. The extremophiles: *Debaryomyces hansenii* and *Zygosaccharomyces rouxii*. The basidiomycetes pathogens: *Cryptococcus neoformans* and *Ustilago maydis*) are presented in **Appendix 1** as phylogenetic tree, based on the 5.8S rRNA of the 17 investigated species and 6 reference, species was built.

3.2 - Cell culture

3.2.1 - Media and plates

Media destined for yeast cultivation were all sterilized by autoclaving. All plates and media were prepared using ultrapure MilliQ water. Yeast extract malt (YM) plates had

Table 3-1 Strains and genomes investigated in this study

n°	Species (3-letter code)	Synonym	Strains under experimental evaluation						Genomes	
			NRRL	DSMZ	TBRC	ATCC	NCYC	CBS	Sequenced strain	GenBank assembly accession
1	<i>Barnettozyma californica</i> (bca)	<i>Williopsis californica</i>	Y-1680			18118	496	5760	UCD 09	GCA_003123585.1
2	<i>Clavispora xylofermentans</i> (cxy)	<i>Spathaspora xylofermentans</i>			BCC 30719				UFMG- HMD23.3	GCA_002105455.1
3	<i>Cutaneotrichosporon curvatus</i> (ccu)	<i>Cryptococcus curvatus</i>	Y-1511			10567	476	570	ATCC 20509	GCA_001712445.1
4	<i>Cyberlindnera saturnus</i> (csa)	<i>Lindnera saturnus</i>	YB-4312						NRRL Y-17396	GCA_003709245.2
5	<i>Lipomyces lipofer</i> (lli)	<i>Waltomyces lipofer</i>	Y-11555	DSM70305		32031		944	NRRL Y-11555	GCA_003705915.1
6	<i>Lipomyces starkeyi</i> (lst)	-	Y-1388		BCC 45247	58680		1807	NRRL Y-11557	GCA_001661325.1
7	<i>Lodderomyces elongisporus</i> (lel)	<i>Saccharomyces elongisporus</i>	YB-4239			11503		2605	NRRL YB-4239	GCA_000149685.1
8	<i>Metschnikowia pulcherrima</i> (mpu)	<i>Chlamydozyma pulcherrima</i>	Y-5941- 53					5534	PRJNA508581	GCA_004217705.1
9	<i>Pichia kudriavzevii</i> (pku)	<i>Pichia orientalis</i>	Y-7551			24210	2658	5147	CBS 5147	GCA_003054405.1
10	<i>Rhodotorula glutinis</i> (rgl)	<i>Cryptococcus glutinis</i>	Y-2502			2527		20	ATCC 204091	GCA_000222205.2
11	<i>Rhodotorula mucilaginosa</i> (rmu)	<i>Sporobolomyces albo-rubescens</i>	Y-17283						JGTA-S1	GCA_003055205.1
12	<i>Rhodotorula toruloides</i> (rto)	<i>Rhodospiridium toruloides</i>	Y-6987					6016	-	-
13	<i>Rhodotorula toruloides</i> (rto)	<i>Rhodospiridium toruloides</i>	Y-1091			10788		921	ATCC 10788	GCA_001-542305.1
14	<i>Scheffersomyces stipitis</i> (sst)	<i>Pichia stipitis</i>	Y-7124		BCC 47637	58376		5773	-	-
15	<i>Scheffersomyces stipitis</i> (sst)	<i>Pichia stipitis</i>	Y-11545			58785		6054	CBS 6054	GCA_000209165.1
16	<i>Scheffersomyces stipitis</i> (sst)	<i>Pichia stipitis</i>			BCC 15191				-	-
17	<i>Schwanniomyces occidentalis</i> (soc)	<i>Debaryomyces occidentalis</i>	Y-2477			26077		2863	-	None available
18	<i>Torulaspora delbrueckii</i> (tde)	<i>Saccharomyces delbrueckii</i>	Y-866			10662		1146	CBS 1146	GCA_000243375.1
19	<i>Wickerhamomyces anomalus</i> (wan)	<i>Pichia anomala</i>	Y-366	DSM5759		8168	432	5759	NRRL Y-366	GCA_000147375.2
20	<i>Yarrowia lipolytica</i> (yli)	<i>Endomycopsis lipolytica</i>			BCC 64401				CLIB 89 (W29)	GCA_001761485.1
21	<i>Yarrowia lipolytica</i> (yli)	<i>Endomycopsis lipolytica</i>			Y203-A *				-	-

Species names listed are the ones used by the providing culture collection at the time of the acquisition. More durable culture collection iD numbers were also specified for each strain as well as at least one synonym. Given synonyms do not constitute a comprehensive list of the existing synonymy. *Y. lipolytica* variant Y203A* was kindly donated by Dr. P. Unrean (Unrean, 2017). In the following manuscript, identification numbers (n°) will be used to differentiate strains and 3-letter codes might be used to refer to species. Matches between experimentally evaluated strains and sequenced strains are highlighted in bold. For the other species, genomes of close relatives were used by default.

the following composition: 20 g/L agar, 3 g/L yeast extract, 3 g/L malt extract, 5 g/L peptone and 10 g/L glucose. Yeast peptone dextrose (YPD) broth was made of 10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose. $_SC^4$ (synthetic complete defined medium $_$ standing for the code of the carbon source used) were prepared following the given proportions: 1.9 g/L yeast nitrogen base (YNB-Formedium, Hunstanton, UK), 0.79 g/L complete supplement mixture (CSM Formedium, Hunstanton, UK), 5 g/L ammonium sulfate and a suitable amount of carbon source (glycerol, glucose, xylose and potassium lactate) introduced to achieve a C/N molar ratio around 9 g/L. Only sodium acetate was introduced regardless of the final C/N molar ratio but according to described tolerated concentrations at a final concentration of 5 g/L. When needed, media pH was adjusted to 6 ± 0.1 (@ 20°C) by the addition of sodium hydroxide (1 M). Nitrogen-limiting (high C/N ratio ~ 60) synthetic complete media was prepared by decreasing ammonium sulfate concentration to 0.72 g/L.

3.2.2 - Cell maintenance and cultivation

All the steps were conducted under aseptic conditions. Strains in 25% glycerol cryostocks (stored at -80°C) were streaked onto yeast extract malt (YM) agar plates. Streaked plates were then incubated at 25°C for 2 to 3 days. These actively growing cultures were typically sealed with parafilm and stored for up to 1 month at 4°C . When needed, inocula were prepared by picking a single colony (diameter around 2 mm) and transferring it into 5 mL of YPD. Inocula were cultivated at 25°C (210 rpm) until an early saturation phase is reached. Cultures were set by transferring enough inocula into a fresh volume of media to achieve an initial OD_{600} of 0.3. Cultures were cultivated at 25°C in 250-mL conical glass flasks non-baffled with a foam stopper or in 50-mL polypropylene sterile disposable centrifuge tube containing respectively 50 mL or 5 mL of appropriate medium. For growth in microtiter plates, adequate amounts of inoculum and media were

⁴ GlySC (glycerol SC), GlcSC (glucose SC), XylSC (xylose SC), LacSC (lactate SC), AcSC (acetate SC)

transferred into each well of clear-U-bottom 96-well microtiter polypropylene plates (ThermoScientific, Loughborough, UK). Each inoculation was performed in 4 replicates. Inoculated plates were then sealed and incubated at 25°C with a 1050 rpm orbital shaking. Whether cells were grown in tube/flasks or microtiter plates, the OD₆₀₀ of cultures were routinely monitored with an appropriate spectrophotometer. In all cases, a blank correction was first applied to measured OD₆₀₀. For culture in microtiter plates, the means of the corrected OD₆₀₀ of the four replicates and the associated standard deviations (SD) were considered.

3.3 - Lipid analysis

3.3.1 - Extraction procedure

For lipid extractions, cells from 40 to 48 mL of cultures were harvested by centrifugation (12 min; 3000xg; 4°C). The resulting pellets were then stored at – 80 °C. When needed, pellets were thawed and vortexed for 1 min. A weight of 775 ± 45 mg of each pellet was transferred into a 50-mL polypropylene sterile disposable centrifuge tube. For cell hydrolysis, 3 mL of HCl (1 M) were added to each pellet. Mixtures were vortexed (1 min) before a 2h incubation at 78°C interrupted every 30 min for 1 min vortexing at room temperature. Hydrolyzed cells were transferred into a separating funnel. Enough water was added to achieve a final volume of 10 ml before adding an equal volume of a chloroform/methanol (1:1; v/v) mix. Funnels were shaken vigorously. Phase separation was allowed and followed by removal of the upper phase.

The lower phase was then washed with 10 mL of 0.1% (w/v) NaCl. After phase separation, the upper phase was removed, and the lower phase washed with 10 mL of MilliQ water. A final phase separation occurred, and the interesting fractions were transferred into 10-mL vials and dried at 65°C using a dry bath. Dried extracts were resuspended into 500 µL of hexane. The resulting solutions were then transferred into 5-mL dark vials and stored at – 20°C until further analysis.

3.3.2 - TLC

A 10 mg/mL erucic acid solution in EtOH and a 20 mg/mL HEAR oil splitting product dissolved in EtOH were used as standards. A staining solution, containing 300 mL of 95% (*v/v*) EtOH, 12 mL of *p*-anisaldehyde, 6 mL of glacial acetic acid and 12 mL of concentrated H₂SO₄ was prepared. Volumes of 5 µL of extracts or standards were loaded on TLC silica plates. Plates were developed in pre-saturated chambers with a mixture of hexane: diethyl ether: acetic acid (70:30:1) as mobile phase. Plates were removed from the development chamber when the migration front was at about 1 cm from the top of the plate. Migration front was scored, and plates were left at room temperature to dry before the use of a detection system. The plate was then quickly submerged into the staining solution and finally dried using a heat gun.

3.3.1 - General procedure for Nile Red staining

For Nile Red assays, two aliquots of 1 mL of culture were harvested by centrifugation (2 min; 10000xg; RT). The latter pellets were then washed twice with 1 mL of phosphate-buffered saline (PBS: 1.23 g/L of KH₂PO₄, 0.416 g/L of Na₂HPO₄, 8 g/L of NaCl, 0.201 g/L of KCl – pH adjusted to 7 by addition of HCl (1M)) and then stored at – 80 °C. Before each experiment, enough PBS:DMSO (1:1, *v/v*) solution and Nile red working solution in DMSO or acetone were prepared. Washed cell pellets were thawed and resuspended in 500 µL to 1 mL of PBS. The OD₆₀₀ of each suspension was determined using a spectrophotometer and, if needed, cell density was adjusted to the desired OD₆₀₀. To a black polystyrene microtiter plate with clear and flat bottom were sequentially added 250 µL of cell suspension or PBS, 50 µL of PBS:DMSO and lastly 50 µL of the Nile red working solution for a total volume of 350 µL (that can be proportionally downscaled to 200 µL) . Components were introduced using a multichannel pipet and well's content was gently mixed after Nile Red addition. For each sample, 4 replicates and 4 blanks were loaded. The plate was then introduced in the plate reader for

measurement at specified parameters including a 10 sec plate shaking before top readings. The microplate was incubated at room temperature and in the dark between readings (when applicable). For repeated measurements the highest fluorescence signal was plotted regardless of the staining time unless the experiment was specifically looking at signal evolution over time. The mean signal of all the blanks on the plate was subtracted from each sample replicate signal for background correction. Background corrected data will be noted RFU_c, RFU_c can further be normalized by division by the OD₆₀₀ of the sample to account for small variation in cell density and give RFU_{co}. Or be standardized by multiplication by a dilution factor and give RFU_{cd}. Unless stated differently, the mean signal of 4 replicates and the associated standard deviation (SD) is presented in plots.

3.4 - Pan-genomics: principles and definitions

The pan-genomic analysis is a homology-based method that compares genes of a pan-genome between them rather than comparing them to external databases. The goal of functional annotation is to associate proteins to functions. When pan-genomics aims at differentiating strain-specific genes from shared genes within a defined group of genomes. Genes of the pan-genome can be grouped into 3 categories: core, accessory, and unique genes. The definition of each category slightly varies from one study to another. Here, unique genes will be defined as specific to one species and homologous to no other gene of the studied pan-genome. Homologous genes of the pan-genome form clusters. When a cluster contains at least one gene of each species, it is considered to be “core” otherwise it is “accessory” (**Figure 3-1**). In other words, core genes find a homologous counterpart(s) in each studied genome.



Figure 3-1 Illustration of the pan-genomic categories

In pan-genomics, it is important to keep in mind that the term homology encompasses multiple concepts. Two sequences are homologous when they share a significant percentage of identity. The homology will be named orthology if it was inherited from a common ancestor through speciation event. The homology will be called paralogy if it results from a locus duplication event. When whole-genome duplication event will lead to ohnology. And finally, homology is also sometimes the result of horizontal gene transfers. Yet, it should be noted that most methods detect homology independently of its source which can sometime be problematic (Correia et al., 2019).

Chapter 4 - Pan-genomic data can improve the functional annotation of unconventional yeasts genomes

4.1 - Introduction

Annotation was defined by Ekblom *et al.* (2014) as the “computational process of attaching biologically relevant information to genome sequence data”. The annotation process encompasses gene prediction, gene annotation⁵ and functional annotation (Ekblom & Wolf, 2014; Yandell & Ence, 2012). Yeast genes have more complex structures than prokaryotic genes and are separated by more intergenic spaces (Song *et al.*, 2015). But compared to higher eukaryotes, they exhibit a higher gene density, fewer introns and lesser repetitive DNA (Wolfe, 2006). Given these specificities, yeast gene prediction and annotation were quite challenging for a while (Proux-Wéra *et al.*, 2012) but now a range of tools trained for yeast genome prediction is available (Hoff & Stanke, 2019; Liu *et al.*, 2018; Min *et al.*, 2017; Proux-Wéra *et al.*, 2012; Zhang, 2002). Nowadays the remaining challenge is the functional annotation of unconventional yeast (Schneider *et al.*, 2011; Song *et al.*, 2015). Indeed, for a gene to be associated with a biological function, this function must have been described and characterized. Ideally, the characterization must have been done in the investigated organism or a close relative for the functional annotation to be accurate. Especially for yeasts that are organisms known to evolve rapidly and thus exhibiting significant genotypic variations even at species level (Song *et al.*, 2015). The current state of characterization of yeast proteins complicates the functional annotation of yeast that are not close phylogenetic relatives of well-characterized species like *S.cerevisiae*. Actually, in 2005, it was evaluated that all the functional annotations were derived from less than 5% of all the known proteins

⁵ Gene prediction is the computational prediction of a coding sequence (CDS) from assembled genomes sequences. When gene annotation is the process of associating each CDS to other genomic features like untranslated regions, stop/start codons, splicing sites

Part II – Methodologies for unconventional yeasts

(Valencia, 2005) and *S. cerevisiae* counts among the most characterized organisms with 85% of its proteins already associated with a loss function ten years ago (Botstein & Fink, 2011). Typical functional annotation methods infer protein functions by identifying proteins from reference databases homologous to the query protein (Chen et al., 2013; Loewenstein et al., 2009; Proux-Wéra et al., 2012; Punta & Ofran, 2008; Ruiz-Perez et al., 2021). Genes exhibiting a high similarity are assumed to have similar functions which can lead to “over-annotation”. “Over-annotation” happens when a function is falsely attributed to an organism due to these homology-based inferences (Chen et al., 2013; Sasson et al., 2006; Valencia, 2005). The accuracy of GEMs that are built based on functional annotations can greatly be affected by “over-annotations”. GEMs are not the only application of functional annotations. They also support functional genomic analysis and transcriptomic analysis so ensuring good quality functional annotations is essential for the subsequent analysis steps.

Yet, there are no standard methods to assess the quality of a functional annotation or to identify suspicious annotations except confronting the functional annotation to the experimental reality by performing genome-based trait predictions for instance (Chen et al., 2013). Such quality control approach is limited by the availability of experimental data and the lack of knowledge surrounding the genetic basis of certain phenotypes. Furthermore, this type of approach would only be limited to genes involved in metabolism or any measurable phenotype. Confronting proteins to different databases of different contents with tools of different sensitivity often results in different functional annotations (Ruiz-Perez et al., 2021). These discrepancies are considered to be an issue (Chen et al., 2013) but with a method to combine and compare functional annotations from different sources, it would be possible to take advantage of the phenomenon. Combining functional annotation from different sources could be a way to increase the three characteristics of annotations: consistency, coverage and accuracy (Chen et al.,

2013). The consistency is the biological relevance of the functional annotation, the coverage is the number of proteins associated with a function and the accuracy is the depth of understanding surrounding the biological role of assigned functions (e.g. the annotation “enzyme with activity X” is less accurate than “enzyme with activity X involved in process Y”). So far, data combination and integration from different sources remain a challenge (Schneider et al., 2011) although a successful attempt has recently been published using bacterial genomes (Griesemer et al., 2018).

This chapter focuses on the functional annotation of unconventional yeasts. More specifically on the improvement of the functional annotation process by combination of multiple annotations from various sources with pan-genomic information. This chapter explores a flexible methodology to generate quality ready-to-use functional annotations.

4.2 - Pipeline development

4.2.1 - Comparing tools for functional annotation

Each annotation tool has its specific nomenclature and the challenge is to find a common language (Chen et al., 2013) with biological relevance. EC numbers, KEGG orthology numbers (KO), Cluster of Orthologous Group (COG) and Gene Orthology (GO) terms are the most common identifiers. EC and KO numbers are particularly attractive because there are commonly used in GEMs reconstructions, but EC numbers come with three major limitations: (1) some proteins are only assigned partial EC numbers during the annotation (e.g. EC:1.2.-.-) (2) some activities are associated to several EC numbers due to substrate ambiguity (Khersonsky & Tawfik, 2010) (3) knowledge of the enzymatic activity does not necessarily mean knowledge of the biological function. When biological relevance is at the center of KO, COG and GO nomenclatures. But COGs offer limited functional information (Chen et al., 2013) compared to KO numbers and GO terms while GO terms are afflicted by a high redundancy (Jantzen et al., 2011). Hence KO numbers

Part II – Methodologies for unconventional yeasts

were thus chosen here as primary connectors unlike the strategy described by Griesemer *et al.* (2018) that based their combination strategy on Enzyme Commission (EC) numbers using enzyme annotation from Kyoto Encyclopaedia of Genes and Genomes (KEGG)(Kanehisa et al., 2014), Rapid Annotations using Subsystems Technology (RAST)(Aziz et al., 2008), Enzyme Function Inference by Combined Approach (EFICAz)(Kumar & Skolnick, 2012) and the Comprehensive Enzyme Information System BRENDA (Chang et al., 2009, 2021). KEGG, BRENDA and EFICAz are general sources of annotation when RAST is specific to prokaryotes.

Here, six candidate tools were investigated as primary sources of annotations for this pipeline: BlastKOALA (Kanehisa et al., 2016), KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007), KEGG Orthology-Based Annotation System (KOBAS) (Xie et al., 2011) Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (EggNOG) mapper (Huerta-Cepas et al., 2016, 2019), NCBI Blastp (Kanehisa et al., 2016) and antiSMASH fungal version (Blin et al., 2019). KEGG being one of the major sources of information for GEMs reconstruction (Schneider et al., 2011; Thiele & Palsson, 2010), two of the candidate tools were naturally picked among the tools hosted on their servers. The other candidate tools are hosted all over the world by: the European Molecular Biology Laboratory (EMBL), the Center for Bioinformatics of Peking University (CBI-PKU), the National Center for Biotechnology Information (NCBI) or the Technical University of Denmark (DTU) (**Table 4-2**). They are all general tools except antiSMASH included for its fungal-specificity, and its use of Hidden Markov marker and profiles (HMMer)(Finn et al., 2011; Johnson et al., 2010).The use of HMMer is a characteristic shared with EggNOGmapper that also use DIAMOND (Buchfink et al., 2015) and Many-against-Many sequence searching tool (MMseqs2) (Steinegger & Söding, 2017) when the other tools all rely on the basic local alignment search tool (BLAST). Ultimately EggNOGmapper was preferred over AntiSMASH fungal version

because it provides more malleable output offering more interconnection opportunities (**Table 4-2**). NCBI BLAST initially appeared like an attractive solution because it was the only option allowing the use fully customizable databases that would have enabled the use of curated fungal databases like AyBRAH (Correia et al., 2019). But NCBI BLAST was also excluded due to a lack of interconnectivity (**Table 4-2**). After a thorough comparison, four tools were chosen: EggNOG, KOBAS, KASS and BlastKOALA. As illustrated in **Figure 4-1**, after retrieving the appropriate sequences (**Table 4-1**) annotations were first obtained from these tools to form a compilation of juxtaposed annotations (**Appendix 3**).

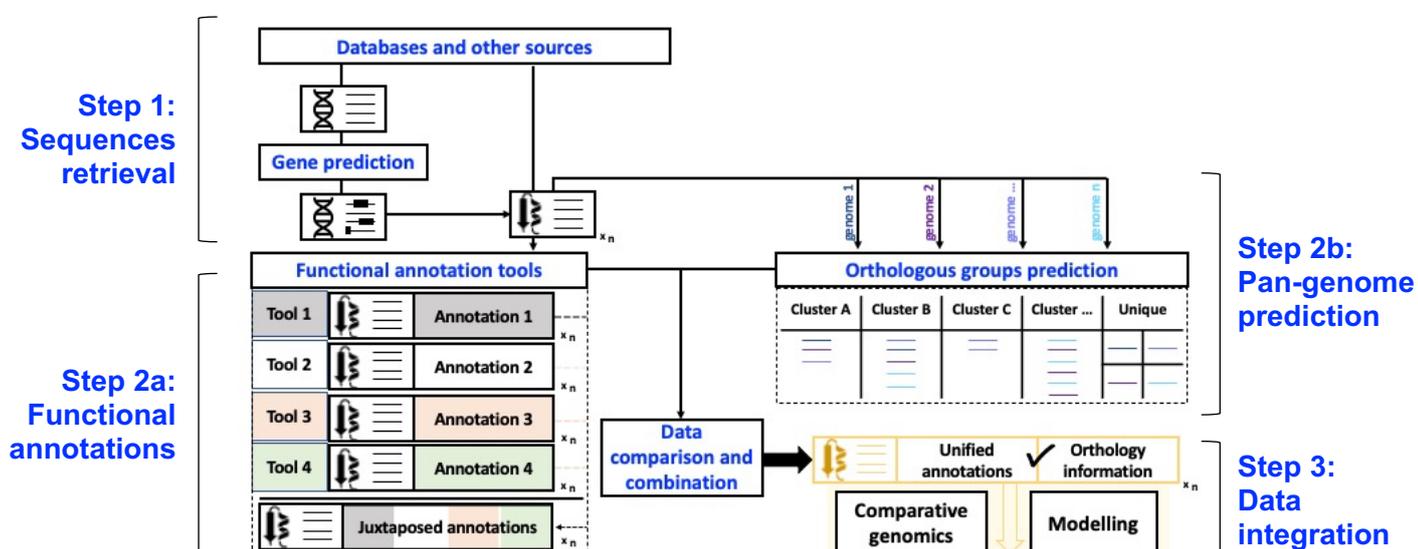


Figure 4-1 Graphical illustration of the annotation pipeline principle

The annotation pipeline was designed to generate unified functional annotations of n yeast genomes by combining functional annotations from various external sources (tools 1 to 4) with orthology information of the studied panel of genomes. The resulting annotations are ready-to-use for comparative genomics or genome-scale modelling applications. The pipeline can use genome assemblies (with an additional gene prediction step) or sets of protein sequences. For illustration purpose, an example of juxtaposed annotation is given in Appendix 3.

Table 4-1 General features of investigated genomes

Species (3-letter code)	Size (Mb)	GC content (%)	Gene and protein prediction	# Proteins	Homology to <i>S. cerevisiae</i> (%)
<i>Barnettozyma californica</i> (bca)	11.67	41.50	Augustus (this work)	5805	80.24
<i>Clavispora xylofermentans</i> (cxy)	15.10	35.30	Augustus (this work)	6001	91.63
<i>Cutaneotrichosporon curvatus</i> (ccu)	19.91	60.70	Augustus (this work)	7805	46.56
<i>Cyberlindnera saturnus</i> (csa)	13.43	42.80	Augustus (this work)	6816	73.97
<i>Lipomyces lipofer</i> (lli)	19.86	48.20	Augustus (this work)	7647	55.77
<i>Lipomyces starkeyi</i> (lst)	21.27	47.00	JGI annotation pipeline (Huntemann et al., 2015; Riley et al., 2016)	8183	57.28
<i>Lodderomyces elongisporus</i> (lel)	15.55	37.00	GeneWise + Glimmer (Butler et al., 2009)	5799	75.53
<i>Metschnikowia aff. pulcherrima</i> (mpu)	15.80	45.89	Modified version of the YGAP (Gore-Lloyd et al., 2019; Proux-Wéra et al., 2012)	5795	73.74
<i>Pichia kudriavzevii</i> (pku)	10.78	38.34	Augustus (this work)	5109	80.60
<i>Rhodotorula glutinis</i> (rgl)	20.48	61.90	GeneMarkS + Augustus + Glimmer (Paul et al., 2014)	2817	49.56
<i>Rhodotorula mucilaginosa</i> (rmu)	20.22	60.50	Augustus (this work)	5965	72.89
<i>Rhodospiridium toruloides</i> (rto)	20.75	61.90	Augustus (this work)	6126	55.52
<i>Scheffersomyces stipitis</i> (sst)	15.44	41.16	JGI annotation pipeline (Huntemann et al., 2015; Jeffries et al., 2007)	5818	79.06
<i>Torulasporea delbrueckii</i> (tde)	9.22	42.04	Yeast genome annotation pipeline YGAP (Gordon et al., 2011; Proux-Wéra et al., 2012)	4970	96.04
<i>Wickerhamomyces anomalus</i> (wan)	26.56	35.10	JGI annotation pipeline (Huntemann et al., 2015; Riley et al., 2016)	6421	79.04
<i>Yarrowia lipolytica</i> (yli)	20.55	48.98	Mapping + YGAP + Snowy-Owl pipeline (Magnan et al., 2016)	7949	57.42

A comparison of each genome with the genome of *S. cerevisiae* (sce) on the KOBAS platform (Xie et al., 2011) was used to determine the percentage of proteins homologous to see proteins

Table 4-2 Comparative table of candidate tools for functional annotation

	BlastKOALA	KAAS	EggNOG-Mapper	KOBAS	NCBI BLAST suite	antiSMASH fungal version
Version	2.2 (2019)	April 2015	Web-based: 2.0.1 (2016) * Standalone: 2.1.5 (2021)	2.1 (2016) or 3.0 (2021)	2.9.0 (2019) to 2.12.0 (2021)	5.0 (2019)
Accessibility	Web-based	Web-based	Web-based and standalone	Web-based and standalone	Web-based and standalone	Web-based and standalone
Host	KEGG	KEGG	EMBL	CBI-PKU	NCBI	DTU
Input	Protein sequences (<7500 entries)	Transcript/ protein sequences	Transcript/protein sequences or assembled genome	Transcript/ protein sequences or BLAST output	Transcript/ protein sequences (<5000 entries)	Assembled genome
Algorithm	BLAST	BLAST, GHOSTX/Z	HMMer + DIAMOND + MMSEQS2	BLAST	BLAST	HMMer but BLAST possible
Database	Predefined	Customizable within a predefined set	Predefined	Customizable within a predefined set	Fully customizable	Predefined
Speed	1 day per genome	< 30 min per genome (from protein sequence)	< 1h per genome (from protein sequence)	< 30 min per genome	1 day per genome	< 1h per genome (from protein sequence)
Output	iD KO1 iDg;Definition[EC #] Score KO2 Score	iD KO	iD seed e-value score hit gene iD iD GO KO BiGG taxonomic scope eggNOG_OG COG Definition	iD Hit iD iDg Hyperlink	BLAST tabular format	Complex HTML or JSON
Post-processing	Minimal	None	Minimal	Medium	Medium	Difficult
Interconnectivity	Medium	Low	High	Medium	Low	Low

*Database version 5.0 (2019). The post-processing is minimal if manually accessible, medium if light programming is required and difficult if heavy programming is required. Output interconnectivity is low for outputs containing only internal references, medium for outputs containing a single type of external reference and high for outputs containing more than one type of external reference. All the abbreviations are described in the abbreviation list.

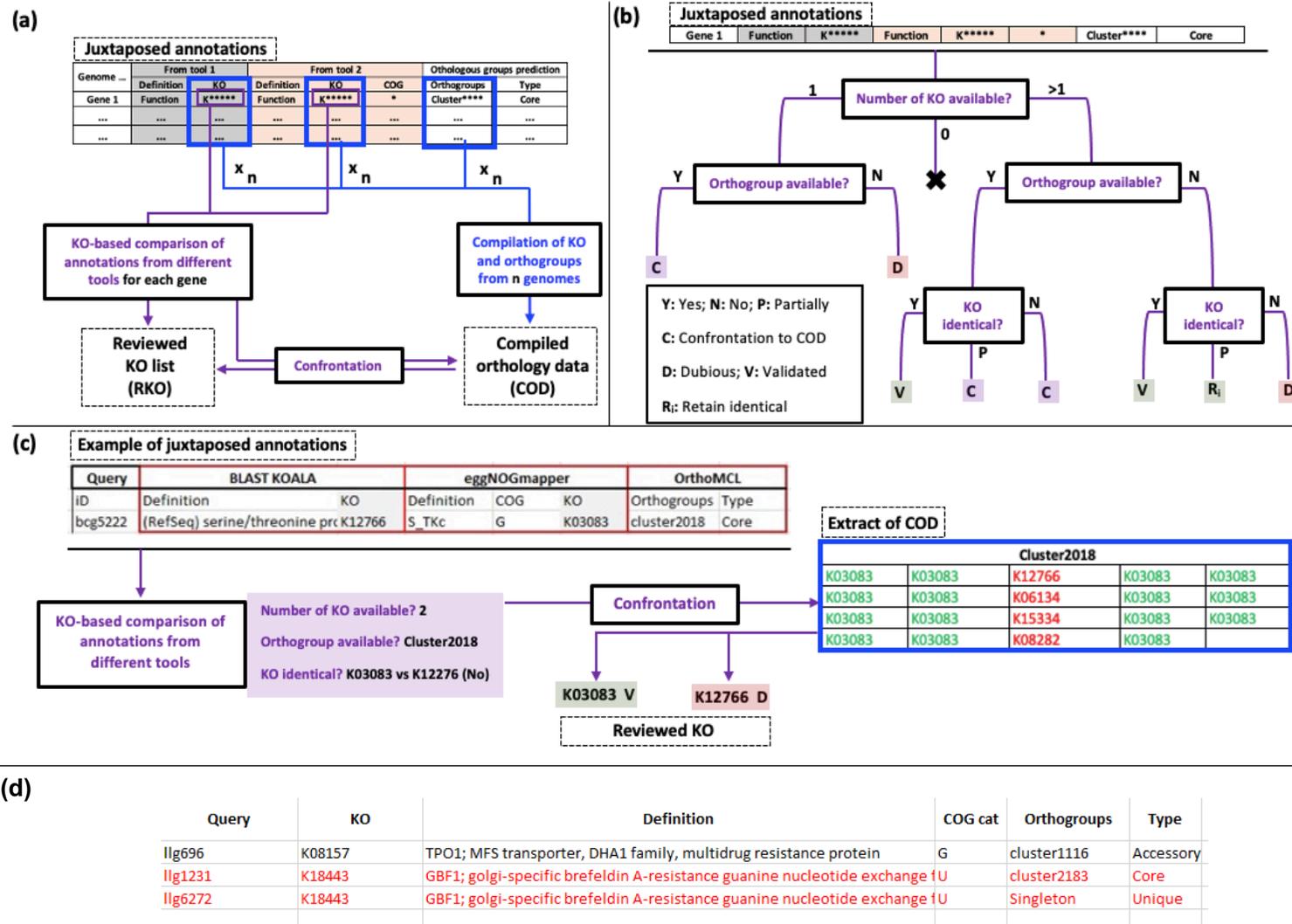


Figure 4-2 Scheme illustrating the combination-validation developed and examples

(a) General principle. Orthogroups is an abbreviation for orthologous groups (b) Decision tree (c) Detailed example (d) Another type of annotation flagged. While llg1231 (length 1480 aa) was correctly associated with the cluster2183 llg6272 (309 aa) is not included in the cluster and was flagged as dubious.

4.2.2 - Combination-validation strategy

At this stage, the main challenge was to determine how to use KOs to compare and combine the different functional annotations. The strategy implemented is described in **Figure 4-2**. A list of orthologous groups (cluster)/KO associations is harvested from the file of juxtaposed annotations to form compiled orthology data (COD) (**Figure 4-2a**). With the assumption that orthologous clusters will contain genes of similar function, KOs represented several times in the same cluster will be validated while KOs represented a single time are considered dubious annotations. Depending on the juxtaposed annotations, several cases are possible. A gene could either be associated with one KO, no KO or several KOs. The KO-based comparison is not possible in the absence of KO. When one KO is available it will be confronted to the COD when the gene is associated to an orthologous group, otherwise it will be marked "dubious" (**Figure 4-2b**). When several KOs are associated with the same gene, identical KOs will automatically be validated (inter-source validation) while discrepant KOs will be confronted to COD if the gene is associated with an orthologous group (**Figure 4-2b**).

The influence of the number of annotation tools was also investigated. The KO coverage was compared achieved with 2 tool was compared to the one achieved with 4 tools (**Figure 4-3**). The use of 4 tools results in an average of 6.5% additional proteins being associated with KOs and the number of proteins associated with several KOs is on average 41.5 time higher. The use of 4 tools alleviates the need for COD confrontation because most of the KOs are reviewed through inter-source validation. COD confrontation is essential when using 2 sources because a large proportion of the genes are only associated with a single KO. The number of genomes to analyze will determine the most suitable strategy because when analyzing a large number of genomes the time can quickly become a constraint. The highest KO coverage recorded was 69.7% of the total proteins (**Figure 4-3: tde-4 tools**) but the average KO coverage was around 55%

indicating that half of the genes were not eligible for a KO-based validation. Among those genes, a lot were associated with non-informative annotations. A list of typical non-informative annotations was put together (Unknown markers list - Examples in **Appendix 3**) and proteins associated with those markers were annotated as “hypothetical proteins”.

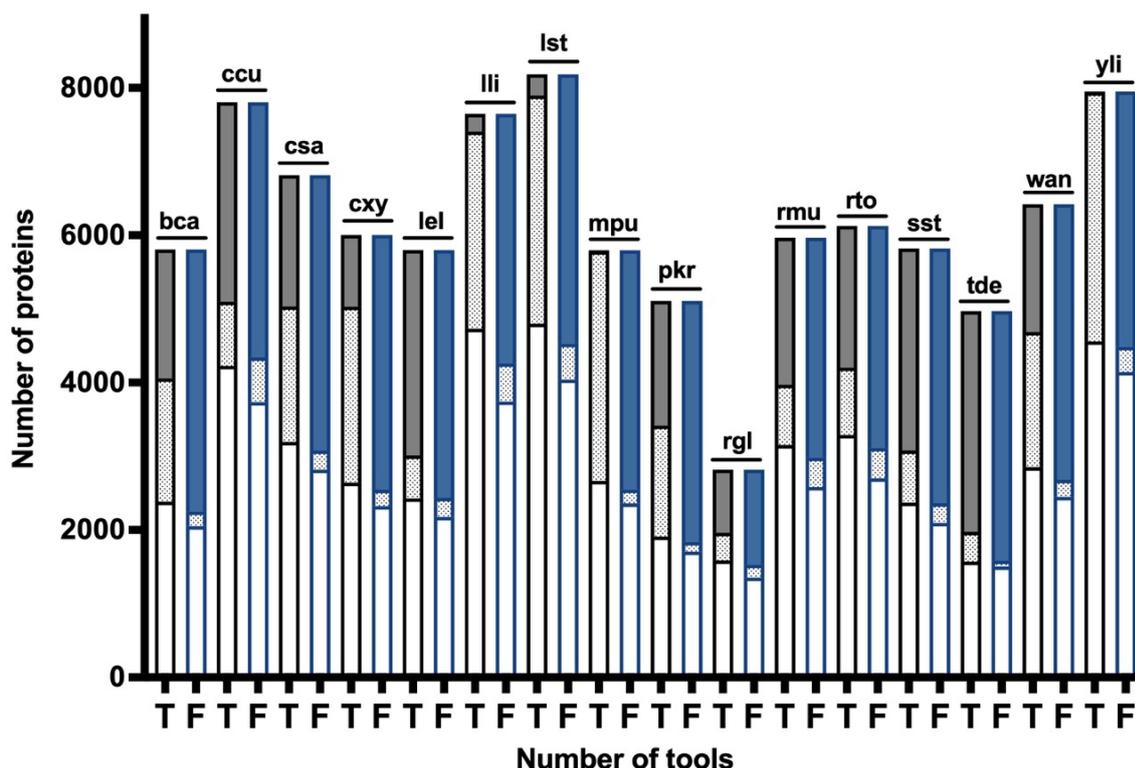


Figure 4-3 Influence of the number of annotation tools on the KO coverage
 T (grey bars) means that 2 annotation tools were used (EggNOGmapper + KOBAS) and F (bleu bars) means that the 4 tools were used (EggNOGmapper + KOBAS + BlastKOALA + KASS). The solid portion of each bar corresponds to genes with several KOs, the dotted proportion represents KO from a single source while the empty proportion illustrates the number of genes non-associated with a KO

Given the limited coverage of KOs, COGs were used as secondary connectors. The only tool providing COG is EggNOG so all COGs were reviewed through a COG version of the COD confrontation. The contribution of COG is minor. For instance, the genome of *B. californica* encodes 5805 proteins out of these 3678 annotations were KO-validated, 3932 were COG-validated and there is an overlap of 3244 annotations between the KO

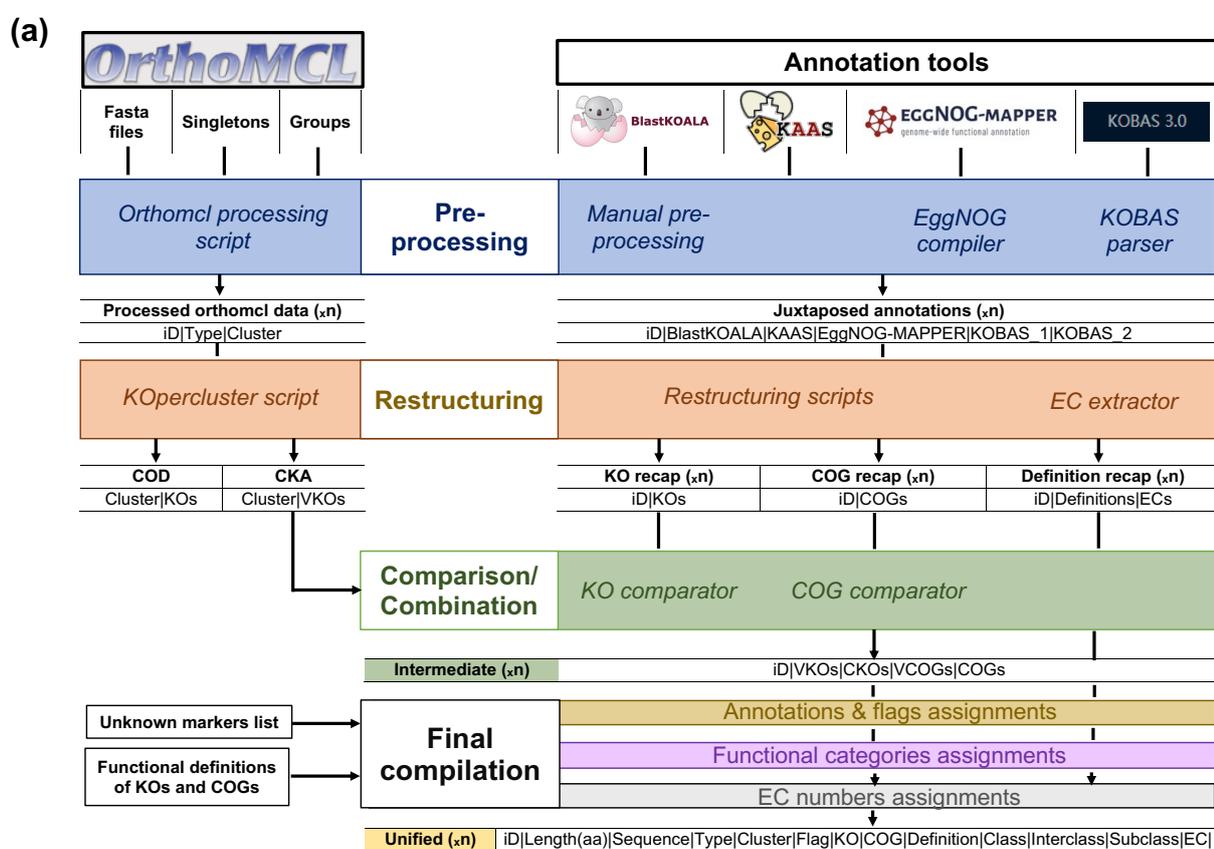
and COG-validated meaning that COGs only add additional information for 688 proteins. Furthermore, the information added by the COG nomenclature is sufficient for comparative genomics but not for model reconstruction.

4.2.3 - Method description

Given the number of tools available for yeast genome prediction (Hoff & Stanke, 2019; Liu et al., 2018; Min et al., 2017; Proux-Wéra et al., 2012; Zhang, 2002), this method mainly aims at improving the quality of functional annotations rather than focusing on gene prediction. But because gene prediction hasn't been performed on all the studied genomes (**Table 4-1**), the described method contains an optional gene prediction step performed using Augustus (Hoff & Stanke, 2019). This tool has been previously successfully used for fungal gene prediction (**Table 4-1**) and is particularly advantageous because it includes pre-trained parameter sets of several fungi alleviating the need for additional experimental input (other than genomic data). With Augustus, the user can either set the existing parameters of a target species or choose to use the parameters of a close relative if the ones for the target species are not available. However, the pipeline is flexible and will admit gene predictions from other tools (**Table 4-1**) as long as they are provided in GFF format.

Genomes or annotations were downloaded from National Center for Biotechnology Information (NCBI) using the accession number provided in **Table 3-1**. When no annotation was publicly available, coding sequences were predicted using Augustus (Hoff & Stanke, 2019). Partial gene prediction was enabled, and pre-trained parameters of the closest relative available were used. Protein predictions of this study were based on the universal genetic code for all strains except for the proteins of *C. xylofermentans* that were predicted using the yeast alternative code. All the proteomes were combined to form a pan-genome that was then clustered into orthologous groups using OrthoMCL (Li et al., 2003). Proteins not included in any cluster form the «unique» portion of the

pan-genome. Other proteins are either in «core» clusters (found across the 16 studied proteomes) or in «accessory» ones. The functional annotation of the 16 proteomes was achieved by running them through online annotation tools and combining the results based on Kyoto Encyclopaedia of Genes and Genomes (KEGG) Orthology (KO) numbers (Kanehisa et al., 2014), Clusters of Orthologous Groups (COG) (Tatusov et al., 2003) and ortholog group prediction of the investigated pan-genome. The annotation tools used are BlastKOALA (Kanehisa et al., 2016), KAAS (Moriya et al., 2007), KOBAS (Xie et al., 2011) and EggNOG mapper (Huerta-Cepas et al., 2016, 2019).



(b)

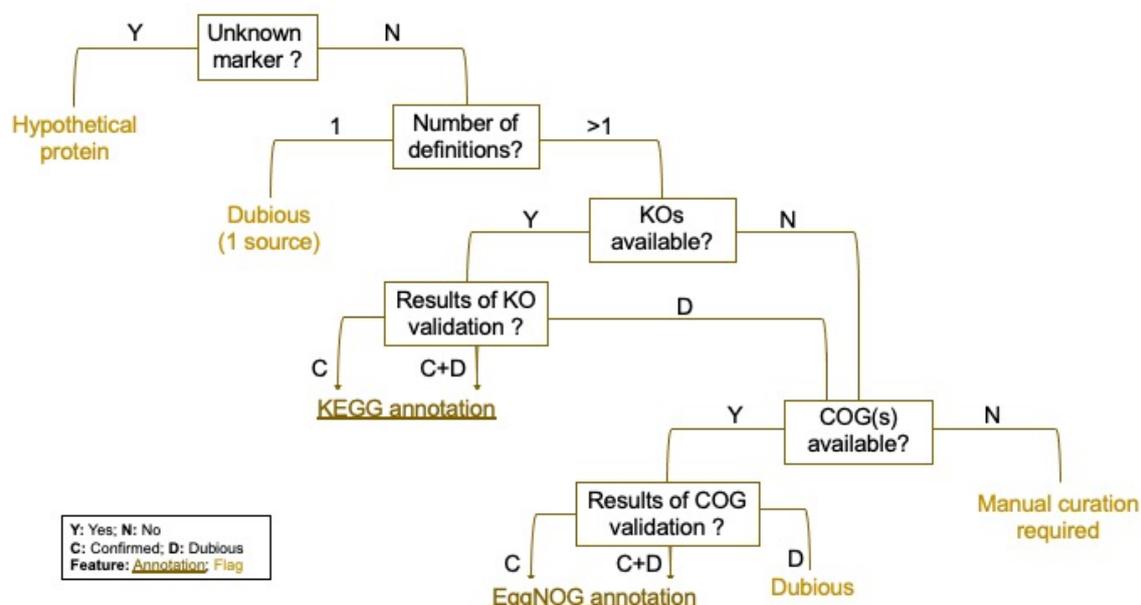


Figure 4-4 Method illustration

(a) Global workflow detailing the scripts in *italic* and describing the outputs at each step
 (b) Decision tree illustrating the final compilation

Several MATLAB scripts were written to pre-process the original annotations as described in **Figure 4-4**. Annotations corroborated by multiple sources or COD confrontation received no flag. If such confirmation was not possible, the annotation was assigned a “Dubious” status or « Dubious (one source) » when only one source returned a result. The combination program returned a “Manual comparison required” flag when no orthologous groups (KO, COG or pan-genomic clusters) were available to automate the comparison. If no matches were found in any of the interrogated databases or the annotation matches a marker from the Unknown list, the protein was assumed to be a “Hypothetical protein”. For COG-validated annotations, EggNOG information was retained. And when the annotation was KO-validated final information was extracted from KEGG database. KO-validation was preferred over COG validation.

4.3 - Results and discussion

4.3.1 - Coverage against representation in the databases

Oleaginous yeasts have been found across various clades of both *Ascomycota* and *Basidiomycota* phyla (Sitepu et al., 2014a) yet there is a clear predominance of the *Ascomycota* strains in the studied group. A preponderance also found when looking at species representation in phylogenomic databases (**Figure 4-5**).

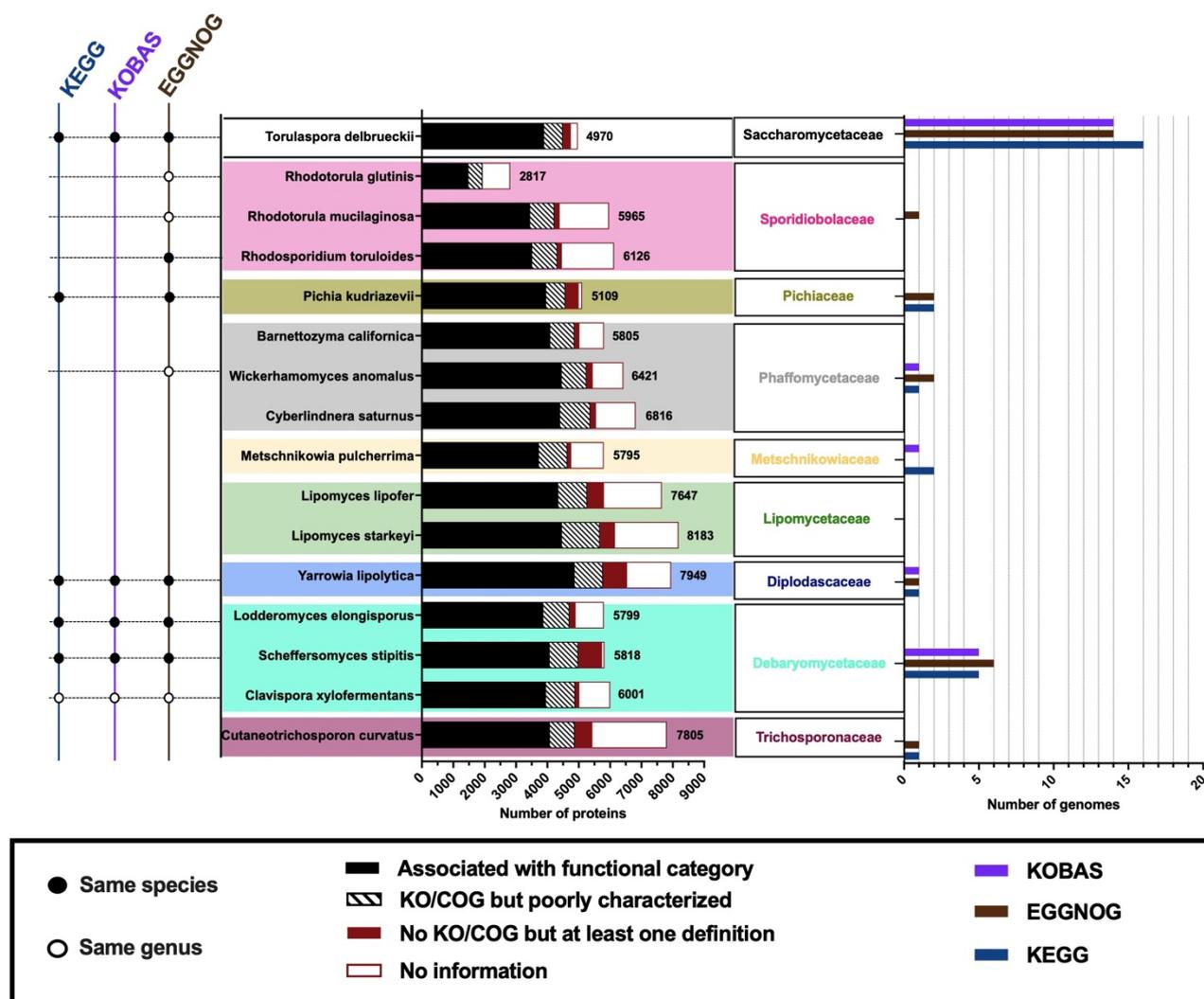


Figure 4-5 Summary of functional annotation quality and phylogenetic representation of unconventional yeast in phylogenomic databases.

Three public databases supported the functional annotation of genomes in this study: KEGG, KOBAS and EGGNOG database. A color was assigned to each database. Yeast representation in these databases was evaluated at species/genus level on the left-hand side of the scheme and the exact number of genomes found in those databases at phylogenetic family level was given on the right-hand side of the scheme. The central part describes the quality of the functional annotation of studied genomes. The total number of protein predicted for each genome was reminded at the end of each bar.

Part II – Methodologies for unconventional yeasts

S. cerevisiae, the most documented yeast, is an *Ascomycetes* species from the *Saccharomycetaceae* family, so research on *Ascomycetes* yeasts was historically privileged (Garay et al., 2014). Since functional annotations are mainly inferred by homology to contents of phylogenomic databases, the taxonomically unbalanced knowledge on yeasts might have consequences on annotation quality considering that more accurate functional information might be obtained from phylogenetically close species. Since the *Saccharomycetaceae* family is the most represented in the databases, a lot of functional information might be inferred from this family so species proximity to the family was evaluated by computing the proportion of each genome homologous to *S. cerevisiae* proteins (**Table 4-1**). Additionally, the phylogenetic distribution of genomes in queried databases was analyzed at both genus and family levels (**Figure 4-5**). For a given species, genus and family representation in the databases, proximity to *Saccharomycetaceae* family and representation of closely related family should be considered in the evaluation of the impact of database contents on the coverage and quality of the functional annotation of studied genomes.

Three databases were routinely interrogated for functional annotation in this study: KOBAS, KEGG and eggNOG databases. Only genomes of *T. delbrueckii*, *Y. lipolytica*, *S. stipitis* and *L. elongisporus* were found in the 3 databases (**Figure 4-5**). But *Y. lipolytica* genome was the only one from the *Diplodascaceae* family while the other 3 species come from families highly represented in the databases (**Figure 4-5**). One genome of *P. kudriazevii* was found in both KEGG and eggNOG while *R. toruloides* was only represented in eggNOG database. *R. toruloides* genome was the only genome of the *Sporidiobolaceae* family although the family is popular for lipids and carotenoids accumulation abilities (Garay et al., 2014). Knowing that *R. toruloides*, *R. glutinis* and *R. mucilaginosa* are from the same genus (**Appendix 1**), the last two have at least one genome of the same genus in the databases (**Figure 4-5**). The 3 *Rhodotorula sp.* and

Part II – Methodologies for unconventional yeasts

C. curvatus are the only *Basidiomycetes* of this study. *C. curvatus*, the species sharing the lowest number of proteins homologous to *S. cerevisiae* (46.56 %) (**Table 4-1**), also belongs to one of the least represented families (*Trichosporonaceae*) (**Figure 1b**). At the other end of the spectrum, *T. delbrueckii* (the only member of the *Saccharomycetaceae* family) (**Appendix 1**) shares 96.04% of its proteins (**Table 4-1**) with *S. cerevisiae* and also benefits from the highest familial representation with 14 to 16 *Saccharomycetaceae* genomes in each database (**Figure 4-5**).

The Debaryomycetaceae family is also quite well-represented with 5 to 6 genomes per database and even a genome of the same genus as *C. xylofermentans* (i.e. *Spathaspora passalidarum*) (**Figure 4-5**). Then the number of genomes per family drops to 2-3 per database for the *Phaffomycetaceae*, *Pichiaceae* and *Metschnikowiaceae* families. The *Pichiaceae* family is equally represented in eggNOG and KEGG databases but absent from KOBAS when the *Metschnikowiaceae* is present in KEGG and KOBAS but not in eggNOG. As for the *Phaffomycetaceae* family, it is represented in the 3 databases with even a genome of the same genus as *W. anomalus* (*Wickerhamomyces ciferrii*) (**Figure 4-5**). *C. saturnus*, *W. anomalus* and *B. californica* are part of the *Phaffomycetaceae* family. This family is usually less represented in the databases (Correia et al., 2019), but recently started to attract interest for its methylotrophs member such as *Komagataella phaffii* (formerly *Pichia pastoris*) (Bernauer et al., 2021) whose genome is actually the only *Phaffomycetaceae* genome found in the 3 databases. Which only leaves one family with no representation in the databases: the *Lipomycetaceae* family. Interestingly, the *Lipomyces* sp. were phylogenetically grouped with *M. pulcherrima* (**Appendix 1**). When *M. pulcherrima*, *L. elongisporus*, *C. xylofermentans*, *S. stipitis*, *S. occidentalis* (*Debaryomycetaceae* and *Metschnikowiaceae*) were described to be part of the CTG clade (Viigand et al., 2018), a taxon characterized by its unique genetic code in which CUG codons encode leucine. Members of the CTG clade seem to share a high homology

Part II – Methodologies for unconventional yeasts

(above 70 %) with *S. cerevisiae* and *M. pulcherrima* makes no exception with a percentage of 73.74 % when the *Lipomyces* sp. only shares between 55 and 58% of their proteins with the model strain (**Table 4-1**). A characteristic *Lipomyces* sp. have in common with *Y. lipolytica* (57.52%) (**Table 4-1**). which is consistent with the fact that despite their apparent phylogenetic distance (**Appendix 1**) they are both considered to be part of the same taxonomic group: the basal Saccharomycotina group (Benini, 2020).

Species of well-represented clades like *T. delbrueckii* or the *Debaryomycetaceae* species exhibited respectively 78.4% and between 69.4 and 70.5% of functionally annotated proteins. Even *P. kudriavzevii* despite the lower representation of *Pichiaceae* was functionally annotated at 77.5% (**Figure 4-5**). Conversely, species from clades with a lesser representation like *Rhodotorula/Cutaneotrichosporon/Lipomyces* species showed proportions of functionally annotated genes between 57.3% and 52.3% (**Figure 4-5**). Yet the functional coverage of *Lipomyces* species is higher than the one of *C. curvatus* despite the absence of *Lipomycetaceae* in the databases. This can be explained by the proximity of the *Lipomyces* sp. to *M. pulcherrima* (**Appendix 1**) and even species from the *Saccharomycetaceae* family to a certain extent as *Lipomyces* sp. exhibits a higher proportion of genes homologous to *S. cerevisiae* than *C. curvatus* (**Table 4-1**). So the difference observed between *Lipomyces* sp. and *C. curvatus*, might just be the reflection of the known unbalance between *Ascomycetes* and *Basidiomycetes* representations. But the relationship between the phylogenetic representation of databases and the proportion of functionally annotated protein is not perfectly linear. Indeed, *Debaryomycetaceae* are more represented than *Pichiaceae* but *P. kudriavzevii* has a higher proportion of functionally annotated genes than all three *Debaryomycetaceae* species (**Figure 4-5**). The same type of conclusion can be drawn when looking at the proportion of genes left unannotated: the 4 *Basidiomycetes* sp. and the *Lipomyces* sp. are the species with the

highest proportion of unannotated genes and *S. stipitis*, *T. delbrueckii* & *P. kudriazevii* have the lowest proportion. But again, the difference between *S. stipitis* and the other two Debaryomycetaceae is striking as *S. stipitis* exhibits 4.6 times more proteins with one match but not associated to any functional category than *L. elongisporus* and *C. xylofermentans* (**Figure 4-5**). This could be the sign of a greater level of information on the strain. If the presence of close relative in the interrogated databases influences the annotation coverage (**Figure 4-5**), it also depends on other factors like the level of characterization of the genomes present in the databases or the quality of gene prediction.

The use of pan-genomic data could make up for the disparities observed in the queried databases. Computed pan-genomic orthologous groups could support the annotation by function inference within pan-genomic orthologous groups rather than based on external databases for example. Indeed, the composition of a pan-genome is much more customizable than orthologous groups from public databases. This could help bypass the poor oleaginous yeast representation in public databases as information could be extrapolated from well-documented species of the pan-genome to less documented ones. For now, in this study, pan-genomic orthologous groups were not used to identify new functions but only to confirm or invalidate functions already at least suggested by one of the public databases. Sources redundancy was previously (Griesemer et al., 2018; Ruiz-Perez et al., 2021) used to improve annotation coverage. In this study, not only the annotation coverage was increased but annotations were also associated with flags resulting of the confrontations of sources (**see section 4.2**). An average of 17.9% of the predicted proteins was left unannotated because no homologs were found in the interrogated databases. When about 63.8 % of the predicted proteins were successfully associated with functional categories from KO and/or Cluster of Orthologous Genes (COG) classifications. The remaining proteins were found to have a match in the

interrogated databases but not characterized well enough to be associated with a function. Most of the annotations (94.3 %) were confirmed by several sources. The remaining 5.7 % comes with lower confidence and might be regarded as dubious.

4.3.2 - Functional characteristics of the pan-genome

The annotation quality can also be evaluated at pan-genomic level. The 16 genomes investigated form a pan-genome of 99 016 genes (**Figure 4-6a**). Interestingly, the majority of the dubious annotations are annotations of unique genes (69.57 % of unique gene annotations are dubious against 0.38 % of core genes and 3.61 % of accessory gene annotations). Species true singularity lies within unique genes, but those genes seem to be more difficultly associated with functions as they also count the highest proportion. The differences in annotation coverage could introduce a bias in the functional pan-genomic analysis considering that genomes with a higher functional coverage could have a higher proportion of functionally annotated unique genes. The genomes with the highest functional coverage are *M. pulcherrima* (78.4%), *S. stipitis* (77.57%), *C. xylofermentans* (70.47%), *C. saturnus* (70.02%) and *R. toruloides* (69.53%) (**Figure 4-6c**). When genomes with the highest proportion of unique annotated (UA) genes are *Y. lipolytica* (9.3%), *C. curvatus* (7.9%), *P. kudriazevii* (6.9%), *L. lipofer* (4.2%) and *T. delbrueckii* (3.82%) (**Figure 4-6c**). This suggests that a higher functional coverage is not correlated to a higher proportion of annotated unique genes. Another bias could arise from the phylogenetic composition of the studied group considering that genes from very close relatives could be sorted into clade-specific accessory clusters rather than into unique genes. The analysis of unique genes should thus be treated with caution.

Grouped functional annotations can be useful to spot singularities. For instance, when the proportion of core genes seems to be the same for most strains, core genes represent a way larger proportion of *R. glutinis* proteins (**Figure 4-6b**). The phenomenon is even

more appreciable when looking specifically at genes associated with metabolic function (Appendix 5).

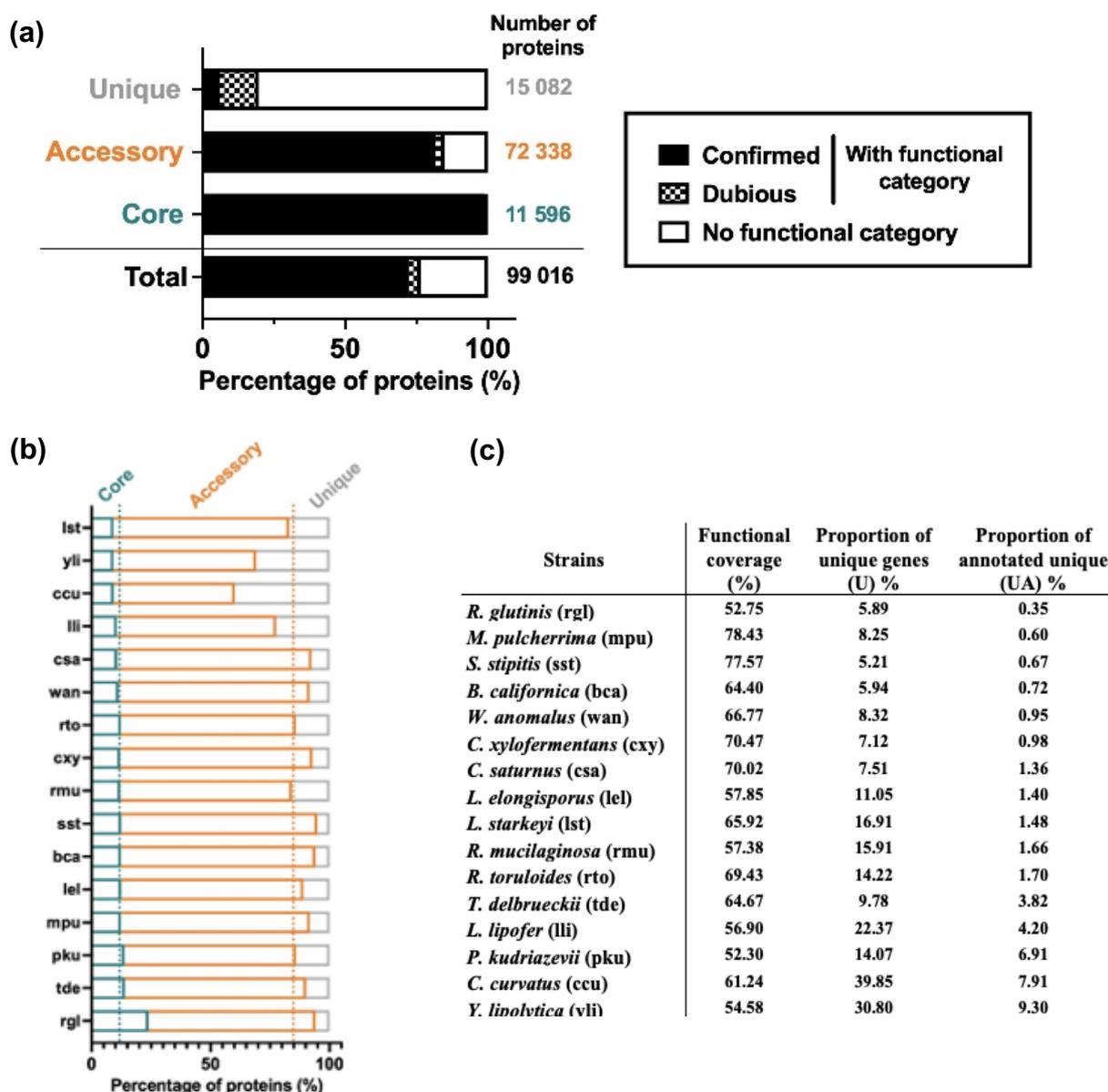


Figure 4-6 Analysis of the pan-genome annotation

(a) Annotation quality with respect to pan-genomic categories. As a reminder: unique proteins are specific to a species, accessory proteins are shared among a few species and core proteins shared across all species considered. (b) Proportion of core, accessory, and unique genes per strain. Percentages given represent fractions of the total number of proteins. Strains are sorted by increasing number of proteins (c) Annotation characteristics. The term “annotated unique” (UA) describes unique genes associated with a functional annotation. Functional coverage, U and UA are fractions of the total number of proteins.

It is also intriguing that the number of proteins of *R. glutinis* (2817 proteins) is about two times lower than the number of proteins of its close relatives *R. mucilaginosa* (5965 proteins) and *R. toruloides* (6126 proteins) for similar genome sizes (20.48 Mb for rgl, 20.75 Mb for rto and 20.28 Mb for rmu). And despite inconsistent protein numbers, *Rhodotorula sp.* do share high GC contents (**Table 4-2**), as expected (Nakase & Komagata, 1971), which suggests that raw sequencing data are of reasonable quality. The low number of proteins of *R. glutinis* must be a result of the methods used for gene prediction as half of the gene predictions (including *R. toruloides* and *R. mucilaginosa*) were performed in this study while existing annotations were downloaded for the other genomes (especially *R. glutinis*).

4.4 - Conclusion: remaining challenges and possible improvements

The aforedescribed method enables the functional annotation of multiple genomes and generates outputs optimized for genome-scale model reconstruction and comparative genomics. The pipeline focuses on the functional annotations but the quality of the input (i.e. gene predictions) used obviously has an impact on the results. I can only encourage one to pick the best annotation available but the lack of standard methods to assess the quality of annotation (Salzberg, 2019; Yandell & Ence, 2012) might make that task quite challenging. Ekblom and Wolf (2014) recommend a qualitative validation and a visual inspection to identify annotation issues while Salzberg (2019) concedes that manual curation cannot always provide an answer. Manual curation is even less conceivable when it comes to analysis large numbers of genomes. In this study, only 16 genomes were compared and which already represents more than 99 000 genes.

Grouped functional annotation offers an alternative to comprehensive manual annotation curation. The method was designed to automatically associate genes with dubious flags to draw attention on potentially problematic annotations that could then, if needed, be subjected to further selective manual curation. Such strategy is particularly

advantageous when working with a high number of genomes as the feasibility of full systematic manual curation decreases with increasing number of genes to review. This approach adheres to Yandell and Ence's recommendations (2012) that suggested an evaluation of the percentage of protein with homology to known domains to assess the quality of annotations. But this kind of approach does not evaluate the likelihood of the predicted structure. Actually, most of the tools routinely used for functional annotation are sequence-based tools when as highlighted by Loewenstein *et al.* (2009), protein 3D structure is often more conserved than their sequences. In the future, the whole annotation process could benefit from integrating more structure-based tools. Meanwhile pipelines could integrate more length-based verifications like the one proposed by the GeneValidator (Drăgan *et al.*, 2016).

Another foreseeable improvement is a slow transition toward more grouped annotations. Indeed, grouped annotation might facilitate error identification and knowledge transfer from one genome to another which should ultimately improve the overall quality of annotations. Furthermore, when dealing with individually annotated genomes, one cannot exclude that part of the differences observed might arise from the annotation method. A transition toward grouped annotation, should ensure a standardization of the annotation method within a group of interest and fairer comparisons. In this spirit, the concept of single reference genome is being challenged in favor of pan-genomic references (Marschall *et al.*, 2018) and protocols describing reference-based gene prediction and annotation (Liu *et al.*, 2018) or multi-genome gene prediction and annotation (Nachtweide & Stanke, 2019) are now available. Yet, current annotation architectures cannot accommodate automated correction propagation (Salzberg, 2019; Yandell & Ence, 2012). Meaning that, if an error is spotted and corrected in one annotation, it does not automatically correct the annotation derived from the reference. For instance, the analysis described in **Chapter 6** revealed that a few dubious *Lipomyces* phospholipases

Part II – Methodologies for unconventional yeasts

C were misannotated because of a homology found with EggNOG seed orthologs 5507.FOXG_07227P0, 5507.FOXG_06598P0 and 5507.FOXG_06661P0 from the EggNOG V4.5. But the erroneous seed ortholog, that was spoiled with transposases traces (source of the mismatch) has since been removed from the database and replaced by 5507.FOXG_03046P0, 5507.FOXG_02246P0 and 5507.FOXG_16985P0 in EggNOG V5. This demonstrates although current annotation tools and methods are not flawless, it is possible to take precautions to limit error propagation. Thanks to EggNOG formidable architecture I was able to track down the source of error, but most tools do not offer the same traceability (Correia et al., 2019) so not only database updates will not be automatically propagated but feedback loops are also impossible when it could greatly contribute to the database curation efforts.

This highlights the importance of considering annotation as “dynamic entities”. The relentless effort of database curators must be saluted for their update work but also for limiting redundancy and slowly improving their interconnectivity. Standardized nomenclatures and identifiers like KOs are gaining in popularity. However other alternatives must be found for annotations not associated with in such nomenclatures. It is likely that text analytics tools (e.g text mining and natural language processing) are going to play an important role in the treatment of non-standardized annotations and maybe just simply contribute to the acceleration of the standardization process like it was observed with EC number databases of which growth can be imputed to text-mining (Chang et al., 2009).

The improvement opportunities are endless. Typically, the more complex and resource-intensive a pipeline is the better accuracy it will achieve. This often happens at the expense of user-friendliness and operating time. So, while developing new annotation tools a trade-off must absolutely be found between speed, accessibility, and accuracy.

Chapter 5 - Reconciling Nile red staining protocols and investigating of the method scalability

5.1 - Introduction

This study combines computational and experimental strategies to explore oleaginous yeasts' potential for industrial applications. While an improved functional annotation method was crucial for the *in silico* analysis, the experimental investigation required a robust screening methods allowing the reproducible monitoring of lipid accumulation and reliable inter-species comparisons. Inter-species comparison is crucial for benchmarking and identification of high-performing strains. Yet most existing methods often require preliminary species-specific optimizations due to species morphological and physiological differences which complicate rigorous inter-species comparisons. The gravimetric determination of lipid content is the most universal method for the comparison of lipid contents across strains. But even for gravimetric determinations, it is recommended to optimize the lipid extraction and especially the cell disruption step while working with yeasts (Zainuddin et al., 2021). Indeed, yeasts (like microalgae) have thick cell walls that has sometimes been described to prevent the release of lipids during extraction (Sitepu et al., 2012). Differences in cell wall/membrane compositions are also problematic when using fluorescent lipophilic dyes to estimate lipid contents because they tend to lead to variable dye diffusion across cell membranes. But at least, fluorescence-based approaches are less time-consuming and tedious than extraction-based procedures. Hence, dye-based indirect estimation of lipid contents has been successfully applied to numerous microorganisms over the past 40 years (Patel et al., 2019; Sitepu et al., 2012).

Among the lipophilic dyes usable for quantitative assays in microplate format, Nile red (9-diethylamino-5-benzo[α]-phenoxazinone) was the only successfully applied to inter-yeast comparisons (Sitepu et al., 2012) but the correlation between fluorometric

Part II – Methodologies for unconventional yeasts

determination and gravimetric measurements was improvable (R^2 of 0.71). Since the use of Nile red (NR) has been reported on a variety of yeasts and fungi, the importance of certain parameters (e.g. cell/NR ratio, staining time, temperature) has already been studied. The significance of other parameters such as the choice of emission wavelength and the cultivation scale remains unclear. A confusion reinforced by discrepancies observed across protocols available. Here, three protocols were chosen as references based on their demonstrated efficacy on strains of various clades. The main parameters of each method were summarized in (**Table 5-1**). The first protocol released in 2004 by Kimura *et al.* highlighted the importance of mixing and suggested some flexibility in OD/NR ratio since various volumes of different cell densities were stained with the same amount of Nile Red. This protocol (Kimura et al., 2004) was designed for consecutive individual sample readings. Hence it didn't include any replicates although it would have enabled a better evaluation of data reliability and a possible statistical comparison. This first protocol was found non-reproducible by Sitepu *et al.* (**2012**) that proposed major improvements to the method while investigating the relevance of parameters such as: temperature, OD/NR ratio, addition of a carrier (e.g., DMSO), media composition, effect of washing. Yet to this day, the first method remains almost twice more cited than the second protocol despite the reedition of the latter in 2019 (Sitepu et al., 2019).

Sitepu *et al.* (2012) designed a microplate-based method including replicates and an OD-based fluorescence correction. They also suggested the use of DMSO to improve strain permeability to NR while emphasizing the strain-specificity of the effect. Again, OD/NR ratio proved to be quite flexible, but the ideal OD_{600} (for a final NR concentration of 5 $\mu\text{g}/\text{mL}$) was set to 1 because it gave the highest signal and probably for normalization purposes. The main bottleneck that Sitepu *et al.* (2012) identified in the first protocol was the variability of time needed to reach fluorescence maximum across different species

Table 5-2 Summary of Nile red staining protocols published between 2004 and 2021 for the estimation of yeast lipid contents
Columns in grey correspond to the latest development while the other 3 columns describe the reference methods.

Author	Volume cultivated	Harvesting and washing	Permeabilization method	NR solvent	Final NR concentration	Sample preparation	Mixing	Staining time	Equipment
Kimura <i>et al.</i> (2004)	100 mL	None	None	Acetone	1 µg (~0.5 µg/ml)	50-100 µL of 0-12 DCW (mg/ml) 2 mL PBS (pH 7) 10 µL NR	In tube	5 min	Spectrofluorometer
Sitepu <i>et al.</i> (2012, 2019)	100 mL	None	DMSO	Acetone	4.2 µg/ml	250 µL OD ₆₀₀ = 1.0 25 µL DMSO:Medium (1:1, v/v) 25 µL of Nile red	Pipetting up and down	20 min kinetic reading	Filter fluorometer (top 50% mirror)
Rostron <i>et al.</i> (2017)	20-50 mL	Washing and resuspension in PBS	DMSO	Acetone	5 µg/ml	250 µL OD ₅₉₅ = 1.0 25 µL DMSO:PBS (1:1, v/v) 25 µL of Nile red	-	20 min kinetic reading	Filter fluorometer (top 50% mirror)
Ramirez-Castrillón <i>et al.</i> (2021)	23 mL	Resuspension in various mixtures	Carrier (variable)	Acetone	25 µg/ml	150 µL OD ₆₀₀ = 0.03/1.0 50 µL of Nile red	Microplate shaking	5 min orbital shaking followed by 1h10 kinetic reading	Spectrofluorometer
Miranda <i>et al.</i> (2020)	On solid media	Washing and resuspension in PBS	DMSO	Acetone	5 µg/ml (assumption)	817 µL OD ₆₀₀ = 1.0 83 µL DMSO:PBS 5 µg of Nile red	In tube	20 min kinetic reading	Filter fluorometer
Zhao <i>et al.</i> (2019)	50 mL	Washing and resuspension in PBS	Preheating 50°C followed by cool-down	Acetone	1 µg/ml	OD ₆₀₀ = 1.0 in PBS Nile red	In tube	5 min in time	Spectrofluorometer
Hicks <i>et al.</i> (2019)	10 mL	Resuspension in fresh media	DMSO	Not specified	15 µg/ml	200 µL O OD ₆₀₀ = 1.0 40 µL DMSO Nile red	Microplate shaking	10 min at 30°C	Filter fluorometer

To be continued on following page..

	Container	Reading mode	Excitation wavelength (λ _{ex})	Emission wavelength (λ _{em})	Single time reading	Repeated measurements		Data processing	Other recommendations
Kimura <i>et al.</i> (2004)	Cuvettes	Spectrum	488 nm	400 - 700 nm	Single time reading	No replicates (cuvettes)	-	Blk = Unstained cells Data = signal - Blk	No removal of NR excess needed Vigorous mixing before reading
Sitepu <i>et al.</i> (2012, 2019)	96-wells black microplates with clear bottom	Endpoint	530/25 nm	590/35 nm	21 readings	4 technical replicates	-	Blk = Unstained cells Data = (signal-blk)/OD ₆₀₀	No need to wash cells with PBS Initial reading OD ₆₀₀
Roston <i>et al.</i> (2017)	96-wells black microplates with individually molded wells	Endpoint	485 nm	595 nm	21 readings	3 technical replicates	-	Blk = Mix with PBS (average of 6 wells) Data = (signal/Blk)/OD ₅₉₅	NR solution protected from light stored at 4°C or -20°C. Fresh working solution before each experiment
Ramirez-Castrillón <i>et al.</i> (2021)	96-wells black microplates with clear bottom	Endpoint	488 nm	585 nm	8 readings	3 technical replicates	-	Blk1 = Mix without cells Blk2 = Unstained cells Data = signal - (Blk1+Blk2)	Monitor fluorescence until fluorescence stabilization to obtain accurate measurements
Miranda <i>et al.</i> (2020)	96-wells black microplates	Endpoint	530 nm	590 nm	21 readings	3 technical replicates	3 biological replicates	Blk = Mix without cells Data = (signal/OD640)-Blk1	Control samples included in each microplate to verify the consistency between readings
Zhao <i>et al.</i> (2019)	Not specified	Endpoint	488 nm (Em slit = 10 nm)	570 nm (Ex slit = 10 nm)	Single time reading	3 technical replicates	-	Blk1 = Mix without cells Blk2 = Unstained cells Data = signal - (Blk1+Blk2)	Staining in the dark
Hicks <i>et al.</i> (2019)	96-wells black microplates	Endpoint	530/25 nm	590/35 nm	21 readings	3 technical replicates	4 biological replicates	Values corrected by cell count	Cell count more accurate than OD ₆₀₀

due to putative different permeabilities to NR. To bypass this limitation, Sitepu *et al.* (2012) recommended a kinetic reading (repeated measurements over time) and compared the strains based on the highest fluorescence signals recorded independently of the staining time. The last reference protocol proposed by Rostron *et al.* (2017), suppresses the variability related to media utilization by analyzing cells suspension in phosphate-buffered saline (PBS) and preconizes 10-sec plate shaking before reading. Their modifications of the method were probably focused on increasing the reproducibility because their other recommendations are related to the type of plate, the wavelength used, data processing and NR solution conservation & handling. The goal of this chapter was to reconcile existing reference protocols by investigating the significance of ambiguous parameters, highlighting the importance of underestimated specifications, and proposing solutions or alternatives when applicable.

Overall, the parameters defining this staining method can be either related to sample preparation, data acquisition or data processing. Sample preparation parameters seem to be the less stringent ones since a given amount of NR can stain various cells concentrations of different lipid contents. Yet, all three procedures described in-flask cell cultivation with working volumes varying from 20 to 100 mL when no more than a few milliliters are needed for the assay. Furthermore, the development of high-throughput approaches set a miniaturization trend when it comes to screening methods, so the scalability of the method was assessed. Regarding data acquisition, Kimura *et al.* (2004) recommended a single-time spectrum acquisition (400-700 nm) when the other two procedures are based on kinetic endpoint readings (single emission wavelength and multiple recordings over time). The first approach takes into account the differences in emission maxima between species while the second considers the differences in permeability to Nile red. A combination of both approaches would surely improve the comparison accuracy. Hence, an alternative to kinetic reading was investigated but the

original approach gave superior results, so the importance of wavelength selection was evaluated. Additionally, good practices pertaining to data acquisition were identified. All the findings were then summarized in a coherent modified method applicable to inter-species comparison and with a few amendments for the monitoring of lipid accumulation. Finally, the relevance of the proposed method was discussed with respect to the latest development in Nile red staining of unconventional yeasts (**Table 5-1**).

5.2 - Material and methods

5.2.1 - Cell culture

For the optimization of Nile red staining strains were mostly grown in glucose synthetic complete with a C/N ratio of 60 (GlcSC) except for the assay described in 1.2.5.g cells were grown in sweetwater (SW15). Media preparation, cell maintenance and culture were performed as described in **Chapter 3 (section 3.2)**.

5.2.2 - Plate testing and reading position selection

Two plates were compared by reading the signal emitted by 8 wells containing: 250 μ L of PBS, 25 μ L of Nile red solution and 25 μ L (60 μ g/mL) of PBS:DMSO (1:1, v/v). Samples were excited at 489 nm (λ_{ex}) and the emission wavelength was set at 535 nm (cutoff 530) (λ_{em}). Signals were measured every 60 sec for 20 min. With the clear bottom plate, both top and bottom readings were performed while only top reading was performed on the plate with solid bottom.

5.2.2.a - Modification of the general procedure

For the following experiments, the general procedure described in **Chapter 3 (section 3.3.1)** was modified as described in **Table 5-3**.

Table 5-3 Summary of modifications made to the general procedure

Detailed in subsection	5.2.3.a	5.2.3.e	5.2.3.c	5.2.3.f	5.2.3.b	5.2.3.d	5.2.3.g
Experiment	Lid and pipettes	Hydrolysis	Blk and scale	Wavelength with preheating	Rows vs columns	Linearity against OD ₆₀₀	Final adjustments
Cultivation scale	-	Tubes or flasks	Tubes or flasks	Tubes	Tubes	Tubes	Flasks
Sample OD ₆₀₀	-	n.d.	1	1	1	1, 2.5, 5, 6.5, 9	5
[NR]final µg/mL	25	25	10	10	5	5	5
Assay volume (µL)	300	300	300	300	200	200	200
λ _{ex} (nm)	530	530	488	488	488	488	488
λ _{em} (nm)	590-700 620	620	520-760	520-760	530	535	535 and 625
Readings	At t = 0, 10, 20 min	Every 20 min for 1h	At t = 0, 20, 30 min	At t = 0, 24, 30, 40, 60 min	Every min for 20 min	Every 3 min for 2h	Every 5 min for 30 min

n.d. not determined ; - not applicable

5.2.2.b - Effect of lid presence and pipette comparison

HEAR oil diluted 20 times with DMSO was used as sample for both experiments. The effect of lid presence was evaluated using a clear polystyrene lid. For the impact of lid presence, the spectral mode was used repeatedly over time and the highest fluorescence signal was plotted regardless of the staining time. For pipette comparison, endpoint mode was used, and the results were plotted against staining time. The assay mixture was only made of 250 µL sample (or control) and 50 µL of Nile working solution (150 µg/mL) for a final volume of 300 µL. Four replicates were analyzed for each sample.

5.2.2.c - Comparison of row and column sample loading

L. elongisporus sample horizontal and vertical loading were compared based on the resulting relative variability across replicates.

5.2.2.d - Selection of a suitable blank and cultivation scale

To investigate blank composition and cultivation scale, *R. toruloides* (12), and *Y. lipolytica* (20) were grown in tubes and flasks. Candidate blanks were unstained cells, assay mixture without cells nor Nile red and assay mixture with Nile red but not cells. Candidate blanks were prepared to have the same final volume as assayed samples. A volume of 250 μL of cell suspension (or candidate blank) and 50 μL of Nile working solution (60 $\mu\text{g}/\text{mL}$) was used for a final assay volume of 300 μL . No replicates for this preliminary experiment.

5.2.2.e - Linearity against OD₆₀₀ for small cultivation scale

R. toruloides (13), *S. stipitis* (15), *W. anomalus*, *Y. lipolytica* (20) and *L. elongisporus* cell suspension were diluted with PBS to OD₆₀₀ of 1, 2.5, 5, 6.5, 9. Aware that permeabilities to Nile red might vary depending on the growth phase, the samples of different cell densities were all prepared by dilution of the same parent sample.

5.2.2.f - Exploring hydrolysis as a permeation method

Washed pellets were thawed and vortexed for 60 sec. One milliliter of aqueous HCl (1M) was added to thawed pellets and mixtures were vortexed for 60 sec. Tubes and a control tube containing only 1 mL of aqueous HCl (1M) were then incubated at 80°C (1050 rpm) for 30 min. Samples were neutralized by addition of 1 mL NaOH (1 M) followed by 60-sec vortexing. Despite sodium hydroxide addition, sample pH remained slightly acidic in tubes formerly containing cells probably because of the contribution of endogenous organic acids. Cells suspensions (and controls) were diluted 1.3, 2, 4, 10 and 20 times with water. Given the introduction of HCl and NaOH for the hydrolysis, the salt concentration was deemed sufficient for the assay and no additional PBS was added. Hence, for the assay, only 250 μL of neutralized cell lysates (or control) and 50 μL of Nile working solution (150 $\mu\text{g}/\text{mL}$) were used for a final assay volume of 300 μL .

For the attempt on 20 stains, cells were grown in flasks in conditions that matches cell growth for lipid extraction and TLC (4 days in GlcSC₆₀).

5.2.2.g - Wavelength comparisons: procedure with preheating

Cell suspensions of OD₆₀₀ 5 were prepared for *R. toruloides* (12), *C. curvatus*, *S. stipitis* (15 and 16), *Y. lipolytica* washed cell pellets. For each strain, a tube containing 300 µL of cell suspension and 24 µL of DMSO was preheated at 70°C (300 rpm) for 10 min. Then, 250 µL of preheated suspension was used as sample for the assay but not additional DMSO:PBS was added so that the assays were only composed of 250 µL and 50 µL of Nile Red solution (60 µg/mL). No replicates for this preliminary experiment.

5.2.2.h - Final adjustments to the optimized method

Method for interspecies comparisons (final wavelength comparison) - Strains were grown in tubes containing 5 mL of SW. One milliliter of culture was harvested after 3 days for each strain. However, for some strains, growth after 3 days did not yield enough biomass for Nile red staining, so no data is available for these strains. The assay was performed as aforescribed but instead, of spectrum acquisition, the fluorescence was monitored at 625 nm and 535 nm at each endpoint reading. The RFU_{cx} of 21 NR-stained strains were sorted into quartiles to compare strain ranking when the fluorescence was read at an emission wavelength of 625 nm or 535 nm. Quartiles were computed using Microsoft Excel. Were placed in the first quartile values falling between the minimum value and the 25th percentile. Values falling between the 25th percentile and the median value were assigned to the second quartile and the ones above the median value but lower than the 75th percentile were assigned to the third quartile. Finally, values higher than the 75th percentile formed the fourth quartile.

Method for lipid accumulation monitoring - To monitoring of lipid accumulation, *Y. lipolytica* (21), *L. elongisporus*, *R. toruloides* and *W. anomalus* were grown in flasks. Samples (volumes between 250 and 750 µL) were harvested, washed, and stored at -80°C

routinely over 10 days. The Nile red strain was performed as aforescribed making sure to record the sample OD₆₀₀, resuspension volumes and dilution factors (when applicable). The latter information was then used to normalize the fluorescence signals.

5.2.3 - Statistical analysis

The relative variability (%) is defined as the ratio between the standard deviation and signal mean expressed as a percentage. When need unpaired t-tests were used to evaluate the statistical ($p \leq 0.05$) difference between compared parameters.

5.3 - Results

The quantitative use of Nile Red staining has been extensively studied (especially in microalgae) and variability in fluorescence readings has been recurrently reported and was the object of several method optimization endeavors (Cirulis et al., 2012; Hicks et al., 2019; Patel et al., 2019). While non-cellular events (Cirulis et al., 2012), Nile Red photobleaching (Lamprecht & Benoit, 2003), manual handling (Morschett et al., 2016) and mixing (De la Hoz Siegler et al., 2012) were identified as major sources of variability, De la Hoz Siegler *et al.* (2012) also observed an “inherent variability of fluorescence reading”. Some of that inherent variability arises directly from the instrument and might be difficult to remove but some manageable parameters might also contribute. This work focuses on the identification of potential sources of variability among accessible settings and parameters while investigating the importance of specifications varying a lot across protocols.

5.3.1 - Parameters related to data acquisition

Some recommendations regarding the type of plates are available. But to the best of my knowledge, the effects of plate type, sample position on the plate and lid presence, on variability were investigated for the first time in this section.

5.3.1.a - Plate and lid

Rostron *et al.* (2017) suggested that the use black plates with individually molded flat clear bottom might reduce fluorescence variability by preventing well-to-well crosstalks typically observed with common black plate clear bottom.

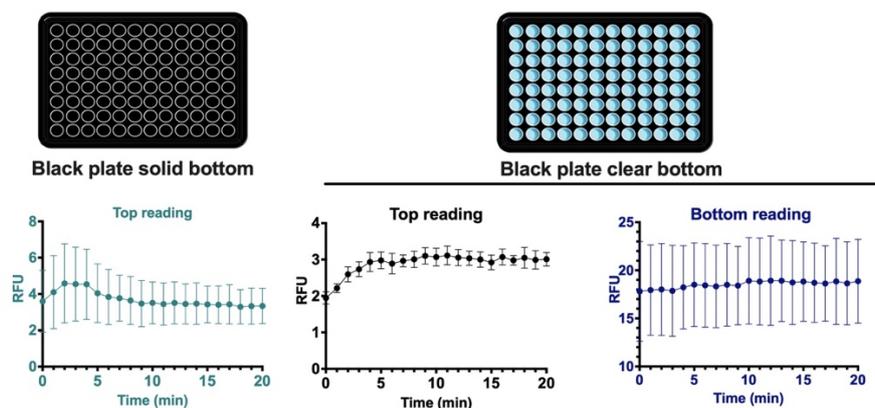


Figure 5-1 - Plate and reading position selection.

Plots representing the evolution of relative fluorescence units (RFU) of 8 technical replicates over time. The mean RFU of replicates is presented here with the associated standard deviation (SD).

The in-house comparison plate comparison (

Figure 5-1) confirmed that high fluorescence was observed with black clear bottom plates when a bottom reading is performed. But using a top reading with this same plate decreases the average variability by 68.98%. When top reading of a black plate with opaque bottom and individually molded wells increases the variability by 4.9times compared to the clear bottom plate. Black clear bottom plate and top reading were thus selected for subsequent steps. The effect of lid presence was also tested.

Covering the plate with a polystyrene lid does not seem to increase inter-replicate variability but decreases the recorded fluorescence intensity (**Figure 5-2a**). Yet, the manufacturer of the spectrofluorometer does not recommend using a plate without a lid when the plate contains an organic solvent or corrosive chemical to preserve the instrument integrity especially if the plate needs to remain in the instrument for a long periods of time (Molecular devices, 2021). To reduce risks, the volatile acetone was

replaced by DMSO in the preparation of Nile Red working solution. This change was also motivated by the observation that acetone evaporation at room temperature might be a source variation leading to the concentration of the Nile Red working solution over time. Later, the total assay volume was also lowered from 300 μL to 200 μL to prevent any precipitation during the vigorous plate agitation.

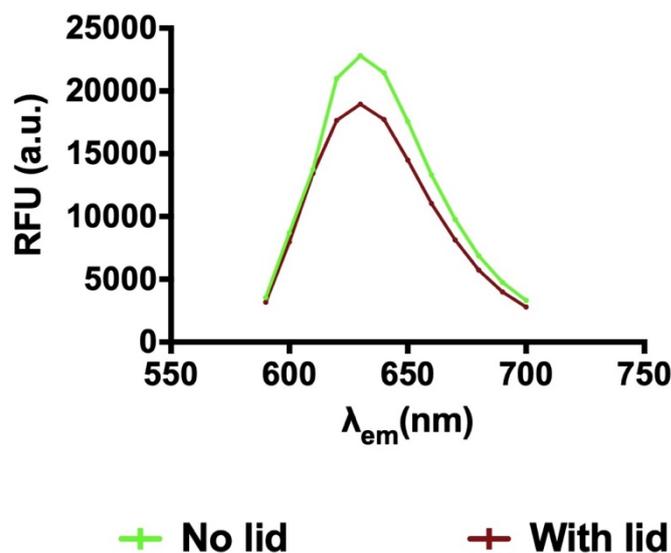


Figure 5-2 - Influence of lid presence.

The mean of replicates ($n=3$ without lid and $n=4$ with lid) and the associated standard deviations are plotted. Error bars are not visible when they are shorter than the size of the symbol.

5.3.1.b - Sample relative position on the plate

The following experiment was led on cellular samples using 4 replicates to mimic real acquisition conditions. The goal was to determine whether relative position of samples on the plate could be a source of variability. The relative variability (standard deviation divided by the mean of 4 replicates) was used as comparison criterion. Loading replicates consecutively in the same row leads to significantly ($p\text{-value} < 0.0001$) more relative variability than loading the replicates in the same column whether the reading mode is spectra or endpoint (**a**). This variability seems to be slightly dependent on the emission wavelength (**a**) and staining time (**b**). The dependency on the staining time is true for

both spectral and endpoint modes although it was chosen to only represent the final time point in **a**. The inter-replicate variability of samples consecutively plated in column shows a stronger dependency to emission wavelength and staining time which results in more dispersed coefficients of variability. Indeed, for samples plated in column the variability will range from 0.27 to 10.97 % of the signal when it represents 4.32% to 10.85% of the signal when samples are plated in row.

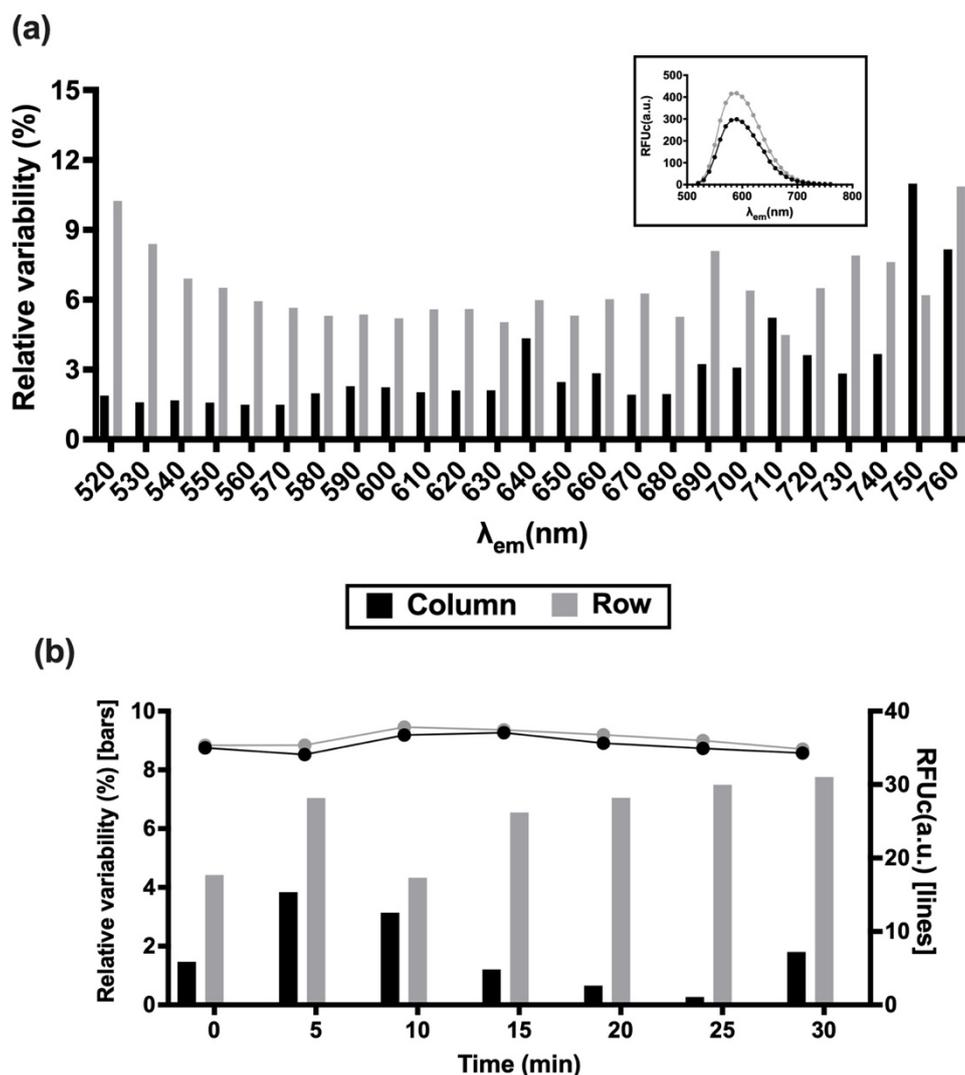


Figure 5-3 - Influence of the relative position of samples on the plate.

Technical replicates ($n=4$) were either plated horizontally (in a row) or vertically (in a column) (a) During spectral acquisition (b) Over time during endpoint readings. In both cases, the evolution of fluorescence signals is given in the upper part of the plots as lines. When the relative variability (%) across replicate was plotted as bars.

This difference suggests that the variability observed in both cases might arise from different sources. Since it might be difficult to identify and suppress the origin of the variability, sample loading in column, that gives more consistent results, should be favored over loading in row. Furthermore, to ensure that no samples were horizontally adjacent it was decided to only load samples in one of two columns limiting the number of samples to 6 per plate since each column can hold 4 blank replicates and 4 sample replicates (total of 8 wells).

Limiting the number of samples per plate might slightly reduce the method throughput but it will not only reduce the variability but should also improve the reproducibility by ensuring a reduced delay between Nile red introduction in the first and last column.

5.3.1 - Parameters related to assay preparation

The main source of variability pertaining to assay preparation are pipetting and mixing. The effect of mixing was evaluated by De la Hoz Siegler *et al.* (2012). They concluded that the variability could be reduced either by in-tube pre-mixing or by a vigorous in-well mixing with assay volume decrease. As for pipetting, Morschett *et al.* (2016) demonstrated that robot-assisted liquid handling was preferable to manual operations. Yet, for labs not equipped with this type of automated systems, monochannel and multichannel pipettes were compared in this section. Additionally, the importance of two other features related to assay preparation was investigated: blank composition and cultivation scale.

5.3.1.a - Pipettes comparison

Both reference protocols n°2 and 3 (Rostron & Lawrence, 2017; Sitepu *et al.*, 2012) highlighted the importance of good pipetting practices in the success and accuracy of the assay. When Rostron *et al.* (2017) recommends using the same pipettes for the assays, Sitepu *et al.* (2019) suggest the utilization of multichannel pipettes. Here,

monochannel (regular) and multichannel pipettes were compared. No significant difference was found between both (**Figure 5-4**). The use of multichannel pipettes leads to slightly more relative variability (on average 2.5 %) than the regular pipettes (relative variability of 1.8%). But multichannel seem to give more stable signals over time. The type of pipette is not a crucial parameter. Yet, multichannel pipettes offer an undeniable kinetic advantage since it allows the simultaneous introduction of Nile Red solution in the 4 blanks and 4 sample replicates which probably contributes to the slightly better signal stability over time.

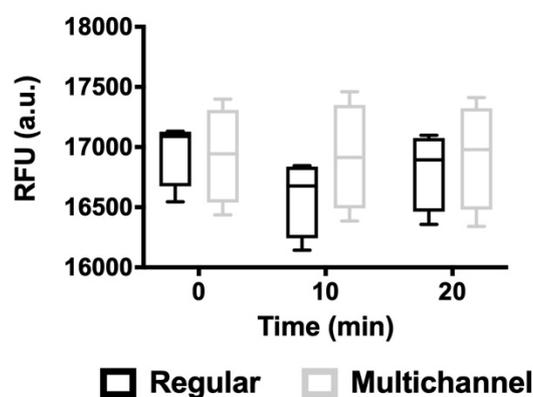


Figure 5-4 - Comparison of multichannel and monochannel pipette for assay preparation.

The box and whiskers plots represent the maximum, minimum and median value of 4 technical replicates.

5.3.1.b - Blank composition

When comparing Nile Red staining protocols different types of blanks or controls are proposed along with different ways to correct the data (**Table 5-1**). Among the most common controls can be found unstained cells, assay mix without cells nor Nile Red and assay mix with Nile Red but without cells. Unstained cells are usually used for background correction of Nile red experiments on microalgal cells due to chlorophyll fluorescence that can interfere with the reading (Hounslow et al., 2017; Huang et al., 2009). The use of unstained cells is even more relevant when the protocols include the removal of Nile red excess after staining and before reading (Cooksey et al., 1987). Yeast

cells being devoid of chlorophyll the necessity of using unstained cells as blank is questionable. Yet, the first reference protocol in yeasts (Kimura et al., 2004) uses unstained cells as control considering Nile red fluorescence background as neglectable since the dye fluorescence tends to be quenched in water. In this protocol (Kimura et al., 2004), an emission wavelength of 488 nm was used and the peaks resulting from the fluorescence of Nile Red interacting with cellular lipids was recorded between 565 and 585 nm for the 6 species tested. Since Kimura *et al.* (2004) performed an emission scan between 400 and 700 nm, a second peak that could be assigned to cells autofluorescence was described around 480 nm.

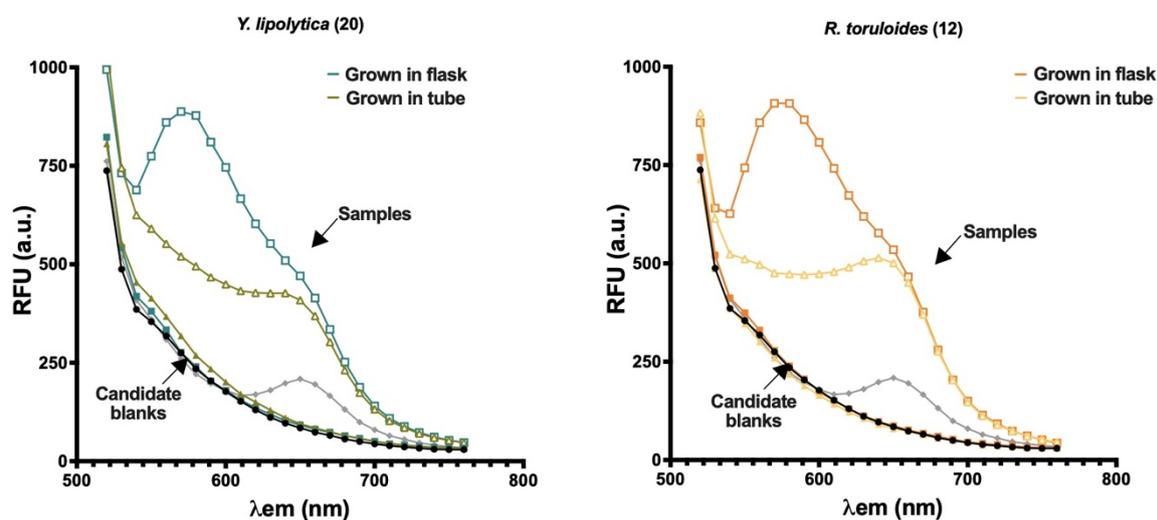


Figure 5-5 - Choice of blank composition.

Fluorescence spectrum recorded for: samples prepared with cells grown in flasks (\square); unstained cells grown in flasks (candidate blank) (\blacksquare); samples prepared with cells grown in tubes (\triangle); unstained cells grown in tubes (candidate blank) (\blacktriangle); mix without cells and Nile red (candidate blank) (\bullet); mix with Nile red but without cells (candidate blank) (\blacklozenge).

Since cell autofluorescence and NR-stained lipid peaks are quite distant, the study didn't demonstrate the relevance of using unstained cells as blank. Some would argue that the use of unstained cells as blank enable events of light scattering by cells to be considered. But are such events dominant over Nile Red fluorescence background? Nile Red background could take two forms: fluorescent aggregates or unspecific interactions.

Indeed, Nile Red has been described to form fluorescent aggregates (Dutta et al., 1996). As for unspecific interactions they were neglectable in Kimura's experiments (2004) because the assay was mainly aqueous but the introduction of organic element (i.e. DMSO) might alleviate the fluorescence quenching reported in water (Sackett & Wolff, 1987). It is probably why Rostron et al. (2017) recommended the use of acellular blanks containing Nile Red.

The most common types of blanks were compared in an experiment where the excitation wavelength was set at 488 nm and emissions were read from 520 to 760 nm. The experiment was conducted using two different strains (including a carotenoid producer) grown in 250-mL flasks and 50-mL tubes to generate cells of different lipid contents and possibly shapes (**Figure 5-5**). The mix without cells but with Nile Red (in DMSO) stands out for its peak at 650 nm which slightly interferes with NR-stained lipid peaks for both strains (**Figure 5-5**). The use of inadequate background correction could lead to misinterpretations and is crucial for cell with lower lipid contents. The mix without cells but with Nile Red (in DMSO) was adopted as blank and corrected data are available in **Figure 5-6**.

5.3.1.c - Cultivation scale

Even though only a few milliliters of cell suspension are required for the assay, the three reference protocols (Kimura et al., 2004; Rostron & Lawrence, 2017; Sitepu et al., 2012) recommend growing the cells in flasks with culture volumes varying between 20 and 100 mL. More developments regarding Nile red staining are available for microalgae than for yeasts (Zhao et al., 2019). And even in microalgae, only one reference reports downscaled cultures in 24-well plates (Chen et al., 2011). This is intriguing considering that growing cells in smaller cultivation vessels might improve the throughput of the assay and/or accuracy. Hence, the importance of the cultivation scale was studied and requirements in terms of cell density were checked. For this purpose, *R. toruloides* (12) and *Y.*

lipolytica (20), two strains belonging to 2 well-documented oleaginous species, were grown in 250-mL flasks (50 mL working volume) and in 50-mL centrifuge tube (5 mL working volume) in nitrogen limiting condition to promote lipid accumulation. For the assay, cells suspensions of the same OD₆₀₀ were stained with the same amount of Nile Red in the 4 cases (Figure 5-6).

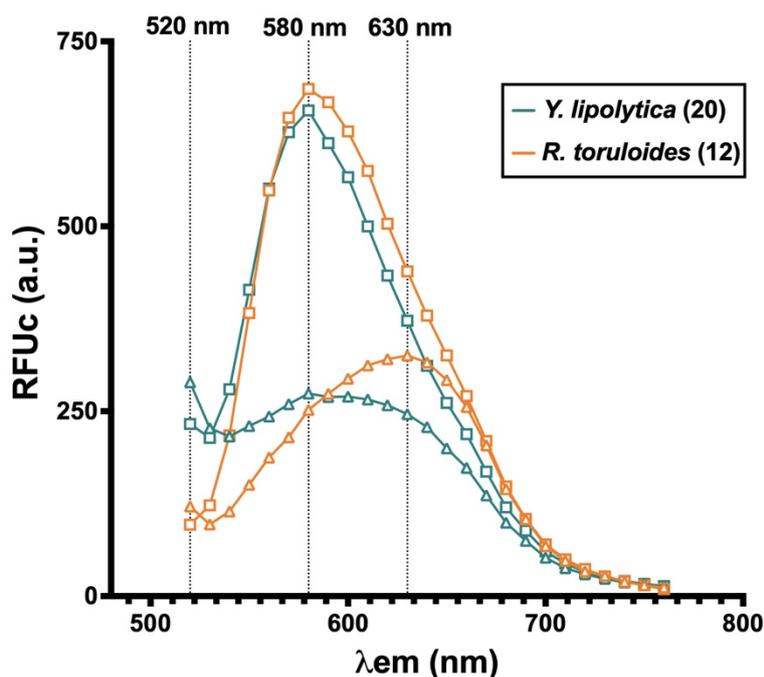


Figure 5-6 - Investigating the effect of cultivation scale on the fluorescence of NR-stained *R. toruloides* (12) and *Y. lipolytica* (20).

Background corrected fluorescence (RFUc) spectrum of NR-stained samples prepared with cells grown in flasks (□) or with cells grown in tubes (△). Strains were both grown under nitrogen limiting conditions either in 250-mL flasks or in 50-mL tubes.

The cultivation scale seems to influence the lipid composition of both strains as suggested by the marked shift in emission maxima observed for *R. toruloides* (12) going from 580 nm (flask) to 630 nm (tube) and by the increased emission at 520 nm when *Y. lipolytica* is grown in tube compared to cells grown in flasks. Based on emission maxima, the NR-stained lipid signal is 2.3 to 2.1-times less intense when cells are grown in tubes as opposed to flasks (Figure 5-6). A change in culture scale might result in variations in lipid content that might alter peak definition but, in the tested conditions, the lipids

accumulated in 50-mL tubes remained detectable by Nile red staining. To make sure of the method sensitivity with samples grown at smaller scale, samples of different cells densities were prepared with different species grown in 5-mL cultures. The species were selected to offer a range of different accumulation abilities.

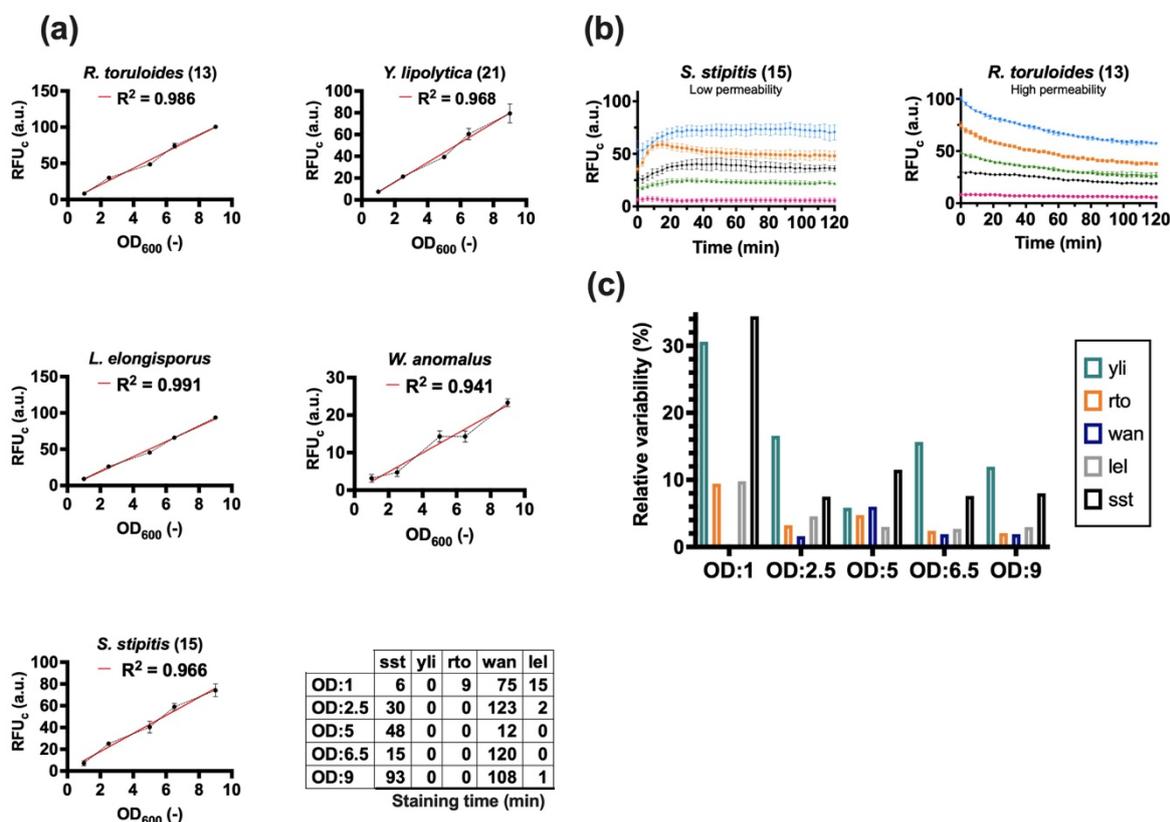


Figure 5-7 - Behaviour during Nile red assay of 5 yeasts grown in 50-mL centrifuge tubes.

(a) Linearity of background corrected fluorescence signal (RFU_c) against OD₆₀₀ and table summarizing the staining times required to reach the emission maxima (b) Evolution over time of the fluorescence of different concentrations of NR-stained *R. toruloides* (13) (example of strain with high permeability to Nile red) and *S. stipitis* (15) (example of strain with low permeability to Nile red).

Even when smaller volumes of culture are grown, the method can accommodate a range of OD₆₀₀ with good linearity between OD₆₀₀ and fluorescence intensities since correlation coefficients between 0.94 and 0.99 were obtained (Figure 5-7a). In the tested conditions, the poorest lipid accumulator seems to be *W. anomalus* for whom recorded intensities (RFU_c) do not exceed 23.3 (a.u.) when the other stained strains exhibited maximal fluorescence intensities between 74.1 and 100.5 (a.u.) (Figure 5-7a). Despite

the good linearity, the method did not enable the distinction between NR-stained *W. anomalus* of OD₆₀₀ 1 versus 2.5 and 5 versus 6.5. Indeed, no significant differences were found between these pairs of fluorescence intensities (**Figure 5-7a**) indicating decreased sensitivity for lower lipid contents.

Repeated measurements over time were performed (kinetic reading) and strains were compared based on their emission maxima regardless of the time required to reach it (**Figure 5-7**). The observation of fluorescence signals over time for samples of different cell densities revealed different behaviors. *W. anomalus* and *S. stipitis* (15) require long staining times, respectively from 75 to 108 min and from 6 min to 93 min. When the emission maximum was reached within the 20 first minutes for all concentrations of *L. elongisporus*, *R. toruloides* (12) and *Y. lipolytica* (21). It should be noted that the time is considerably reduced for *S. stipitis* (15) if, instead of considering the time required to reach the maximum, is considered the time to reach a plateau (**Figure 5-7b**). Yet the evolution of NR-stained *W. anomalus* and *S. stipitis* (15) fluorescence over time follows a sigmoid trend when an exponential decay is observed for the other strains (**Figure 5-7b**). This suggests that *W. anomalus* and *S. stipitis* (15) have a lower permeability to Nile Red (or are less sensitive to DMSO permeation). Looking at **Figure 5-7b**, it seems like strains with lower permeability to Nile Red exhibit higher variability in readings. But no correlation was found since the strains exhibiting the highest average relative variability (16.1 %) is *Y. lipolytica* (21) and the lowest is *W. anomalus* with an average of 2.3% of relative variability (**Figure 5-7c**). Interestingly, most of the strains exhibit the highest variability in readings when they are diluted to an OD₆₀₀ of 1 with relative variability varying from 34.4% (*S. stipitis* (15)) to 9.8% (*L. elongisporus*) except for *W. anomalus* exhibiting its highest variability (6%) when diluted to an OD₆₀₀ of 5 (**Figure 5-7c**).

Fluorescence data corresponding to lower cell densities (OD_{600} of 1) seems less reliable when cells are grown in tubes compared to cell grown in flasks. The data was background corrected so the signal is not noise but it is extremely low and usually associated with higher standard deviations. It should be reminded that the reference protocol n°2 is based on the analysis of samples grown in flasks, diluted to an OD_{600} of 1 and gives consistent results. Hence, flask-cultivation might be preferable to analyze lipid contents in the early stages of growth (low cell densities and lipid content). But smaller cultivation scale might be compatible with Nile red monitoring at later growth stages. For *Y. lipolytica* and *R. toruloides*, downscaling the culture by a factor 10 decreased the fluorescence intensity by half (**Figure 5-6**). Such a drop in fluorescence intensities, resulted in lowered sensitivity since the method is not discriminating for very low lipid contents (**Figure 5-7c**). While growth in 50-mL tubes seems sufficient to discriminate between low and high lipid accumulators further downscaling of the cultivation scale might result in an even bigger loss of sensitivity. For optimal results, we recommend analyzing samples of OD_{600} between 5 and 9 when growing samples in 50-mL tubes.

5.3.2 - Addressing remaining bottlenecks

Cell permeability to Nile Red has been the main bottleneck in the quantitative use of lipophilic dye for lipid content estimations in yeasts and microalgae. Methods have been proposed to improve cell permeability such as: the use of permeabilizers or dye carrier (i.e. DMSO, glycerol, ethanol, acetone, isopropanol, glutaraldehyde, ethylenediaminetetraacetic acid and ethylene glycol), sample pre-heating, combinations or grinding in liquid nitrogen (Patel et al., 2019; Rumin et al., 2015). Some of these methods exhibited good results but the effect is strain-specific or was only evaluated on a limited number of species. When considering the permeability limitation, a kinetic reading (repeated measurements over time) remains the safest option but is incompatible with the use of spectral mode. Indeed, repeated spectrum acquisition would lead to premature photobleaching explaining why kinetic readings are mainly executed with

endpoint readings (at fixed emission wavelength). But fixing the emission wavelength prevents the comparison of stains based on their respective emission maxima. A single-time reading would be preferable to enable the use of emission spectrum and avoid the accelerated fluorescence quenching. Single-time reading could be considered if a better permeabilization method was found. So acid hydrolysis was investigated as a permeabilization method on 20 yeast strains. The importance of wavelength selection for inter-species comparison was then investigated.

5.3.2.a - Permeabilization

Among the methods available for microalgae, one albeit promising was (to the best of my knowledge) never extended to yeasts: the staining of lysed cells proposed by Montalbo-Lomboy *et al.* (Gerde *et al.*, 2012; Montalbo-Lomboy *et al.*, 2014). The objective was to investigate this approach in yeasts to bypass the permeability issue. The method was optimized, tested on 20 strains and its main limitations were highlighted.

Montalbo-Lomboy *et al.* proposed a lysis based on the use of a surfactant and sonication. Here acid hydrolysis was used as the lysis method because it gave good results for the investigated strains prior to lipid extraction (see section 3.2.1) for TLC. Additionally, acid hydrolysis limits the number of chemicals added to the systems and potential unspecific interactions with Nile red. Sample hydrolysis was performed using HCl (1M) at 80°C for 30 mins. Lysates were then neutralized with NaOH (1M) with a vigorous mixing -that turned out to be crucial - before analysis using a final NR concentration of 50 µg/mL (working solution in DMSO). The first evaluation of the optimized method was done on samples grown in flasks to avoid any concentration-related issues. For most strains, a first emission peak around 650 nm was recorded during the first reading (t₀) (a). The emission maxima recorded for each strain at t₀ are presented in b.

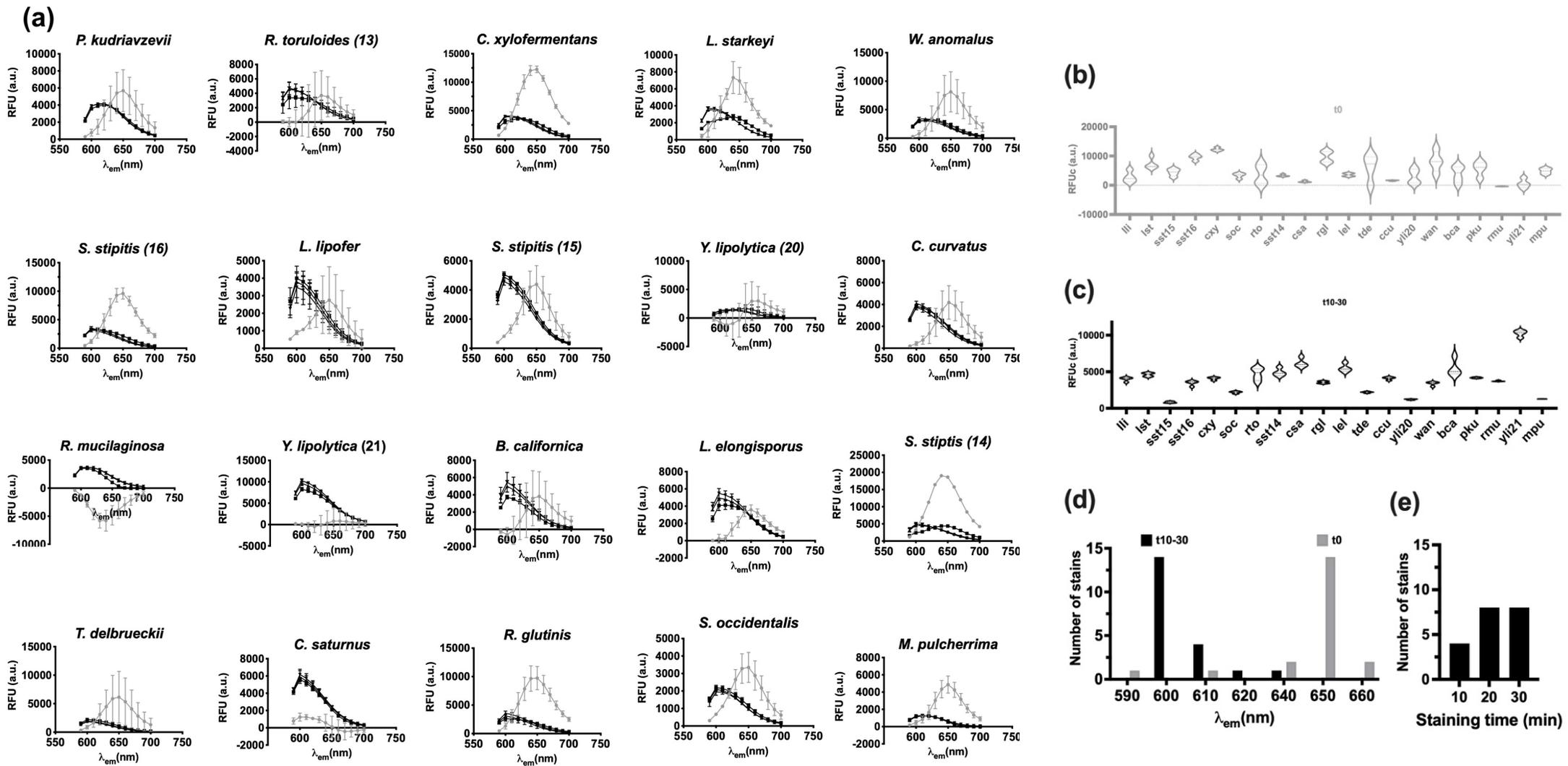


Figure 5-8 - Exploring the use of acid hydrolysis as a permeation method to improve the efficiency of NR staining

(a) Evolution over staining time ($t = 0, 10, 20$ and 30 min) of fluorescence (RFUc) spectra of 20 strains hydrolysed and NR-stained. (b) Violin boxplot (minimum, 25th percentile, median, 75th percentile, maximum values) representing the maximal RFUc reached for the 4 technical replicates of each strain during the initial reading. (c) Violin boxplot representing the maximal RFUc reached for the 4 technical replicates of each strain during the reading at $t=10, 20$ or 30 min (d) Summarises the distribution of emission wavelengths at which data presented in plots c were recorded (e) Summarises the staining time required to reach the RFUc presented in plots c.

The maximum emission wavelength of Nile Red interacting with polar lipids is higher than the maximum emission wavelength of Nile Red interacting with neutral lipids (Greenspan et al., 1985). So this first peak might correspond to Nile Red interaction with membranes and cells debris explaining the important variability recorded on this initial measurement (an average of 39.2% of relative variability on the emission maxima) **(c)**. But the hydrolysis totally changes Nile Red environment so peak identification solely based on previous emission wavelength reports is not possible and will require further investigation.

Then emission maxima shift towards 600 nm is observed for readings after 10, 20 and 30 min (t10, t20 and t30) **(b)**. This peak might correspond to Nile Red interaction with neutral lipid and is associated with a lower variability (an average of 7.6 % of relative variability on the emission maxima) **(d)**. But strains like *R. toruloides*, the three *S. stipitis* strains, *C. saturnus*, *L. elongisporus*, *B. californica*, *W. anomalus* still exhibit an over average variability (between 20.8 and 8.1% of relative variability) on the emission maxima recorded between 10 and 30 mins **(d)**. Out of these 8 strains exhibiting high variability between 10 and 30 min, 6 of them reach their emission maxima on the last reading (t30) so it is possible that the variability is the indication that the staining is not stable yet. Despite these few strains and unlike what is observed with cellular samples, the fluorescence remained constant for most strains from t10 to t30 (no significant differences between the emission maxima) **(b)**. Hydrolysis was used here to homogenize staining time and it turns out strains still require different times to reach the emission maxima: 4 strains require 10 min, 8 strains need 20 min and the same number at least 30 min **(f)**. The results are not ideal, but it is encouraging that the signal remains quite stable over t10 to t30 for most strains. The hydrolysis also influenced the emission wavelengths as 14 strains share the same emission maxima **(e)**. Then we checked whether the method would be able to account for subtle variation in

lipid content at lower concentrations. To do so, the *W. anomalus* and *S. stipitis* (15) were grown in 50-mL tubes, hydrolyzed, diluted and NR-stained.



Figure 5-9 - Linearity of fluorescence signal (RFU_c) against sample concentrations of strains with presumably low permeability to Nile Red (a) *W. anomalus* (b) *S. stipitis*.

Even though the fluorescence intensities were higher than the one observed when intact cells were stained (**Figure 5-7**), no linear relationship between fluorescence intensity and lysate concentration was found (**Figure 5-9**) when a difference was observed with intact cells (**Figure 5-7**). So, despite a few encouraging findings, the Nile red staining of lysate was abandoned because of its lack of sensitivity.

5.3.2.b - Emission wavelength selection

As identified in **Figure 5-6**, different species admit different emission maxima due to differences in lipid contents. Kimura *et al.* (2004) took that aspect into account when proposing a method based on the acquisition of the full emission spectrum. Then Sitepu *et al.* (2012) and later Rostron *et al.* (2017) recommended the use of kinetic readings (repeated measurement over time) that is hardly compatible with the acquisition of the full emission spectrum. Indeed, during spectrum acquisition, the sample is repetitively excited and such repeated excitation if extended over a long period of time might lead to premature fluorescence extinction/photobleaching. The use of acid hydrolysis as a permeabilization method would have enabled single time point reading and the use of spectral recording instead of endpoint readings. Since this approach failed, a careful

selection of the emission wavelength must be achieved. To illustrate the problem, the emission spectra of 5 different strains dyed with Nile Red were recorded between 520 nm and 760 nm (**Figure 5-10**).

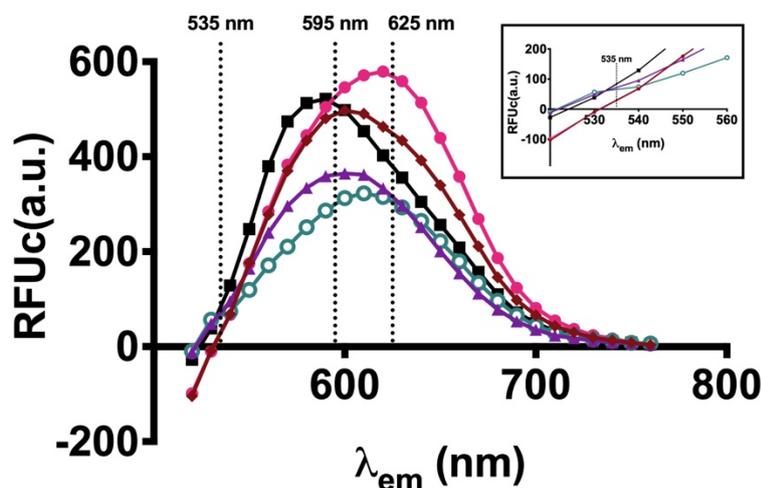


Figure 5-10 - Emission spectra of 5 NR-strained strains.

The dashed lines indicate candidate emission wavelengths for a transition from spectra towards endpoint readings. Emission maxima were recorded at 590 nm for *Y. lipolytica* (20) [■], 600 nm for *C. curvatus* [◆] and *S. stipitis* (15) [▲], 610 nm for *S. stipitis* (16) [○], and 620 nm for *R. toruloides* (12) [●].

Choosing to compare strains at a given wavelength instead of their respective emission maxima will inevitably lead to underestimations. For instance, in **Figure 5-10** comparing strains at 595 nm does not enable *Y. lipolytica* (20), *C. curvatus* and *R. toruloides* (12) discrimination. Comparing them at 625 nm aligns the two *S. stipitis* strains while a comparison at a wavelength lower than 595 nm groups *C. curvatus* and *R. toruloides* (12). The impact of emission wavelength selection was thus investigated. For that purpose, 21 strains were grown in 50-mL tubes for 3 days, NR-stained and analyzed with an excitation wavelength of 488 nm and emission wavelengths of either 535 nm or 625 nm. This experiment confirmed that tube culture is not suitable for the analysis of strains in their early growth phases since 3 of the 21 strains did not yield enough biomass for the Nile Red assay after 3 days of cultivation (**Figure 5-11**).

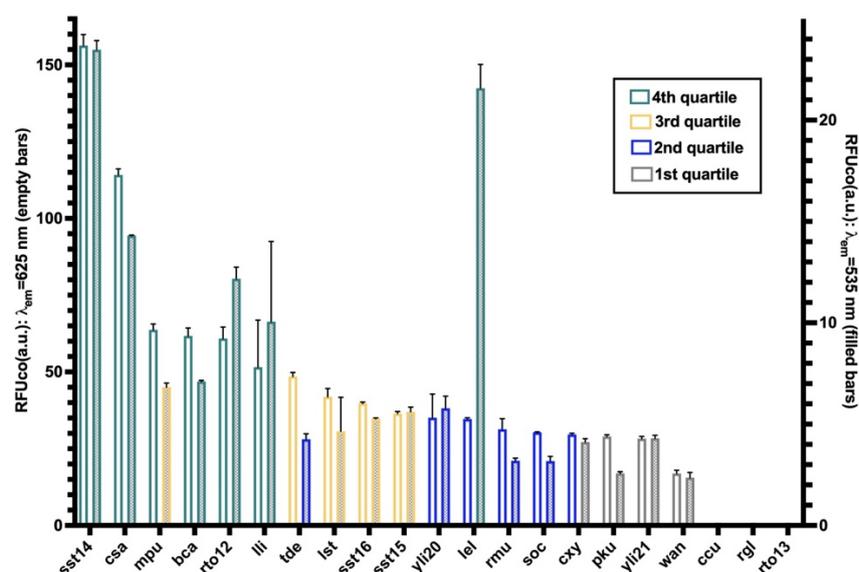


Figure 5-11 - Comparing the fluorescence of NR-stained strains at two emission wavelengths (625 nm and 535 nm).

Strain 3-letter codes are given in this figure (please refer to **Table 3-1** for complete names). Stained strains were sorted into quartiles according to their fluorescence. Bars colours indicate the quartile, and the emission wavelengths are specified by bar fullness (625 nm = empty bars; 535 nm = filled bars).

As for ranking, 80.95% of the strains are in the same quartile whether 625 nm or 535 nm is used as emission wavelength. And among the 4 (*M. pulcherrima*, *T. delbrueckii*, *L. elongisporus* and *P. kudriavzevii*) that are not in the same quartile, 3 are in the following quartile. Only *L. elongisporus* exhibit a huge difference in quartile (**Figure 5-11**). Different emission wavelengths represent different type of lipids. In an attempt to target a specific type of lipids, Chen et al. (2011) dyed separately monoacylglycerols, diacylglycerols, triacylglycerol, sterol esters and free fatty acids with Nile Red. This experiment demonstrated that even though each type of lipids admitted emission maxima at different wavelengths, all the peaks were spanning from 540 nm to 700 nm. So, the signal perceived when staining a cell is a combination of Nile Red interacting with all these types of lipids each with a different contribution to given wavelengths. *L. elongisporus* was the only strain with a major decoupling of fluorescence intensity

measured at 625 nm and 535 nm so its lipid accumulation over time in a different media was monitored to see if the decoupling was a kinetic effect or strain-specific effect.

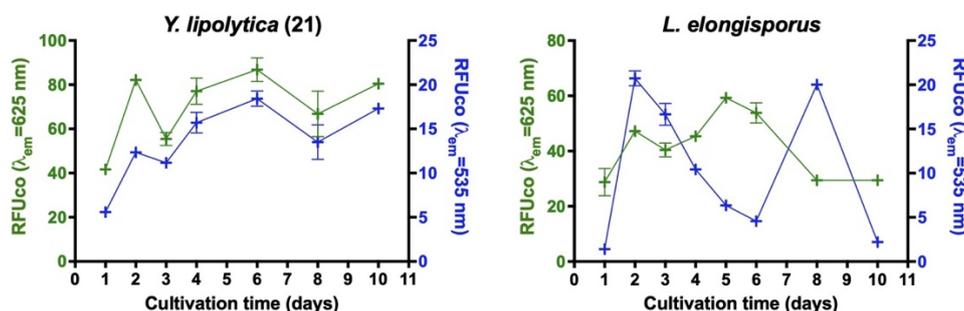


Figure 5-12 - Using Nile Red staining to monitor *Y. lipolytica* (21) and *L. elongisporus* lipid accumulation over time at two emission wavelengths (625 nm and 535 nm)

For this monitoring of lipid accumulation, *Y. lipolytica* (21) which exhibited consistent measurements between 625 nm and 535 nm (**Figure 5-11**) was used as a control (**Figure 5-12**). Consistency between 625 nm and 525 nm was confirmed at all points of the monitoring period for *Y. lipolytica* (21) when the decoupling in *L. elongisporus* appeared to be inherent to the strain. Since measuring the fluorescence of NR-stained cells at 2 wavelengths can draw attention to unusual lipid compositions, kinetic reading at 625 nm and 535 nm was adopted. Comparing strains based on fluorescence measurement at sub-optimal wavelengths give coherent sorting into quartiles for most strains indicating that the emission wavelength only becomes a determinant criterion when a more accurate ranking is required. This explains the wide variety of emission wavelengths proposed in the literature.

5.3.3 - Method summary

5.3.3.a - General recommendations and data normalization

Good practices

The influence of several parameters on the quality of yeast quantitative Nile Red staining was investigated in this study but parameters pertaining to Nile Red handling and storage were not detailed despite their importance. It is common practice while using

Nile red for microscopy to perform staining in the dark (Rostron & Lawrence, 2017) but despite a few papers recommending Nile Red storage in dark vials (Greenspan et al., 1985; Rumin et al., 2015), this aspect seems to have been left out of Nile Red staining for spectrofluorimetric measurements. We would like to emphasize the importance of this practice. Nile red should be prepared and preserved in amber vials. Fresh working solutions should be prepared for each experiment, but a stock solution can be preserved at -20 °C. If possible, Nile Red should be introduced in the assay right before fluorescence reading in a closed plate reader that shall remain closed for the length of the kinetic reading. If for some reason, Nile Red is introduced to the plate and then transported to the plate reader, the plate should be covered with foil or transported in an opaque box to isolate it from ambient light.

Data processing

Fluorescence data are typically normalized by cell density or by volume. Nile Red stained lipid fluorescence normalized by cell density reflects lipid contents while fluorescence data reflect titers when they are normalized by sample volume. For simplicity, OD is often used as an estimate of cell density especially when numerous samples are screened since the alternative would be cell enumeration which can be cumbersome and time-consuming in the absence of an automated platform. Yet, OD determination is affected by cell sizes and shapes and might not reflect the same number of cells from one species to another so OD normalization should only be used when comparing cells of similar morphology (Hicks, 2019). Here, we resorted to volume-based normalization to compare the panel of 21 yeasts and OD normalization was only used to monitor lipid accumulation in given strains. It should be noted that even for a given strain morphological changes during growth could affect the accuracy of OD-based normalization and thus, when possible, cell enumeration should be preferred although OD-based normalization remains a widespread practice (Morschett et al., 2016).

5.3.3.b - Rapid screening of a panel of unconventional yeasts

The optimized yeast screening protocol is summarized as follows:

- 1) Yeasts can be cultivated in volumes ranging from 5 mL to 100 mL for at least 24 hours (25 °C; 210 rpm).
- 2) Typically, 1 mL of culture is harvested and washed twice with PBS before storage at -80°C. The earliest the sample is harvested; the highest cell density will be required to make up for the low lipid contents. So, make sure to harvest enough cells to be able to prepare 700 µL of samples at $OD_{600} = 1$ (sample grown in flask) or 5 (sample grown in tube).
- 3) Samples were thawed and resuspended in 700 µL PBS to the desired OD_{600} . The OD_{600} should be recorded at this point for further normalization.
- 4) For each sample, 4 blanks and 4 replicates were loaded in the same column of a black 96-well with clear bottom. Each assay was made of 166 µL of cell suspension or PBS (blank), 17 µL of freshly prepared DMSO:PBS (1:1 v/v) and 17 µL of Nile Red working solution (60 µg/mL). Leave one column out of two empty and only add Nile Red right before reading.
- 5) Perform a 30 min kinetic reading (5 min increment) on non-empty wells with an excitation wavelength of 488 nm and at emission wavelength 625 nm (cutoff 610) and 535 nm (cutoff 530). With a 10 sec shaking before each reading. The fluorescence intensities should decrease within 30 min if not, extend the reading period.
- 6) The mean of all blanks signal on the plate was withdrawn of each sample fluorescence signal for background correction (RFUc). For each sample, the mean of the 4 replicates was taken and the value was multiplied by the appropriate dilution factor (RFUcd).

5.3.3.c - Application of the method to the monitoring of lipid accumulation

The optimized yeast lipid accumulation protocol is summarized as follows:

- 1) For the monitoring of lipid accumulation, yeasts should be cultivated in at least 50 mL volumes. Since the method involves daily sampling, the largest the culture volume is the lowest the impact of sampling will be. We suggest adjusting sample volume to culture OD₆₀₀ by starting with samples of 750 µL and decreasing as the culture OD₆₀₀ increases. For this method, not only sample OD₆₀₀ but also culture OD₆₀₀ should be recorded to allow the overlay of lipid accumulation profile to the strain growth curve.
- 2) For subsequent steps (2-5) see 5.3.3.b
- 3) The mean of all blanks signal on the plate was withdrawn of each sample fluorescence signal for background correction (RFUc). For each sample, the mean of the 4 replicates was taken and the value was normalized by dividing it by the OD₆₀₀ of the sample to account for small variation in OD₆₀₀ (RFUco).

5.4 - Discussion and concluding remarks

As exposed in this chapter variability in NR staining can originate from unexpected sources. In 2017, Rostron *et al.* was already highlighting the importance of good pipetting practices in reducing variability (standard deviation) in fluorescence assay. This finding was confirmed by Morschett *et al.* (2016) that improved the precision of their assay by 8 to 2 % by automating their method. If no sophisticated robots are available, the use of multichannel pipettes can be adopted mostly for kinetic motives because no significant differences were found in terms of variability between multichannel and monochannel pipettes. By transitioning to multi-channel pipettes an acetone evaporation issue arose because the Nile red working solution was kept in loosely closed reservoirs for the length of experiments.

Since the influence of acetone evaporation has been suspected before (Rostron & Lawrence, 2017), acetone was replaced by DMSO increasing thus the total amount of carrier in the assay. An increase in DMSO final concentration was also proposed by Hicks *et al.* (2019) (**Table 5-1**) in a modified version of Sitepu *et al.* protocol (2012). Hicks *et al.* (2019) demonstrated that samples achieved maximum fluorescence at 40 μL DMSO instead of 12.5 μL suggested in the original method. Another trend found in recent method optimization releases, is sample staining in microcentrifuge tubes before distribution into microplates for fluorescence reading (Miranda *et al.*, 2020; Zhao *et al.*, 2019) (**Table 5-1**). This simple step might have a tremendous impact on staining because both methods achieved coefficient of determinations (R^2), for the linear correlation between gravimetric determination and fluorescence reading, above 0.9 when the second reference protocol Sitepu *et al.* (2012) only achieved a R^2 of 0.7. This suggests that the mixing step might be a limiting factor when staining the cells directly in 96-wells plate. Based on De la Hoz Siegler *et al.* work (**2012**), a vigorous plate shaking combined with a reduction of assay volume was implemented here, but I can only encourage one to use in-tube mixing if the number of samples allows it. Additionally, Zhao *et al.* also emphasized the importance of maintaining strict dark conditions. Finally, correcting the fluorescence using cells count rather than OD_{600} was found to be more accurate than OD_{600} (Hicks *et al.*, 2019) but still suffers from the same shortcoming as both are dependent on cells physiological properties (Morschett *et al.*, 2016).

This method optimization was approached with the idea of further downscaling the method and find a way to overcome the difference of permeability to Nile Red to facilitate strain comparison. Since Nile Red staining of cell lysates did not give the expected results, the use of a dye carrier (i.e DMSO) combined with kinetics readings remains the most acceptable way to manage yeast difference of permeability to Nile Red. This work demonstrated that the choice of an emission wavelength was not a limiting

factor in inter-species comparison. Indeed, even though different species admit different emission maxima, comparing species based on a wavelength different from their emission maxima is enough to distinguish low lipid accumulators from high accumulators.

Furthermore, measuring the fluorescence at two distinct wavelengths can help identify singularities in lipid composition. Finally, regarding the scale of the assay, this work confirms that a reduction of cultivation volume from 20 to 100 mL to 5 mL is possible if accompanied by an increase in the minimal cell density requirements for the assay. Even though a reduction of the cultivation scale is possible, it will not lead to a throughput increase because it was found that reducing the number of samples on plates was preferable to limit inter-well crosstalks during fluorescence reading. Among recent development in yeast Nile Red staining, only Miranda *et al.* (2020) (**Table 5-1**) proposed changes to the cultivation scale by limiting the liquid culture to the preparation of inoculum and designing a screening based on solid media culture. This protocol (Miranda *et al.*, 2020) enables the simultaneous growth of more strains than the one using liquid cultures, but the applications are limited. When Nile red assays based on liquid cultures proved to be applicable not only to inter-species comparison but also to the monitoring of lipid accumulation and have the potential to be integrated into high-throughput processes and medium optimization pipelines (Jordan & Stettler, 2014).

Part III – New insights into oleaginous yeasts metabolism

Chapter 6 - Combined genomic and phenotypic approach to explore the diversity of oleaginous yeast with high potential for lipid production

6.1 - Introduction

Despite demonstrated high-performances, most oleaginous yeasts remain less documented and amenable than traditional industrial species which hinders their enhancement and industrial utilization. Widening our knowledge and command of unconventional oleaginous yeasts has become an important stake for the establishment of microbial oils. A better understanding of oleagenicity and other industrially relevant traits could facilitate the development of processes based on unconventional strains or inspire new engineering strategies for traditional industrial species.

With the increasing availability of microbial genomes, pan-genomics stood out as a powerful discipline able to support functional genomic, population genomic, evolutionary genomic, phylogenomic and even metagenomic analysis (Guimarães et al., 2015; Marschall et al., 2018). Yet so far, microbial pan-genomics studies have been driven by evolutionary or medical interest (Bosi et al., 2016; McCarthy & Fitzpatrick, 2019a; Seif et al., 2018; Song et al., 2015; Sun et al., 2015). To the best of my knowledge, only a few studies used pan-genomic to investigate the biotechnological potential of bacterial species and it was for food applications (Chun et al., 2017; Hao et al., 2011). Aside from *S. cerevisiae* genomes, yeast genome sequencing has historically been dominated by clinical interest (Dujon, 2015; Piškur & Langkjær, 2004) so the lack of genomes might have prevented the implementation of pan-genomic analysis for unconventional yeasts. But a recent interest shift towards the sequencing of strains with great biotechnological potential is now allowing such approach. Hence, in 2016 a study comparing 29 genomes from biotechnologically relevant *Ascomycetes* yeasts was published and gave valuable

insight into the genetics of *Ascomycetes* carbon utilization, mating, phylogeny and synteny (Riley et al., 2016). The same year, the first comparative genomic analysis of oleaginous yeasts was released but the study only included 9 species (Shen et al., 2016). The studies that followed were more focused on a given clade or species (Fakankun et al., 2021; Li et al., 2020; McCarthy & Fitzpatrick, 2019b) when phylogenetic data suggest that oleaginicinity might have emerged multiple times during evolution (Sitepu et al., 2014a). Since it is very likely that mechanisms underlying oleaginous properties are not shared across all oleaginous species, an inter-clade comparison could help highlight interesting singularities.

While the potential of pan-genomics contribution to the genomes functional annotation process was demonstrated in **Chapter 4**, its discriminating power will be exploited in this chapter. A comparative analysis of 16 genomes from both *Ascomycetes* and *Basidiomycetes* oleaginous yeasts was performed. The conservation of industrially relevant traits, such as carbon utilization and strain stress management abilities, was first investigated using pan-genomics. The pan-genomics results were then confronted with phenotypic data. And finally, a comparison of genes involved in lipid metabolism was achieved to identify singular genetic features that might contribute to oleaginicinity.

6.2 - Materials and methods

6.2.1 - Cell culture

Media preparation, cell maintenance and culture were performed as described in **Chapter 3 (section 3.2)**. YPD was also used for the thermotolerance evaluation (incubation at 25°C and 37°C). YPD supplemented with 1M or 2.3 M NaCl was used for the osmotolerance assessment. For the carbon utilization tests, pH-adjusted nutrient-sufficient synthetic complete media containing either glucose, xylose, lactate, acetate or glycerol, as carbon sources (see **section 3.2.1** for compositions), were used. Cells were

grown in microtiter plates for carbon utilization, osmotolerance and thermotolerance evaluation, otherwise, centrifuge tubes or flasks were used.

6.2.2 - Lipid extraction and thin layer chromatography

For the lipid extraction, cells were grown into glucose nitrogen-limiting media (GlySC₆₀) for 4 days in 250-mL flasks. Cells from the whole working volume were harvested. The extraction and TLC were performed as described in **Chapter 3 (section 3.2)**.

6.2.3 - Principal component analysis

Clusters and unique genes associated with the functional category “Lipid transport and metabolism” were listed. For each species, the number of genes in the clusters of interest was computed and formatted as a matrix that was analyzed by principal component analysis using MATLAB.

6.3 - Results and discussion

6.3.1 - The pan-genomic analysis reveals conserved functional categories

Functional categories with high representations in unique genes might have been exposed to higher selective pressure and possibly play a role in individual strain environmental preferences and adaptation. If high proportions of unique genes are associated with low proportions of core genes, the categories are poorly conserved across species. These non-conserved categories are “Metabolism of cofactors and vitamins” (Cat. 7), “Xenobiotic biodegradation and metabolism” (Cat. 8), “Signal transduction” (Cat. 12) and to a lesser extent “Mitochondrial biogenesis” (Cat. 14) which supports of the aforementioned environmental adaptation hypothesis (**Figure 6-1**). The implication of “Mitochondrial biogenesis” (Cat. 14) is less obvious than the contribution of the other two categories but “Mitochondrial biogenesis” (Cat. 14) has been described to be induced by stress or in environmental signals (Bouchez & Devin, 2019). When a significant proportion of

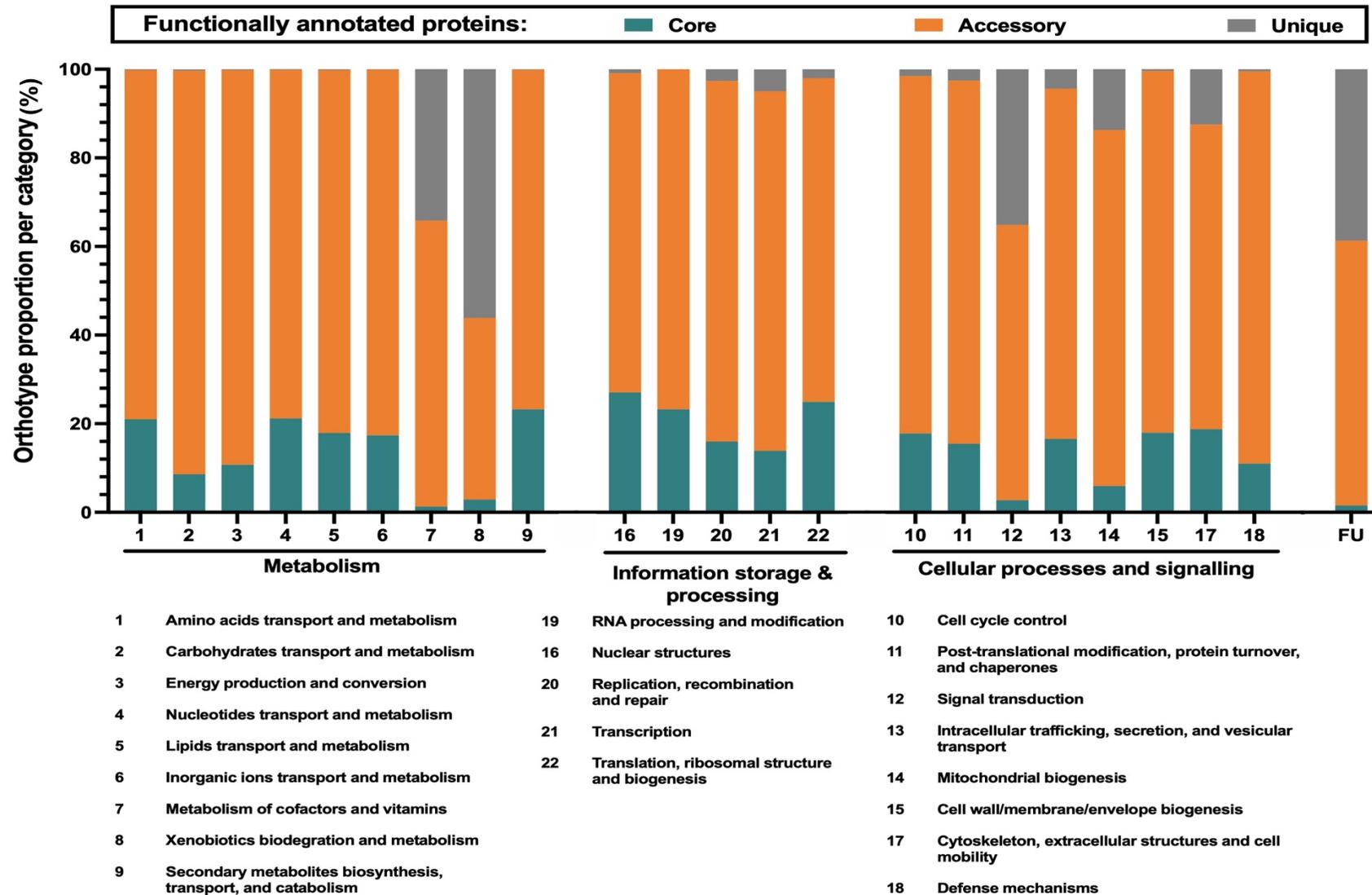


Figure 6-1 Functional distribution of pan-genomic genes.

Each bar represents a functional category except the FU bar for genes of unknown function. All the functional categories are listed under the plot. For each functional category is given the percentage of core genes (shared across all the studied species) in teal, accessory genes (shared across a few species) in orange and unique genes (specific to a species) in grey.

genes is combined with a higher proportion of core genes, the function can be assumed to be more widely conserved. Like it is the case for “Cytoskeleton, extracellular structures and cell mobility” (Cat. 17) (**Figure 6-1**).

Finally, categories mainly enriched in accessory genes (low proportions of unique and core genes), like “Carbohydrate transport and metabolism” (Cat. 2), “Energy production and conservation” (Cat. 3) and “Defense mechanisms” (Cat.18), are shared mainly across a few strains perhaps across closely related species (**Figure 6-1**).

All the other categories exhibit similar profiles even if categories like “Replication, recombination, repair” (Cat. 20) and “Transcription” (Cat. 21) seem slightly enriched in unique genes and depleted in core genes compared to the other information storage and processing (ISP) categories which might suggest an increased contribution to adaptation phenomena compare to translation, maintenance of nuclear structure or RNA processing and modifications. The decoupling between “Transcription” (Cat. 21) and “RNA processing and modification” (Cat. 19) (**Figure 6-1**) seems to indicate that proteins involved in the transcription process admit more variability than proteins responsible for RNA processing which would be consistent with the described important divergence observed between transcription factors of related yeasts (Borneman et al., 2007; Zheng et al., 2010).

6.3.2 - The phenotypic characterization nuances pan-genomic predictions

6.3.2.a - Carbon utilization

The 3 categories (Metabolism, ISP and CPS)⁶ represent each about a third of all functionally annotated genes, but they might weigh differently on phenotypes. Confronting pan-genomic predictions with phenotypic data could help us understand

⁶ ISP and CPS respectively stand for « Information Storage and Processing » and « Cellular Processes and Signaling »

the contribution of each category. Carbon utilization is a particularly important criterion for microbial oil production. It was evaluated as the major bottleneck in microbial lipid production since feedstock was described to account for up to 70% of the total production cost (Muniraj et al., 2015). That is understandable knowing that it takes 5 tons of sugars to produce 1 ton of oil (Wynn & Ratledge, 2005). The latter observation has prompted the exploration of “zero-cost” carbon sources although tradeoffs between product price and raw material cost need to be found to generate a margin sufficient to ensure process viability (Ratledge, 2002).

Inter-strain variability is often reported when it comes to carbon utilization abilities so phenotypic screenings seem to indicate that carbon utilization is highly strain-specific when pan-genomic data suggest that carbon utilization abilities might be quite conserved across limited groups of species given the accessory gene representation in “Carbohydrate transport and metabolism” (Cat. 2), “Energy production and conservation” (Cat. 3) (**Figure 6-1**). Inter-strain variability was sometimes imputed to the fact that strains of the same species evolved in different environments (Opulente et al., 2013, 2018), to simple genetic drift (Warringer et al., 2011) or to experimental discrepancies resulting from differences in parameters such as media composition in cofactor and vitamins as suggested by Klein et al. (Klein et al., 2017) when discussing the ability of *S. cerevisiae* to grow on glycerol. Inter-strain variability being the result of variations in media composition would be consistent with the low conservation of “Metabolism of cofactors and vitamins” (Cat. 7) genes (**Figure 6-1**).

Here the growth of 21 strains (including several variants for some species) was experimentally compared in 5 industrially relevant⁷ carbon sources with the same

⁷Wastes or by-products are industrially relevant carbon sources. Lignocellulose is mainly made of xylose and glucose. Lactic acid and acetic acid are often found in wastewaters of food industry (Casey et al., 2010). And the relevance of glycerol was discussed in Chapter 1.

nutrients to verify whether or not inter-strain and inter-clade variability were still observed.

Table 6-1 Summary of carbon utilization abilities

	Glucose SC	Xylose SC	Lactic SC	Acetic SC	Glycerol SC
<i>M. pulcherrima</i> (8)	+				+
<i>Y. lipolytica</i> (21)	+		+	+	+
<i>Y. lipolytica</i> (20) *	+	+	+	+	+
<i>P. kudriavzevii</i> (9)	+	+		+	+
<i>C. saturnus</i> (4)	+	+	+		+
<i>L. elongisporus</i> (7)	+				+
<i>S. stipitis</i> (16)	+	+			+
<i>C. xylofermentans</i> (2)	+	+			+
<i>S. occidentalis</i> (17)*	+	+	+	+	+
<i>R. mucilaginosa</i> (11)*	+	+	+	+	+
<i>S. stipitis</i> (15)	+	+			+
<i>W. anomalus</i> (19)	+		+	+	+
<i>T. delbrueckii</i> (18)	+				+
<i>B. californica</i> (1)*	+	+	+	+	+
<i>C. curvatus</i> (3)*	+	+	+	+	+
<i>R. toruloides</i> (13)	+	+			+
<i>R. glutinis</i> (10)*	+	+	+	+	+
<i>R. toruloides</i> (12)*	+	+	+	+	+
<i>L. lipofer</i> (5)	+	+			±
<i>L. starkeyi</i> (6)*	+	+	+	+	+
<i>S. stipitis</i> (14)	+	+			+

Nutrient sufficient conditions and initial pH adjusted to pH = 6 ± 0.1. Growth curves recorded for each carbon source were compared to the growth curves obtained in synthetic complete (SC) without carbon source. A condition resulting in a significant growth was indicated by a “+”. Black boxes represent the absence of growth and the symbol “±” means that growth was very slow. Names were assigned an asterisk if the strain can grow in all the tested media.

Eight strains were able to grow on the 5 tested carbon sources: *R. toruloides* n°12, *C. curvatus*, *R. mucilaginosa*, *R. glutinis*, *Y. lipolytica* n°20, *S. occidentalis*, *B. californica* and *L. starkeyi* (Table 6-1). Those assimilation assays also revealed that all strains are at least able to grow on 2 of the tested carbon sources. The most common being glucose

and glycerol. This experiment was discriminating for the two *R. toruloides* and *Y. lipolytica* strains and for closely related species like *L. lipofer* and *L. starkeyi* confirming that carbon source preferences are not even conserved at species or clade levels despite controlled experimental parameters. This dissociation between genotypic and phenotypic findings might be explained either by different regulatory circuits or by high evolutionary divergence. In both cases, close relatives would share complete utilization pathways at the genetic level but the pathways would have either different levels of functionality across strains or be differentially regulated which would be consistent with the great diversity observed in “Signal transduction” (Cat. 12) genes (**Figure 6-1**).

6.3.2.b - Thermotolerance and halotolerance

Given the important variability observed for the category “Signal transduction” (Cat. 12) (**Figure 6-1**), it is very likely that traits relying directly on genes of these categories would present high inter-clade variability. Signal transduction is involved in the cellular response to environmental stimuli and stresses so this functional category can be involved a multitude of cellular functions. Among those traits, two highly industrially sought-after abilities are the halotolerance and thermotolerance. Halo tolerance is defined as the ability of an organism to withstand and thrive in ionic concentrations higher than the one required for its growth (Antón, 2011). Thermotolerant organisms can grow at a wide range of temperatures spanning from 8°C to 42°C when thermophilic strains can only grow at high temperatures (starting at 25°C) (Arthur and Watson, 1976). Given the interval of temperature considered here (25°C to 37°C), strain thermotolerance will be discussed rather than strain thermophilia.

Osmotic and thermal shock have been described as modulators of lipid production with repercussions on lipid nature and/or concentrations (Sitepu et al., 2014b). Furthermore, the osmotic shock has also been described to stimulate the production of small polyol osmoprotectants (Křęgiel et al., 2017) that could be valorized as co-products.

Investigating strains' tolerance to salt or heat is thus relevant since these stresses can influence lipid production. Additionally, both thermotolerance and halotolerance are industrially interesting properties because they offer some attractive perspectives regarding process design like: the possibility of growing halotolerant strain in seawater rather than freshwater (Zaky et al., 2018), the limitation cooling of lignocellulosic feedstocks enzymatic saccharification a process that usually takes place at 50°C (Radecka et al., 2015) or opportunities of contamination risks reduction (Lamers et al., 2016; Mukherjee et al., 2017).

Thermotolerance was assessed by comparison between the growth in rich medium at 25°C and 37°C. Out of the 21 strains evaluated, only 10 were able to maintain some growth at 37°C (Table 6-2). And from those ten, only 3 did not seem affected by the temperature change: *P. kudriavzevii*, *S. stipitis* (15) and *R. toruloides* (13). The results of the two *Y. lipolytica*, the two *R. toruloides*, and the three *S. stipitis* also suggest that temperature tolerance is a trait quite conserved among strains of the same species even though the extent of the tolerance admits small inter-strain variations. This finding should enable the transposition of tolerated cultivation temperatures from one strain to another for a given species but extrapolating this information to another species of the same genus might not be recommended. Both *R. toruloides* and *R. mucilaginosa* can maintain some growth at 37°C (Table 6-2) when more phylogenetically distant *Basidiomycetes* *R. glutinis* and *C. curvatus* (Appendix 1) cannot (Table 6-2).

Looking at the CTG clade, *S. occidentalis*, *S. stipitis*, *C. xylofermentans* outperformed (Table 6-2) extended relatives like *L. elongisporus* and *M. pulcherrima*. The two *Lipomyces* sp. performed similarly. Yet, *W. anomalus* and *C. saturnus* behaved differently despite their phylogenetic proximity. Indeed, *C. saturnus* behaved more like *B. californica* (Table 6-2), the other member of the *Phaffomycetaceae* family.

Table 6-2 Overview of halotolerance and thermotolerance

Strains	Halotolerance						Thermotolerance	
	YPD		YPD +1 M NaCl		YPD +2.3 M NaCl		YPD 25°C	YPD 37°C
	Time	OD ₆₀₀	Time	OD ₆₀₀	Time	OD ₆₀₀		
<i>M. pulcherrima</i> (8)*	1 day	1.94	1 day	1.84	4 days	1.83	+	-
<i>Y. lipolytica</i> (21)	1 day	2.16	2 days	2.08	-	0.24	+	-
<i>Y. lipolytica</i> (20)*	1 day	2.16	2 days	2.09	> 7 days	1.52	+	-
<i>P. kudriavzevii</i> (9)☐	1 day	1.99	2 days	1.74	-	0.04	+	+
<i>C. saturnus</i> (4)	1 day	2.00	2 days	0.66	-	0.10	+	-
<i>L. elongisporus</i> (7)*	1 day	2.07	1 day	2.08	3 days	1.93	+	+
<i>S. stipitis</i> (16)	1 day	2.07	2 days	1.65	-	0.14	+	+
<i>C. xylofermentans</i> (2)	1 day	2.01	2 days	1.97	-	0.12	+	-
<i>S. occidentalis</i> (17)	1 day	1.99	2 days	1.89	-	0.11	+	+
<i>R. mucilaginosa</i> (11)*	1 day	1.95	2 days	1.86	6 days	1.93	+	+
<i>S. stipitis</i> (15) ☐	1 day	2.05	2 days	1.94	-	0.10	+	+
<i>W. anomalus</i> (19)*	1 day	2.06	2 days	2.00	> 5 days	1.89	+	+
<i>T. delbrueckii</i> (18)*	1 day	1.94	2 days	1.92	4 days	1.75	+	-
<i>B. californica</i> (1)	1 day	2.05	2 days	1.92	-	0.08	+	-
<i>C. curvatus</i> (3)	1 day	2.08	2 days	2.06	-	0.07	+	-
<i>R. toruloides</i> (13) ☐	1 day	1.88	3 days	1.91	-	0.07	+	+
<i>R. glutinis</i> (10)	2 days	1.88	3 days	1.92	-	0.08	+	-
<i>R. toruloides</i> (12)	2 days	1.84	3 days	1.90	-	0.09	+	+
<i>L. lipofer</i> (5)	3 days	1.67	-	0.13	-	0.08	+	-
<i>L. starkeyi</i> (6)	3 days	1.84	-	0.11	-	0.12	+	-
<i>S. stipitis</i> (14)	1 day	2.07	2 days	1.94	-	0.13	+	+

Halotolerance: Was given the time needed to reach an early saturation and the highest OD reached by the strain in the tested conditions. All the time and OD summarized in this table were rounded. Noteworthy strain names are followed by an asterisk. **Temperature effect:** noteworthy strains were associated with the following symbol ☐. A “-” was assigned when no growth was observed in the tested temperature. A “+” was assigned when growth was recorded, and the “+” was framed when the strain was able to maintain the same level of growth at 25°C and 37°C.

Halotolerance was evaluated using YPD supplemented with 1M and 2.3 M NaCl (**Table 6-2**). The 2.3 M NaCl concentration appeared to be the most selective as only 6 strains were able to maintain some growth at this salt concentration: *Y. lipolytica* (20), *M. pulcherrima*, *R. mucilaginosa*, *T. delbrueckii*, *L. elongisporus*, *W. anomalus*. At 1 M NaCl, despite an initially stunted growth, the relatives *R. torulooides* (12 and 13) and *R. glutinis*, seem to produce slightly more final biomass when grown in YPD +1 M NaCl compared to regular YPD. They share this characteristic with no other strains among the 21 evaluated strains not even *R. mucilaginosa*. On the other side, *M. pulcherrima* and *C. xylofermentans* maintained the same growth rate in YPD + 1M but reached hardly lower final optical density at 600 nm (OD₆₀₀) compared to their growth in YPD. The 10 remaining strains showed a slower initial growth rate at 1M NaCl, although *C. curvatus*, *T. delbrueckii* and *L. elongisporus* ultimately reached the same final OD₆₀₀ as the one reached in YPD. Finally, *Lipomyces* strains were not even able to grow in YPD + 1M NaCl. Looking at both halotolerance and thermotolerance, strains from very different clades seem to be able to tolerate the harshest conditions which would suggest that both traits are not clade-specific and be consistent with the pan-genomic prediction if it wasn't for the very clear familial trends observed when investigating the thermotolerance. Regarding halotolerance, there are fewer familial similarities, and they are subtler suggesting that response to temperature stimuli is probably more conserved than the response to high NaCl concentrations indicating that even within a given functional category, some paths might be more conserved than others.

6.3.2.c - Biomass production

Fast growth and high biomass yields can make up for average lipid contents, hence the ability to quickly generate high cell densities can be extremely advantageous. Yet the rational enhancement of this ability is more difficult than the improvement of product accumulation because growth depends on a lot of different factors. More than the other

three traits considered, biomass production might reflect the conservation of genes from metabolism, ISP and CPS processes. Yet the good conservation of the categories like “Cell cycle control” (Cat. 10) and “Cell Wall/ Membrane/ Envelope biogenesis” (Cat. 15) (**Figure 6-1**) might balance the poor conservation of other involved categories. The growth of the 21 strains was thus compared in rich glucose-containing media to assess the biomass production potential of each strain (**Figure 6-2**). As a lot of parameters can affect the biomass yield including medium composition, the type of cultivation vessels, the aeration, the temperature, the length of the experiment and biomass production capabilities were evaluated in 2 different set-ups. In the first condition, strains were grown in a rich glucose-containing media (YPD) in 50-mL centrifuge tubes until cells reached an early saturation phase. For the second condition, cells were cultivated for 4 days in flasks containing a glucose-based defined media with a C:N ratio of 60. The optical OD₆₀₀ of each culture was monitored over time and the highest OD reached were plotted (**Figure 6-2**).

In condition 1 (**Figure 6-2**), 3 strains reached the saturation phase after 24h out of which *C. xylofermentans* (2) was able to yield the highest cell density. A cell density still low compared to the one (48 OD units) achieved by *T. delbrueckii* (18) in the same conditions after 2 days of incubation. Growth in condition 1 was a barely discriminating parameter for *S. stipitis* strains. The three strains achieved quite high OD₆₀₀ (between 30-46 units) even though two of them n°14 and 15 reached the early saturation phase after 2 days while the other, *S. stipitis* (16), took 3 days to reach the same stage. And another member of the CTG clade, *L. elongisporus* (7), like *S. stipitis* strains, was even able to attain the same OD₆₀₀ as *S. stipitis* (16) with the speed of *S. stipitis* (14 and 15).

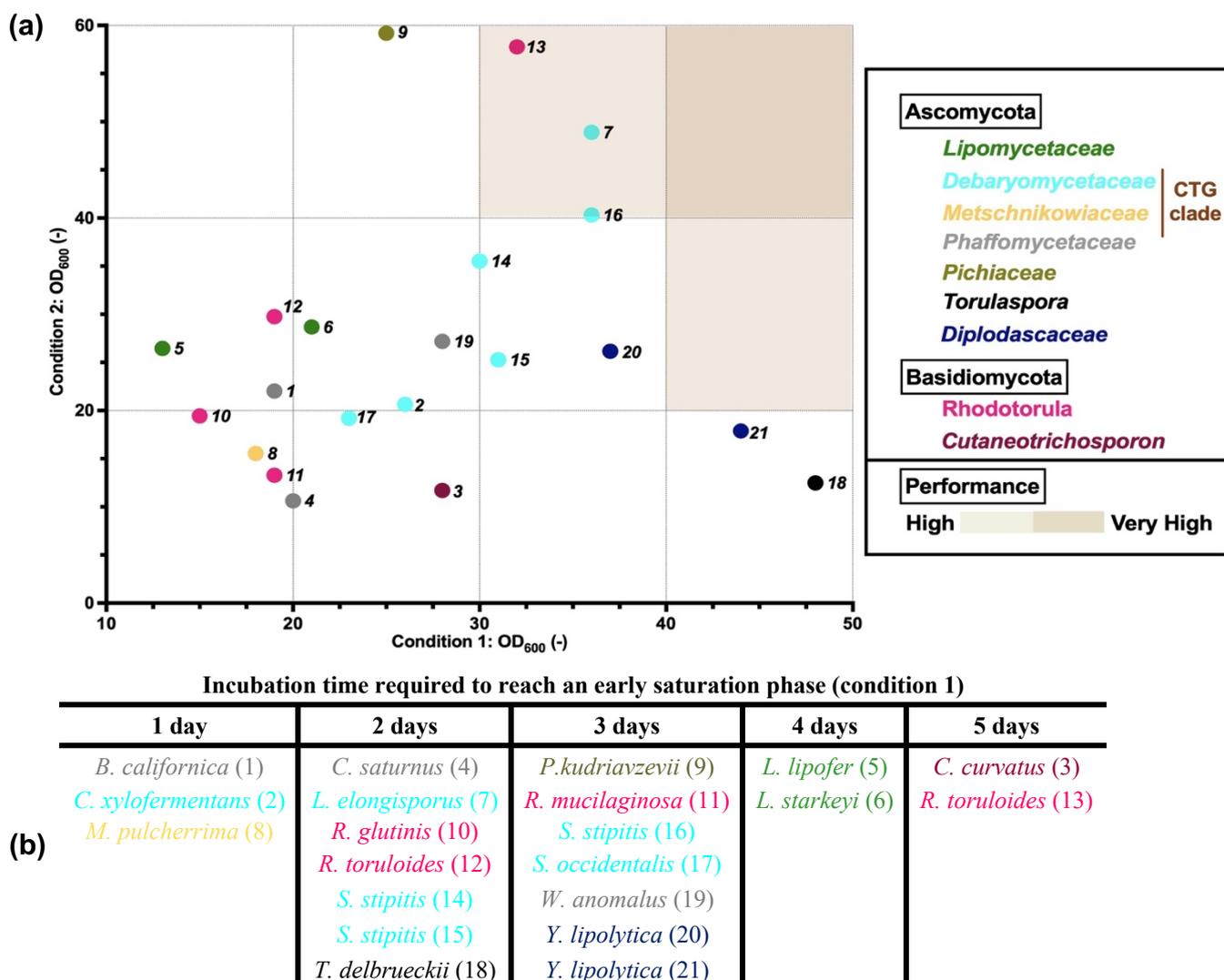


Figure 6-2 Biomass production in glucose-containing media.

Strain taxonomic distribution was color-coded. (a) Cell density: Condition 1 (x-axis) corresponds to growth in 5 mL of yeast peptone dextrose (YPD) using 50-mL centrifuge tubes as vessels. Condition 2 (y-axis) corresponds to 4 cultivation days in 250-mL flasks in 50 mL of high C:N molar ratio glucose synthetic complete medium (GlcSC₆₀). The results of both experiments were plotted to facilitate high-performance species identification. (b) The time needed for strains to reach an early saturation phase in condition 1.

The two *Y. lipolytica* strains reached simultaneously saturation phases characterized by very high ODs (37 and 44 units). But *R. toruloides* (13) growth in YPD, stood out from the growth of its relatives. *R. toruloides* (12), *R. mucilaginoso* (11) and *R. glutinis* (10) exhibited similar growth in condition 1 as they only need 2-3 days to reach an early saturation phase characterized by quite low OD₆₀₀ (between 15-21). When *R. toruloides*

(13) early saturation was characterized by a much higher OD₆₀₀ (33 units) but requires 5 days of incubation (**Figure 6-2**).

Interestingly, in condition 2, *R. toruloides* (13) achieved one of the highest cell densities. A performance matched by *S. stipitis* (16), *L. elongisporus* (7) and *P. kudriavzevii* (9) (**Figure 6-2**). Despite the OD₆₀₀ differences observed between *R. toruloides* (13) and *R. toruloides* (12) that reached respectively OD₆₀₀ of 58 and 30 units, they both yielded higher OD₆₀₀ in condition 2 compared to condition 1. On the contrary, the two *Y. lipolytica* strains achieved lower OD in condition 2 when the 3 *S. stipitis* strains were only mildly affected by the change of condition. When combining information from both experiments, *S. stipitis* (16) and *L. elongisporus* (7) are standing out. Given these 2 strains are part of the same clade and that some behavioral similarities were observed among close relatives, the phylogenetic conservation of this trait was investigated. Clustering attempts did not reveal any correlation between performances and taxonomy. The ability to quickly yield high cell densities does not appear to be a clade-specific trait. Furthermore, the experiment did confirm the significant influence experimental design has on the outcome of this type of test.

6.3.3 - Diving deeper into genes related to lipid metabolism

6.3.3.a - Overview of current knowledge

Visualizing the neutral lipid extracts of the investigated strains on a TLC (**Figure 6-3a**) and knowing that the functional category “Lipid metabolism and transport” (Cat. 5) is enriched in accessory genes (**Figure 6-1**), it becomes obvious that strains may admit variations in lipid metabolism and possibly clade-specific variations. A few mechanisms have already been described but do not apply to all the oleaginous yeasts so there is still a lot to be deciphered about oleaginicities in yeasts. As aforementioned, in certain strains,

limited access to nitrogen promotes lipid production which imposes a very specific timing on lipid accumulation that is often coupled to growth (Beopoulos et al., 2008). Nitrogen-induced lipid accumulation tends to coincide with the saturation phases and since it was here demonstrated that growth-related features are poorly conserved across strains, it would not be surprising that when comparing lipid accumulation in various yeasts, the first differences are kinetic. As nitrogen depletion is tightly related to cell duplication rate and strains exhibit different growth rates (**Figure 6-2**), these differences in growth rate will result in differences in nitrogen-induced lipid accumulation. And those differences will be even more important that different yeasts have been described to be sensitive to different levels of nitrogen limitation (often discussed in terms of C:N ratio).

An optimal initial C/N ratio of 72 was described for *L. starkeyi* NRRL Y-11557 (Riley et al., 2016) when *R. toruloides* is sensitive to C/N ratio from 30 to 120 (Nicaud et al., 2014) for instance. This suggests potential differences in nitrogen-sensing and response to starvation mechanisms. C/N ratio was chosen as an indicator because yeast lipid content is also highly dependent on carbon availability. In the events of carbon depletion, cells tend to degrade their storage compounds for survival. Differences in lipid degradation might also explain high lipid contents. *L. starkeyi* was described to display a very low degradation of endogenous lipids during carbon-starvation, a characteristic they don't share with *C. curvatus* (formerly *C. curvata*) and *R. toruloides* (Holdsworth et al., 1988). Nevertheless, other features were also identified in *R. toruloides*: the presence of 2 acyl carrier protein (ACP) whom tandem usage could greatly improve the stain fatty acid synthase (FAS)(Zhu et al., 2012) and the unique class of perilipin-like protein probably involved in lipid droplet formation. Actually, the perilipin-like proteins found in *R. toruloides* are different from the type (MPL1) identified in *Y. lipolytica* (Zhu et al., 2012).

In order to go beyond these well-known mechanisms, an identification of the major genetic differences between investigated species was undertaken. Only genes annotated as involved in lipid metabolism or homologous to genes with such annotation were considered. Individual gene representation in each species was derived from the pan-genomic orthology information and used to build a matrix that was then analyzed using a principal component analysis (PCA). Inter-species distances were thus determined only based on the number of absent/present homologous and unique genes presumably related to lipid metabolisms. Genetically speaking, *Y. lipolytica* and *C. curvatus* are the most distant species when it comes to lipid metabolism (**Figure 6-4b**).

6.3.3.a - Comparative analysis of genes related to lipid metabolism: principal component 1 (PC1)

The differences between *Y. lipolytica* and the rest of the cohort are particularly well represented by the first principal component (PC1) that accounts for 32.8% of the global genetic variations in the lipid metabolism of the 16 species. Genes involved in glycerolipids metabolism, lipid degradation and sphingolipids metabolism mainly contribute to PC1 (**Figure 6-3b**). Indeed *Y. lipolytica* genome seems enriched in genes involved in glycerolipids metabolism, especially 27 strain-specific genes organized in 2 clusters (cluster3615 and cluster3140) (**Figure 6-3b**) that are unfortunately poorly annotated. In addition, *Y. lipolytica* genome also contains 16 genes coding for lipases (Class 3) all clustered (cluster1048) with homologous proteins from all the studied species except from the 4 *Basidiomycetes sp.* and *T. delbrueckii* (**Figure 6-3b**). Outside this cluster, *Y. lipolytica* genome encodes 2 additional lipases of the same class for a total of 18 genes when *W. anomalus* genome only contains 7 of these genes and the other species

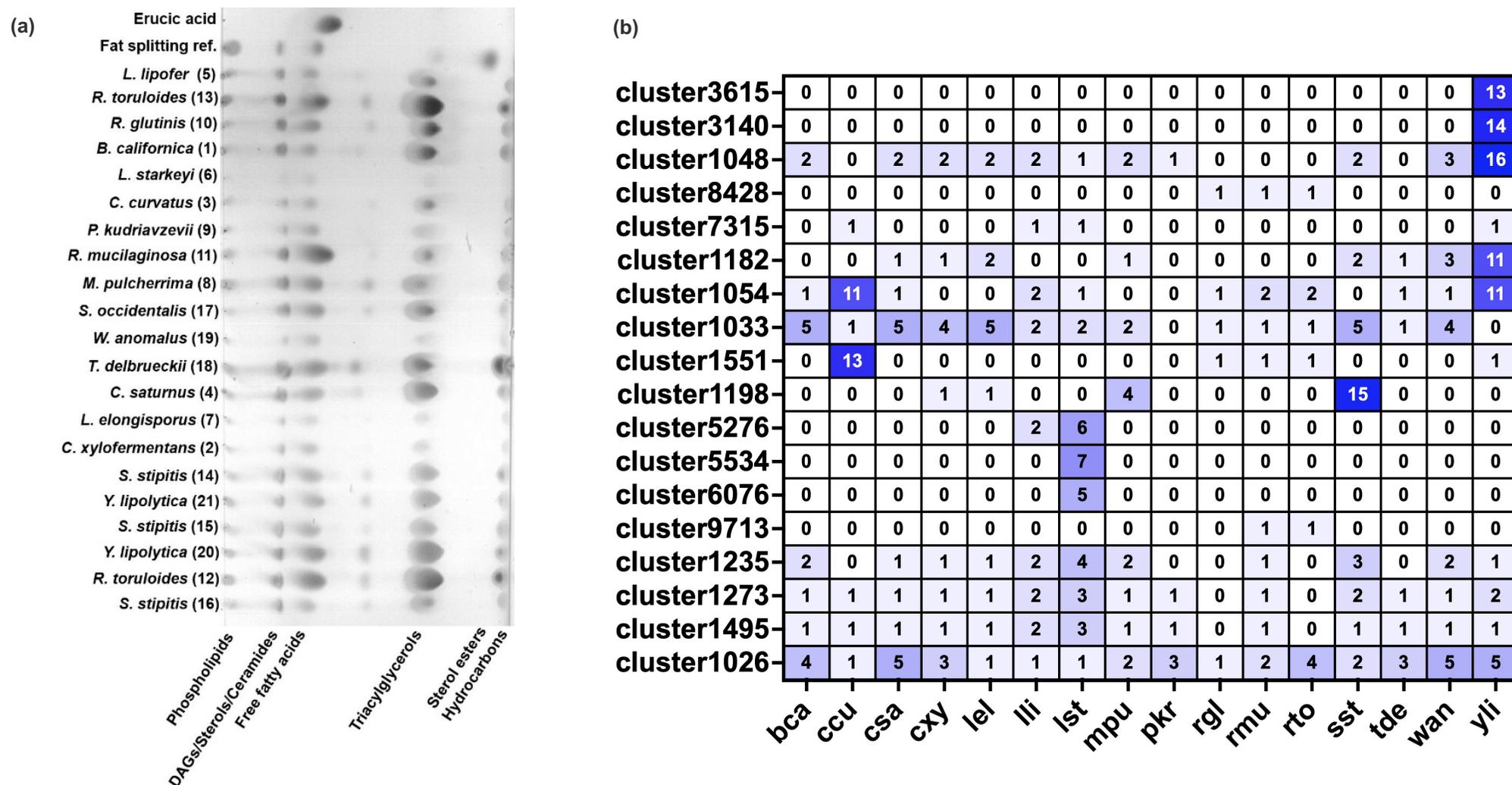


Figure 6-3 Oleaginous yeasts lipid metabolism at glance

(a) Thin-layer chromatography (TLC) of lipid neutral extracts from the strains identified in Table 4-2 and grown in a glucose-based media under nitrogen-limiting conditions. The TLC was developed in hexane:diethyl ether:acetic acid (70:30:1) and *p*-anisaldehyde was used as detection system. Due to their similar polarities, sterols, DAG and ceramides give spots of similar retardation factor (Rf). DAG = diacylglycerol; TAG = triacylglycerol; Fat splitting ref. = high erucic acid rapeseed oil hydrolyzed 2h with a porcine lipase used here as standard along with erucic acid. (b) Heatmap representing for each strain the number of genes represented in clusters related to lipid metabolism and highly contributing to inter-species differences. Points are labelled using species 3-letter codes listed in Table 4-2.

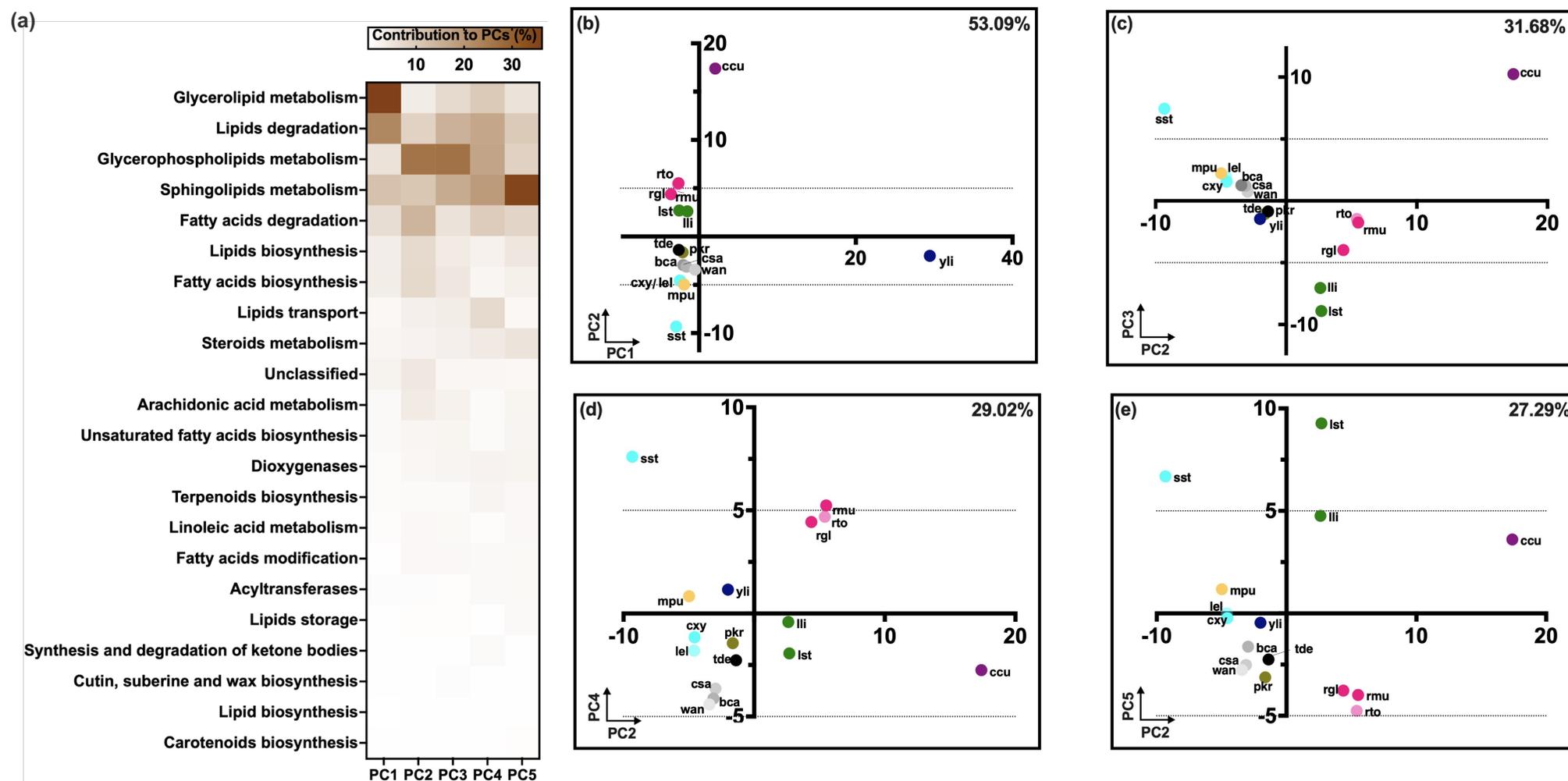


Figure 6-4 Principal component analysis of genes associated with lipid metabolism for 16 unconventional yeast species

Analyzed genomes and species 3-letter codes are listed in **Table 4-2**. (a) Contribution (%) of each functional category to different principal components (PC). (b-e) PCA score plots. The contribution of the represented PCs is as follows: PC1(32.78%); PC2(20.31%); PC3(11.37%); PC4(8.71%); PC5(6.98%); PC6(4.45%); PC7(3.33%) (for more refer to **Appendix 4**). The cumulative contribution (%) of the 2 PCs displayed is given for each plot. To help the visualization of scale changes y-axis -5 and +5 were marked with dotted lines on each plot.

from 1 to 3 genes except *T. delbrueckii* that does not contain genes encoding for this specific class of lipases. These lipases are sorted into various clusters including one totally specific to *Rhodotorula* species (cluster8428) (**Figure 6-3b**). Interestingly, the only lipase of class 3 of *C. curvatus* (the other *Basidiomycetes*) was clustered (cluster7315) with genes from the two *Lipomyces* species and *Y. lipolytica*. These numerous lipases might support *Y. lipolytica*'s ability to utilize exogenous lipids as carbon sources but also play a role in cellular lipid homeostasis. Actually, 5 of the 16 lipases of cluster1048 are part of the LIP family whose contribution to *Y. lipolytica* hydrophobic substrate utilization has been previously discussed (Fickers et al., 2005, 2011; Meunchan et al., 2015) but the other lipases remain to be functionally characterized. It is another degradation enzyme that distinguishes *Y. lipolytica* sphingolipid metabolism from the others as the strain genome encloses 11 genes encoding for sphingomyelin phosphodiesterases genes grouped with 1 to 2 genes per other species in a cluster (cluster1182) from which genes from *B. californica*, the 4 *Basidiomycetes sp.*, the 2 *Lipomyces* and *P. kudriazevii* are excluded (**Figure 6-3b**). These strains seem to be equipped with another class of sphingomyelin phosphodiesterase and some strains have both types but none have more genes encoding sphingomyelin phosphodiesterases than *Y. lipolytica*.

6.3.3.b - Comparative analysis of genes related to lipid metabolism: principal component 2 (PC2)

The difference between *C. curvatus* and the rest of the cohort are particularly well represented by the second principal component (PC2) that accounts for 20.3% of the global genetic variations in lipid metabolism of the 16 genomes (**Figure 6-4b**). *C. curvatus* genome stands out for its high content in putative oxalate:CoA ligase [EC:6.2.1.8]. Ten genes of *C. curvatus* encoding enzymes of this type are gathered in cluster1054, including two exhibiting domains suggesting both oxalate:CoA ligase

[EC:6.2.1.8] and either 4-coumarate:CoA ligase [EC:6.2.1.12] or hydroxyacid-oxoacid transhydrogenase [EC:1.1.99.24]. Those 10 oxalate:CoA ligase-like genes admit another paralogous gene (cluster1054) associated with a long-chain acyl-CoA synthetase activity [EC:6.2.1.3]. Enzymes represented in the cluster1054 are various but all seem to involve a CoA ligase activity so genes from this cluster probably participate in CoA stock management. Four activities are predominantly represented (by decreasing occurrence): oxalate:CoA ligase [EC:6.2.1.8], 4-coumarate:CoA ligase [EC:6.2.1.12], phenylacetyl-CoA ligase [EC:6.2.1.30] and long-chain acyl-CoA synthetase [EC:6.2.1.3]. *C. xylofermentans*, *L. elongisporus*, *S. stipitis* and *M. pulcherrima* (the 4 species of the CTG clade) genomes do not seem to contain genes encoding any of the later activities (4-coumarate/phenylacetate/oxalate:CoA ligases) and the same goes for *P. kudriazevii*. Strains like *B. californica*, *C. saturnus*, *W. anomalus* (the 3 Phaffomycetaceae sp.), each count one oxalate:CoA ligase [EC:6.2.1.8] in cluster1054 like *L. lipofer* and *L. starkeyi* except that *L. starkeyi* genome also presents an additional enzyme with unclear specificity as it was annotated as both homologous to oxalate:CoA ligase [EC:6.2.1.8] and 4-coumarate:CoA ligase [EC:6.2.1.12]. *R. toruloides* and *R. glutinis* also contain oxalate:CoA ligase [EC:6.2.1.8] activities but *R. mucilaginoso* contains a protein with an ambiguous annotation and long-chain acyl-CoA synthetase [EC:6.2.1.3]. Only *Y. lipolytica* genome harbors nine phenylacetyl-CoA ligase [EC:6.2.1.30], one oxalate:CoA ligase [EC:6.2.1.8] and one 4-coumarate:CoA ligase [EC:6.2.1.12] for a total of genes in cluster1054 that is equal to *C. curvatus* representation in the cluster (**Figure 6-3b**). Furthermore, *C. curvatus* genome contains additional 14 oxalate:CoA ligase [EC:6.2.1.8] clustered in 3 different groups including cluster1551 that would have been *Basidiomycetes* specific if the cluster didn't encompass one ambiguously annotated gene from *Y. lipolytica* (**Figure 6-3b**). *Rhodotorula* sp. each has one oxalate:CoA ligase [EC:6.2.1.8] associated to cluster1551 (**Figure 6-3b**).

The involvement of long-chain acyl-CoA synthetase [EC:6.2.1.3] genes in lipid metabolism is obvious but it is not the main difference represented by cluster1054. Indeed, long-chain acyl-CoA synthetase activity is largely distributed across the 16 genomes over several other clusters including cluster1033 (another orthologous group highly contributing to PC2). So the long-chain acyl-CoA synthetases of cluster1054 do not per se constitute a singularity. However, closer scrutiny of the other activities (4-coumarate/phenylacetate/oxalate:CoA ligases) revealed possible misannotations. Phenylacetate:CoA ligases were never described in yeast and *Y. lipolytica* putative phenylacetate:CoA ligases share low to no identity with bacterial phenylacetate:CoA ligases (Uniprot IDs: P76085, O33469, Q9L9C1, P76085) but significant sequence similarities were found with a putative fungal phenylacetate:CoA (Uniprot id: D3GE78) that was also annotated as “4-coumarate-CoA ligase-like”. Even higher percentages of identity were found between *Y. lipolytica* putative phenylacetate:CoA and 4-coumarate:CoA ligases from *Arabidopsis thaliana* (Uniprot ids: Q42524, Q9S725, Q9LU36, F4I9T8). The same was observed for oxalate:CoA ligases from *C. curvatus*. They are organized in two orthologous clusters (1054 and 1551), proteins of each cluster share more identity with 4-coumarate:CoA ligases from *Arabidopsis thaliana* (Uniprot ids: Q42524, Q9S725, Q9LU36, F4I9T8) than with proteins of the other cluster. Further experimental testing should help resolve these specificity questions. An oxalate:CoA ligase was described in *S. cerevisiae* (Foster & Nakata, 2014) but oxalate formation in yeast was never characterized and barely more is known about oxalyl-CoA utilization that supposedly releases CO₂ and CoenzymeA (CoA) (Foster & Nakata, 2014). In some fungi, oxalate formation was described from oxaloacetate (Gadd et al., 2014) while in plants, oxalate can be derived from xanthine (Hafez et al., 2017), glycine or serine through glyoxylate oxidation (Brzica et al., 2013; Hafez et al., 2017). This echoes phenylacetate formation from phenylacetaldehyde derived from phenylalanine via the Ehrlich pathway (Vuralhan et al., 2003) and p-coumaric acid (4-coumarate) production from phenylalanine or

tyrosine (Liu et al., 2019). Phenylacetate, 4-coumarate, oxalate formation from amino acids or nucleic acids are coupled with ammonia release. Some of those pathways might not be active in yeasts but if they are, enzymes from cluster1054 could play a greater role than just contributing to CoA homeostasis: they could also participate in the utilization of by-products of ammonia recovery.

Genes involved in glycerophospholipids metabolism, fatty acids degradation and sphingolipids metabolism mainly contribute to PC2 (**Figure 6-4a**). Sphingolipids metabolism contribution to PC2 is mainly related to *S. stipitis* high representation in sialidase-1 [EC:3.2.1.18] (also known as acetylneuraminyl hydrolases). *S. stipitis* genome presents 13 sialidase-1 genes with no homology with any of the 4 human sialidasases. Sialidasases are enzymes with broad substrate specificity catalyzing sialic acid hydrolysis in substrates like glycans, glycolipids and glycoproteins (Eneva et al., 2021). Focusing on sphingolipids, sialidasases might be involved in ceramide formation from gangliosides (Miyagi & Yamaguchi, 2012). The presence of sialic acids was reported on the cell surface glycolipids of some microorganisms but sialic acid metabolism in fungi appears to be mainly studied with respect to pathogenicity, morphogenesis and saprophytism (Eneva et al., 2021).

Sialidasases are rarely found in the other investigated genomes. The cluster1198 regrouping the 13 sialidasases of *S. stipitis* is a CTG clade-specific cluster that also contains one sialidase from *C. xylofermentans* and homologous genes from *L. elongisporus* (1 gene) and *M. pulcherrima* (4 genes) annotated as “cell wall protein” or “hyphally regulated cell wall protein”. The homology found between the (hyphally regulated) cell wall proteins and the sialidasases reinforces the possibility of sialidasases being involved in cell wall sialic acid metabolism and hyphal formation but whether these enzymes have a preference for glycolipids or glycoproteins remains unclear. Only one sialidase was found

outside of cluster1198. It belongs to *L. lipofer* (llg1429) and appears to be a unique gene of the pan-genome. Unlike *S. stipitis* sialidases, llg1429 exhibits between 26.8 and 29.8 % identity with the four human sialidases and 77.9% identity with *Aspergillus fumigatus* sialidase (KDNase) (Telford et al., 2011). If llg1429 cellular localization matches the one of the KDNase (Yeung, 2015) it is probably not a cell wall protein.

6.3.3.c - Comparative analysis of genes related to lipid metabolism: principal component 3 (PC3)

The contribution of functional categories to PC2 suggests that *Y. lipolytica*, *T. delbrueckii*, *P. kudriazevii*, the 3 *Phaffomycetaceae*, *C. xylofermentans*, *L. elongisporus* and *M. pulcherrima* have very similar glycerophospholipids metabolism, fatty acids degradation and sphingolipids metabolism (**Figure 6-4b**). Furthermore, these functional categories seem to unite *Lipomyces* and *Rhodotorula* and isolate *C. curvatus* and to a lesser extent *S. stipitis* for aforescribed reasons (**Figure 6-4b**). It should be noted that the genetic overlapping of genes involved in those categories is so good for very close relatives like *C. xylofermentans* & *L. elongisporus* or *R. toruloides* & *R. mucilaginosa* that they are difficultly discernible by projection on PC2 (**Figure 6-4b**).

Functional categories contributing to PC2 also strongly contribute to PC3 and especially glycerophospholipids metabolism and sphingolipids metabolism (**Figure 6-4a**) so much that half of the 10 clusters contributing the most to PC2 also are also among the 10 contributing the most to PC3. Differences between PC2 and PC3 lie in categories such as glycerolipids metabolism and fatty acid degradation that respectively contributed more to PC3 and PC2 (**Figure 6-4a**). Given the functional categories contribution similarity observed between PC2 and PC3, genomes distribution according to these two principal components also isolates *C. curvatus* and *S. stiptis* while regrouping *Y. lipolytica*, *T. delbrueckii*, *P. kudriazevii*, the 3 *Phaffomycetaceae*, *C. xylofermentans*, *L.*

elongisporus and *M. pulcherrima* on one hand and *Lipomyces* and *Rhodotorula* on the other hand (**Figure 6-4c**). Among the 10 clusters contributing the most to PC3, one is *Lipomyces*-specific (cluster5276) and two are exclusive to *L. starkeyi* (cluster5534 and cluster6076) (**Figure 6-3b**). Cluster5276 and cluster5534 mainly contain hypothetical and dubiously-annotated proteins but draw the attention to the over-representation of phospholipases C (PLC) in *Lipomyces* species. *L. starkeyi* genome counts 8 phospholipases C with high confidence annotations and 12 with dubious annotations status and *L. lipofer* presents 7 PLCs with high confidence annotations and 8 with dubious annotations status.

The analysis of *L. starkeyi* and *L. lipofer* PLCs revealed scars of transposases as in both genomes 7 genes (6 dubious and 1 confirmed) shared high identity (43.45 to 84.21%) with 8 to 35% of the total length of a transposase-like protein from the fungi *Metarhizium robertsii* (ncbi iD: KID80896.1). This finding is consistent with the results of another study showing transposon remains surrounding the many phosphatidylinositol-specific phospholipase-C genes identified in the fungi *Ceratocystis cacaofunesta* and possibly constituting hints on the evolutionary mechanisms of expansion of this gene family (Molano et al., 2018).

Even with the exclusion of dubious PLC, *Lipomyces* encompass an important number of PLC when most strains have 2 PLCs with high confidence annotations except for: *Rhodotorula sp.* that contains none, *P. kudriazevii* that have 3 additional PLCs including one with dubious annotation status and *Y. lipolytica* & *C. curvatus* that also include 1-2 dubious PLCs. No genes from *Rhodotorula sp.* were annotated “phospholipase C” or were evaluated homologous to other PLC of the pan-genome but a few were loosely annotated “Phospholipase/Carboxylesterase”. Those “Phospholipase/Carboxylesterase” might ensure the functions of PLC in *Rhodotorula sp.* Yet the absence of PLC is probably viable since scPLC1 deletion in *S. cerevisiae* results in weakened but viable mutants

(Barman et al., 2018). PLC catalyzes the hydrolysis of phosphatidylinositol 4,5-bisphosphate, yielding diacylglycerol (DAG) and inositol-1,4,5-trisphosphate (IP3). PLC and the 2 second messengers it produces are entangled in a lot of signaling pathways including pathways regulating cell cycle, nutrient sensing, cell wall integrity or response to various stresses such as high temperature or osmotic shock (Barman et al., 2018; Djordjevic, 2010). Regarding lipid synthesis, the DAG can contribute to triacylglycerol (TAG) synthesis and the IP3 will be involved in the regulation of Mth1-induced glucose catabolite repression that can have an influence on acetyl-CoA homeostasis from glucose which can in turn influence lipid accumulation (Galdieri et al., 2013). IP3 influence on acetyl-CoA homeostasis is probably carbon source dependent because not all carbon sources can elicit a carbon catabolite repression/derepression. As for PLC's impact on lipid accumulation, PLC activities have been linked to carotenoid accumulation in *Neurospora crassa* (Barman & Tamuli, 2015) and induced by nitrogen starvation in *Mortierella alpine* to support membrane phospholipids recycling into TAG (Lu et al., 2020). Similar types of recycling phenomena (autophagy and related macrolipophagy) leading to increased release of free fatty acids during nutrient starvation (Garcia et al., 2018; Singh et al., 2009) have been described in yeasts including *R. toruloides* (Zhu et al., 2012). The autophagy observed in both *R. toruloides* and *M. alpine* is very likely mediated by the mTOR signaling pathway (Lu et al., 2020; Zhu et al., 2012) and scPLC1 have been described to interact with Tor2p (Barman et al., 2018) which suggests PLC might influence lipid accumulation at multiple levels. The fact that *Lipomyces* species are bearing 4 times more PLC than the other investigated species might be correlated with their well-established lipid accumulation abilities (Dien et al., 2016) and could also have ramifications on their responses to stress.

6.3.3.d - Comparative analysis of genes related to lipid metabolism: PC4 and beyond

Similarly, to what is observed between PC2 and PC3, half of the 10 clusters contributing the most to PC3 are also among the 10 contributing the most to PC4. In terms of functional categories contributions: a higher contribution of categories glycerophospholipids metabolism and fatty acids/lipids biosynthesis is recorded for PC3 compared to PC4 when PC4 seems to be more affected by categories like fatty acids degradation and lipid transport than PC3 (**Figure 6-4b**). A projection according to PC4 finally separates *Lipomyces* and *Rhodotorula*, gathering this time the *Lipomyces sp.* with *S. stipitis* while the remaining of the studied group is represented similarly by PC4 even though the PC4 enables a small distinction between *R. torulooides* and *R. mucilaginosa* (**Figure 6-4e**).

Differences observed between *R. torulooides* and *R. mucilaginosa* remain slim: over the 444 clusters and unique genes considered for this comparison, the 2 genomes only differ on 68 features including 7 unique genes specific to *R. torulooides* and 3 unique genes specific to *R. mucilaginosa*. The differences between *R. torulooides* and *R. mucilaginosa* contributing the most to PC4 are the presence of *R. mucilaginosa* gene in clusters1235, 1273 and 1495 when *R. torulooides* is not represented in those clusters (**Figure 6-3b**). And cluster1026 contains 2 genes from *R. mucilaginosa* against 4 from *R. torulooides* (**Figure 6-3b**). The cluster1026 is a core cluster containing mainly alcohol dehydrogenases [EC:1.1.1.1]. Genes of *R. mucilaginosa* in cluster1026 are one alcohol dehydrogenases [EC:1.1.1.1] and a hypothetical protein sharing sequence identity with mannitol-1-P dehydrogenase [EC:1.1.1.17] when *R. torulooides* counts 2 alcohol dehydrogenases [EC:1.1.1.1] and 2 protein sharing sequence identity with mannitol-1-P dehydrogenase [EC:1.1.1.17]. Focusing only on alcohol dehydrogenases [EC:1.1.1.1], 6 additional enzymes were found in *R. torulooides* and 4 more in *R. mucilaginosa*. A

duplication event probably occurred but it is hard to evaluate the consequences of such events on phenotypes especially because alcohol dehydrogenases are versatile and redundant. The cluster1495 regroups lipases ATG15 [EC:3.1.1.3] (also known as tweenesterase), an activity that *R. mucilaginosa* shares with all the other genomes except *R. toruloides* and *R. glutinis* (**Figure 6-3b**). Phenotypically, the absence of tweenesterase will probably result in a reduced range of hydrophobic substrate utilization abilities.

The last two clusters (1273 and 1235) not only contribute to the differences between *R. toruloides* and *R. mucilaginosa* but also explains the similarity between *Lipomyces sp.* and *S. stipitis*. Only *Lipomyces sp.*, *S. stipitis* and *Y. lipolytica* have more than one copy of acetyl-CoA C-acetyltransferase [EC:2.3.1.9] grouped in cluster1273 (**Figure 6-3b**). Most of the other strains only count one acetyl-CoA C-acetyltransferase [EC:2.3.1.9] except for *W. anomalus* genome that potentially encodes a second that seems to have been partially predicted (61 aa) and for *R. glutinis* and *R. toruloides* genome that encodes none. Variations in acetyl-CoA C-acetyltransferase [EC:2.3.1.9] representation might not have an impact on lipid production since the reaction catalyzed by the enzyme (condensation of 2 acetyl-CoAs into acetoacetyl-CoA) (Lynen & Ochoa, 1953) can also be performed by acetyl-CoA acyltransferase 1/2 [EC:2.3.1.16] (Haapalainen et al., 2007). None of *R. toruloides* proteins was annotated acetyl-CoA C-acetyltransferase [EC:2.3.1.9] but 3 acyltransferases 1/2 [EC:2.3.1.16] were found so the possible absence of acetyl-CoA C-acetyltransferase [EC:2.3.1.9] might represent a genetically important difference between *R. toruloides* and *R. mucilaginosa* but not have a huge impact on their respective phenotypes.

Finally, cluster 1235 includes 4 genes from *L. starkeyi* and 3 from *S. stipitis*, 2 from *L. lipofer*, *B. californica*, *W. anomalus*, *M. pulcherrima* when the other strains are only

represented by 1 gene (rmu/yli/cxy/lcl/csa) or none (ccu/pkr/rgl/rto/tde) (**Figure 6-3b**). The involvement of cluster 1235 in lipid metabolism is less clear because the primary function of the Major Facilitator Superfamily (MFS) transporter of the *dha1* family is to act as drug:proton antiporters and confer resistance to some drugs/xenobiotics (dos Santos et al., 2014). But the knock-out of some proteins of this family has been described to affect lipid homeostasis in *C. albicans* (K. Redhu et al., 2016) and the co-regulation of membrane lipid composition and drug transport to achieve optimal response to drug exposure has been suggested (dos Santos et al., 2014) although the involved signaling pathways remain elusive. To summarize, *Lipomyces sp.* and *S. stipitis* are brought closer by their content in MFS transporter of the *dha1* family and acetyl-CoA C-acetyltransferase but as described singularities (PLC for *Lipomyces sp.* and sialidases for *S. stipitis*) are contributing to *Lipomyces* species and *S. stipitis* isolation from the rest of the cohort according to PC4.

S. stipitis segregation according to PC5 is still mainly related to the sialidase representation which causes sphingolipids metabolism to be the main contributors to PC5 (**Figure 6-4a**). *S. stipitis* sialidases are gathered in cluster1198 that contributes 32.1% to PC5. Cluster2377 follows cluster1198 with a contribution of 2.9% to PC5 which only represent 0.2% of the overall differences in genes related to lipid metabolism as PC5 only accounts for 6.9% of the global genetic differences between strains. So, after PC4 features highlighted by the PCA only reflect minor variations between strains.

6.4 - Concluding remarks

After supporting the functional annotation efforts in **Chapter 4**, pan-genomic information proved to be especially useful for the identification of singular genomics features. Although uniqueness is an informative indicator it can sometimes reflect the phylogenetic composition of the studied group and is afflicted by the lesser reliability of unique gene annotation. Yet, the more reliable annotation of core and accessory genes

highlighted differences in lipid metabolism and gave new insight into the level of conservation of industrially relevant traits. Industrially relevant traits typically include robustness (McMillan & Beckham, 2017), strain performances in terms of growth and products accumulation (Kitcha & Cheirsilp, 2011; Lamers et al., 2016), the carbon utilization abilities (Sitepu et al., 2014c). Carbon utilization and evidence of robustness such as thermotolerance and halotolerance are advantageous for economic considerations as they will influence the cultivation conditions and production costs. The analysis revealed that despite a few familial trends no strong evidence supported the conservation of these industrially relevant features at genus level. The exploration of one genus should thus not be favored over the others. Additionally, properties extrapolation within a genus should be avoided and experimental re-assessment is recommended as often as possible except maybe for the use of cultivation temperatures at species level.

The rest of the analysis focused on annotated genes related to lipid metabolism. Previous phenotypic strain characterizations highlighted species like *Y. lipolytica* a model strain when it comes to oleaginicinity (Beopoulos et al., 2009) and *C. curvatus* which was the first oleaginous yeast to be used at industrial scale for CBE production (Cohen & Ratledge, 2015). Genetically, both species are quite distant from the other investigated species in terms of lipid metabolism. The comparative analysis enables us to go beyond the typical nitrogen-induced lipid accumulation scheme and associate *Y. lipolytica* with high numbers of lipases (class 3), sphingomyelin phosphodiesterases and putative phenylacetyl-CoA ligases. When *C. curvatus* was associated with high numbers of acyl-CoA synthase-like proteins. *Lipomyces* species were noticed for their putative content in phospholipase C, an enzyme that might influence lipid accumulation in many ways. Minor differences were also identified such as the important sialidase-1 content of *S. stipitis* that might play a role in cell wall modification. This type of feature might seem less important when it comes to oleochemicals production but the cell wall and

membrane composition can affect product extractability and so they should not be neglected.

Despite the low contribution of unique genes to PCs, genomes exhibiting the highest number of unique lipid genes (**Appendix 5**) (i.e. *Y. lipolytica*, *C. curvatus* and *Lipomyces sp.*) also stand out during the comparison by PCA except for *P. krudriazevii*. The later strain has a high number of unique genes involved in metabolism but was often associated with *T. delbrueckii* by PCA although both species are not part of the same taxonomic family. Both species having a low number of genes involved in lipid metabolism and no close relatives in the studied group, they are often excluded from a lot of accessory clusters which could cause them to be often grouped by PCA since the number of common/different genes was used as observation for the singular value decomposition. Similarly, the PCA also set *C. curvatus* apart from the other basidiomycetes species (*Rhodotorula*) and surprisingly *Rhodotorula sp.* lipid metabolism seems to be genetically closer to one of *Lipomyces species*. Yet, the analysis also confirmed the genetic proximity of taxonomically close species when it comes to lipid metabolism. Indeed, *C. saturnus* & *B. californica*, *R. toruloides* & *R. mucilaginosa* and *C. xylofermentans* & *L. elongisporus* were barely separable by projection on the PCs considered.

Although PCA, insights were gained regarding lipid-related functional category, mapped categories remain very broad. A better characterization of proteins from unconventional yeasts should enable a more accurate categorization. Database contents are continuously curated offering the prospect of rapid improvements. Despite these annotation inaccuracies, *Lipomyces sp.* encouraging genetic features contrast with the low intensities observed on the TLC (**Figure 6-3a**). Experimental comparisons albeit rapid and convenient can give very different results depending on how they are executed as shown

by the results of the biomass evaluation in two conditions (**Figure 6-2**). This type of discordance should prompt us to increase the number of combined phenotypic & genomic approaches so that a species potential is no longer exclusively limited to its experimental performances but also takes a genetic dimension into account. Genetic information unlike transcriptomic or proteomic data should not be too affected by experimental design but could be impaired by inter-strain variability. Overcoming inter-strain variability would require a better overlap between sequenced genomes and commercially available strains.

When genomes are available, comparative genomics can offer new insights into unconventional yeast genetics but cannot completely replace experimental strain comparison. Existing genetic information can support bottom-up engineering approaches and pave the way to exclusively computational trait prediction while improving the reconstruction of genome-scale models. This combined comparative phenotypic::genomic analysis offered interesting perspectives on traits relevant to the industrial production of oleochemicals by oleaginous yeast. But genomic data can also be used for the reconstruction of genome-scale models (GEM). GEM could then participate in a transition toward systematic computational strategies for biomanufacturing or contribute to a deeper functional exploration of certain genes by complementing tedious knock-out experiments. Yet, the lack of diversity in sequences availability and lack of characterization of unconventional yeasts is limiting the potential of omics-based explorations. For instance, the sialidases found in *S. stipitis* are not documented in yeasts, they were never described in *S. cerevisiae* despite a few fungal reports. Due to the low identity share with human sequences, inferring function would be inaccurate. As a lot of information cannot be inferred from well-documented species, unconventional yeast genomes must be subjected to more combined phenotypic/omics endeavors to reduce the knowledge gap and improve our understanding of industrially relevant microbial processes.

Part IV – Toward the enhancement of sweetwater utilization

Chapter 7 - Sweetwater: underrated crude glycerol as feedstock for sustainable lipid production in unconventional yeasts

7.1 - Introduction

When it comes to microbial lipid production (single cell oil), yeasts stand out for their amenability, versatile carbon utilization, short duplication times and potential for high lipid accumulation (Thevenieau & Nicaud, 2013). Hence, numerous studies have already investigated yeast lipid production from biodiesel-derived crude glycerol but yields remain low possibly due to the inhibitory effects of certain impurities and especially methanol (Gao et al., 2016; Raimondi et al., 2014; Samul et al., 2014). Impurities are not the only aspect limiting crude glycerol microbial utilization as high glycerol concentrations also have adverse impacts on cell growth (Muniraj et al., 2015). When using crude glycerol for bioconversions, an efficient solution is to dilute the waste to lower both glycerol and impurities concentrations (Ito et al., 2005). This solution, albeit impractical at an industrial scale, initiated reflections around biodiesel-crude glycerol feeding strategies (Karamerou et al., 2017; Poontawee & Limtong, 2020; Signori et al., 2016). In parallel, impurities removal and crude glycerol purification techniques have multiplied (Isahak et al., 2015) to improve crude glycerol usability or try to lower the refining costs when other sources of crude glycerol possibly more suitable for microbial growth could be investigated.

Because of the predominance of biodiesel-derived crude glycerol, glycerol by-products from other industrial sources were overlooked despite their good potential. Aside from biodiesel production, crude glycerol is historically discharged by two other processes: fat saponification and fat splitting. Biodiesel-derived crude glycerol valorization is essential for the economic viability of the process when the other two processes depend less on the by-product valorization. Yet the escalating excess of unused crude glycerol poses a

disposal issue with both environmental and economic consequences. Especially for fat splitting that generates 30 % of the global glycerol production compared to 6 % for the soap industry (Ciriminna et al., 2014; Pagliaro, 2017). Post-fat-splitting glycerol refining facilities are currently hardly profitable, but they are still operational. Using these facilities, sweetwater of different purity levels could be used as feedstock for lipid production (**Appendix 6**). The crudest version SW15 has a 15% (w/v) glycerol concentration. It is characterized by its low pH and high content in brown fatty residues (**Appendix 6**). Cleaner versions of the waste can also be obtained with glycerol concentrations of respectively 35% (w/v) and 85% (w/v): SW35 and SW85. Whatever form of sweetwater is considered, given its methanol-free composition, the fat-splitting waste might be more compatible with use as feedstock than biodiesel crude glycerol. Finding new uses for sweetwater could alleviate the competition with biodiesel-derived crude glycerol for refined glycerol production but should certainly tackle the liability that sweetwater disposal represents. In this regard, to the best of my knowledge, no study has earlier examined the ability of a panel of yeasts to biotransform sweetwater into lipids although sweetwater was previously valorized using the yeast *Starmerella bombicola* to produce sophorolipids (Wadekar et al., 2012).

In this chapter, the suitability of sweetwater as feedstock was evaluated in a range of yeast species looking at both growth and lipid production. The tolerance to increasing glycerol concentrations of each strain was first assessed as this aspect has previously been identified as a limiting factor in crude glycerol use. The growth in the crudest form of the waste was then tested in presence and absence of nutrients. Then, the effect of sweetwater initial pH adjustment was investigated. Finally, the lipid production in sweetwater was compared across strains using the screening method described in **Chapter 5**. Which revealed that sweetwater can promote lipid accumulation in certain strains. **Chapter 6** highlighted the importance of combined phenotypic/genotypic

studies but since the genotypes of genes involved in glycerol metabolism (data not shown) didn't reflect the phenotypic differences observed the investigation was taken up to the transcriptomic level. This sweetwater effect was further investigated by comparing the transcriptomic changes induced in the most promising strain opposing growth in sweetwater to growth in refined glycerol with and without nutrients.

7.2 - Materials and methods

7.2.1 - Media and sweetwater-based preparations

Sweetwater batches used in this study are described in **Chapter 3** (section **3.1.1**). For the evaluation of pH effect, the pH of SW15 was adjusted to 6 ± 0.1 (@ 20°C) by the addition of sodium hydroxide (1 M). SW1.5SC solutions were prepared by aseptically diluting sterile SW15, SW35 or SW85 with sterile 2X SC nutrient mix (3.8 g/L Yeast Nitrogen Base, 1.58 g/L complete supplement mixture and 1.44 g/L ammonium sulfate) and sterile ultrapure water. The SW35SC was prepared by direct dissolution of the SC nutrient mix (1.9 g/L Yeast Nitrogen Base, 0.79 g/L complete supplement mixture and 0.72 g/L ammonium sulfate) in non-sterile SW35 followed by autoclaving.

GlySCs were prepared as described in **Chapter 3** (section **3.2.1**) using microbiology-grade glycerol. The ammonium sulfate amount was adjusted to ensure an initial C/N molar ratio of 60 whether 15, 20, 80 or 160 g/L of glycerol (tolerance to glycerol) was used. A 150 g/L glycerol media was also prepared without nutrients with ultrapure water and microbiology-grade glycerol (transcriptomic analysis).

7.2.2 - Cell culture

Other media preparation, cell maintenance and culture were performed as described in **Chapter 3** (section **3.2**). Growth-related comparisons were performed in microplates. Then, kinetic parameters were determined using the R package GrowthCurver (Sprouffske & Wagner, 2016) for each replicate individually. Replicates that could not be fitted to the following logistic growth equation were excluded.

$$N_t = \frac{CC}{1 + \left(\frac{CC - N_0}{N_0}\right)e^{-rt}}$$

Where N_0 is the initial population size and N_t the population size at a time t . CC represents the carrying capacity which is the maximum possible population size in a given environment. As for r , it is the intrinsic growth rate of the population. The parameters computed for the remaining replicates were used for statistical comparison. When more than 2 conditions were compared, an ANOVA followed (when applicable) by Tukey's post-hoc test was performed. For the lipid analysis and transcriptomic experiments, centrifuge tubes or flasks were preferred to microplates. Inocula were washed twice before cells resuspension in the appropriate medium.

7.2.3 - Nile red staining

Nile red staining was performed as described in **Chapter 5 (section 5.3)**. Technical replicates were averaged and associated standard deviations were plotted using GraphPad Prism for time-course monitoring of lipid accumulation. Only the mean was plotted for the comparison of lipid content at 3 and 10 days in SW15 and SW1.5SC but statistical difference was evaluated using an ANOVA followed (when applicable) by Tukey's post-hoc test.

7.2.4 - Lipid extraction and TLC

For the lipid extraction, *R. toruloides* (12) was grown into SW15 and SW1.5SC for 4 and 10 days in 250-mL flasks. Cells from the whole working volume were harvested. The extraction and TLC were performed as described in **Chapter 3 (section 3.2)**.

7.2.5 - RNA extraction and sequencing

For transcriptomic analysis, cells were grown into SW35>15, SW35>1.5SC, 15% (w/v) glycerol and 1.5% (w/v) glycerol with SC. The cultures were allowed to reach a mid-exponential phase. Three aliquots of 10 mL of culture were then harvested by

centrifugation for each condition (5000xg, 5 min). Cells were then suspended in 1 mL of RNALater[®] to stabilize the RNA (Ambion, Life technologies, Carlsbad, USA). The suspensions were mixed for 1 hour at 4°C on a tube rotator. Cells were harvested by centrifugation (12.000xg, 5 min) and pellets were stored at -80°C. Given the density of RNALater, 14 mL of sterile DEPC-treated ultrapure water was sometimes added to dilute the RNALater and facilitate cells sedimentation. The procedure was executed in an RNA-free environment. Dedicated tips and tubes were used, and surfaces (biosafety cabinet, bottles, and pipettes) were decontaminated with RNaseZAP (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's instructions. A volume of 1 mL of TRIzol reagent (Ambion, Life technologies, Carlsbad, USA) was added to frozen pellets. Pellets were thawed on ice and resuspended in the TRIzol reagent. The resulting suspensions were transferred in tubes containing 250 µL of chilled acid-washed glass beads 425-600 µm (Sigma, St Louis, USA). Tubes were vortexed 15 s before 15 min cell disruption at 4°C using a TissueLyser II (Qiagen, Hilden, Germany) at 30 Hz. After cell disruption, tubes were centrifuged for 5 min at 12.000xg and 4°C. Supernatants were transferred to fresh tubes to which 200 µL of chloroform were added. The solution was vortex 15 s and incubated at room temperature for 5 min. The RNA extraction was performed according to the TRIzol manufacturer's instructions. Extracts were resuspended in 50 µL RNase free-water and treated with the Turbo DNase kit (Invitrogen) for removal of DNA contamination. Finally, samples were concentrated by ethanol precipitation as described by Green and Sambrook (Green and Sambrook, 2016) for a final volume of 30 µL (RNase free water). At each step, sample concentration and purity were monitored by NanoDrop. The quality evaluation was completed by running 1.5% (w/v) agarose gels after the extraction and concentration steps. Before library preparation, extract quality was confirmed by Bioanalyzer Agilent 2100 (Agilent, Palo Alto, USA) and NanoDrop 2000 (ThermoFisherScientific, Waltham, USA). PolyA mRNA enrichment was performed using oligodT beads then the mRNAs were

fragmented and cDNA library was synthesized Illumina TruSeq Stranded mRNA Library Preparation kit (Illumina, San Diego, USA) according manufacturer's instructions. A paired-end sequencing on NovaSeq with 150 bp (Illumina, San Diego, USA) read length. Library preparation, sequencing, adapters and low-quality reads trimming and quality control were performed by NovogeneAIT.

7.2.6 - Transcriptomic analysis

Trimmed reads were aligned to *R. toruloides* NRRL Y-1091 genome assembly (NCBI accession ID: GCA_001542305.1) using STAR v2.7.7 (Dobin et al., 2013). RSEM v1.3.3 (Li and Dewey, 2011) quantified the expression level based on a genome annotation generated by AUGUSTUS v3.3.3 (Stanke and Morgenstern, 2005) and functionally annotated using the method described in **Chapter 4 (section 4.2.3)**. When needed protein subcellular localization was predicted using the tool BUSCA (Savojarado et al., 2018). Expression levels are given in counts of Fragments Per Kilobase of transcript per Million mapped reads (FPKM) or Transcripts Per Million (TPM). DEseq2 v1.30.0 (Love et al., 2014) from Bioconductor 3.12 was used for count processing and differential expression analysis (DE).

First, data were filtered and only genes associated with TPM above 10 for more than 3 samples were kept. After DE, genes with adjusted p-value (p_{adj}) < 0.01 were considered significantly differentially expressed. Then data were normalized to correct the differences in sequencing depth by variance stabilizing transformation (VST) method (Anders and Huber, 2010). And normalized data were used for hierarchical clustering and principal component analysis (PCA). For the PCA and hierarchical clustering, the biological replicates were not aggregated. And the z-scores of normalized counts of relevant genes across conditions were used as indicators of inter-replicate variability.

7.2.7 - Genomic analysis

Some relevant genes were mined in an attempt to highlight inter-species genetic differences that could explain the phenotypic differences observed. Mining was achieved based on functional annotations, KO numbers (Mao et al., 2005) and by homology to other yeast species equivalents from the AYbRAH database (Correia et al., 2019).

7.2.8 - Phenol-sulfuric sugar assay

The presence of sugar in sweetwater was evaluated using a phenol-sulfuric assay. For the assay 200 µL of sample (blank, glucose standards, SW undiluted, SW diluted 100 or 100 times; water) were mixed with 200 µL of 5% phenol solution (Sigma-Aldrich, Saint-Louis, USA) and carefully supplemented with 1 mL of concentrated sulfuric acid (Sigma-Aldrich, Saint-Louis, USA). Samples were then heated at 80°C for 30 min. Once the samples have cooled down their absorbance was read at 490 nm. A 15% refined glycerol solution was used as blank and to prepare the glucose standards (final concentration varying from 10 to 70 µg/mL).

7.3 - Results and discussion

7.3.1 - Investigated strains and their glycerol tolerance

Twenty-one strains of 17 species (**Table 1-2**) were selected based on their high potential for industrial lipid production. Commercially refined glycerol was first used to determine strain tolerance to high glycerol concentrations without interference from sweetwater impurities. Strains were grown in 2, 8 and 16 % (*w/v*) of commercial glycerol supplemented with appropriate nutrients for yeast growth (SC mix) (**Table 7-1**) respectively be noted Gly2SC, Gly8SC and Gly16SC. The resulting growth curves were described in terms of computed carrying capacities (CC) and time required to reach half of the CC (tmid). All the strains were able to utilize Gly2SC but their responses to higher glycerol concentrations revealed a wide range of tolerances (**Table 7-1**). The most sensitive strains *L. starkeyi* and *L. lipofer* not only exhibited no growth in Gly16SC

but their growth in Gly8SC was characterized by the highest t_{mid} recorded at this concentration (i.e. 88.9 h and 77.7 h) (**Table 7-1**).

Table 7-1 Tolerance to refined glycerol.

Average carrying capacities (CC), times needed to reach half of the carrying capacity (t_{mid}) of the 4 technical replicates and associated standard deviations are summarized in the following table. The statistical analysis is provided in **Appendix 7**.

Strains	2% Glycerol SC (Gly2SC)		8% Glycerol SC (Gly8SC)		16% Glycerol SC (Gly16SC)	
	t_{mid} (h)	CC	t_{mid} (h)	CC	t_{mid} (h)	CC
<i>B. californica</i> (1)	18.09±4.38	1.47±0.02	17.99±4.72	1.63±0.06	25.98±1.88	1.61±0.04
<i>C. xylofermentans</i> (2)	19.93±0.45	0.92±0.04	33.95±10.95	1.33±0.03	28.5±2.78	1.34±0.05
<i>C. curvatus</i> (3)	21.08±0.29	1.18±0.01	22.93±0.38	1.34±0.02	30.88±2.68	1.16±0.11
<i>C. saturnus</i> (4)	16.45±0.19	1.5±0.02	17.52±0.07	1.78±0.02	22.16±0.99	0.82±0.11
<i>L. lipofer</i> (5)	50.65±13.13	1.17±0.3	77.66±1.26	1.34±0.04	-	-
<i>L. starkeyi</i> (6)	46.8±1.4	1.31±0.01	88.85±0.55	1.55±0.03	-	-
<i>L. elongisporus</i> (7)	27.71±3.96	1.53±0.12	40.43±8.69	1.44±0.22	18.54±0.43	1.67±0.04
<i>M. pulcherrima</i> (8)	20.69±2.62	1.48±0.05	18.05±0.05	1.7±0.03	18.52±0.06	1.68±0.02
<i>P. kudriavzevii</i> (9)	22.5±0.83	1.38±0.03	30.73±0.29	1.76±0.06	78.38±8.69	1.72±0.1
<i>R. glutinis</i> (10)	21.14±0.14	1.23±0.04	22.75±0.87	1.42±0.02	-	-
<i>R. mucilaginosa</i> (11)	24.63±3.44	1.46±0.07	26.9±10.2	1.81±0.04	29.63±7.08	1.69±0.03
<i>R. toruloides</i> (12)	23.78±0.32	1.35±0	23.78±0.06	1.84±0.03	73.61±0.54	1.84±0.03
<i>R. toruloides</i> (13)	25.91±3.45	1.35±0.11	24.31±0.54	1.62±0.05	84.08±4.71	1.78±0.1
<i>S. stipitis</i> (14)	22.22±2.29	1.52±0.04	22.57±1.59	1.87±0.03	70.72±1.15	1.73±0.03
<i>S. stipitis</i> (15)	26.35±6.21	1.42±0	29.89±2.1	1.77±0.05	29.89±2.1	1.55±0.05
<i>S. stipitis</i> (16)	20.86±3.71	1.42±0.06	30.03±6.18	1.87±0.05	43.8±4.18	1.82±0.05
<i>S. occidentalis</i> (17)	19.68±3.87	1.07±0.01	19.5±2.94	1.56±0.07	54.37±4.63	1.56±0.03
<i>T. delbrueckii</i> (18)	17.93±2.41	0.91±0.03	18.2±0.09	1.42±0.06	43.48±2.44	1.44±0.04
<i>W. anomalus</i> (19)	16.6±0.08	1.54±0.03	17.85±0.02	1.89±0.02	19.35±0.28	1.84±0.02
<i>Y. lipolytica</i> (20)	20.78±0.94	1.46±0.09	19.48±0.02	1.74±0.02	23.09±1.17	1.82±0.04
<i>Y. lipolytica</i> (21)	17.29±0.4	1.52±0.07	17.1±0.17	1.74±0.03	17.82±0.1	1.62±0.07

R. glutinis was also unable to grow in Gly16SC but only required 22.7 h to reach half of its carrying capacity in Gly8SC (**Table 7-1**) which suggest a higher tolerance to high glycerol concentration than *Lipomyces* species. *C. saturnus*, *C. curvatus*, *B. californica*, *S. stipitis* (n°14, 15 and 16), *P. kudriavzevii* and *R. toruloides* (n°12 and 13) demonstrated an increased tolerance to glycerol. Like *R. glutinis*, these strains exhibited little or no

significant tmid difference in Gly2SC and Gly8SC. However, in contrast to *R. glutinis*, they were able to grow in Gly16SC.

But their growth in Gly16SC was characterized by lower carrying capacities and/or growth rates than the parameters observed in Gly2SC or Gly8SC. Despite *C. saturnus* ability to grow in Gly16SC, the strain displays the lowest ratio of carrying capacities in Gly16SC and Gly2SC (i.e. $CC(16\%)/CC(2\%)=0.5$) which suggests a tolerance to Gly16SC higher than *R. glutinis* and the *Lipomyces* species but lower than the other strains. Then, with tmid in Gly16SC and Gly2SC ratios (i.e. $tmid(16\%)/tmid(2\%)$) close to 1, *W. anomalus*, *M. pulcherrima*, *Y. lipolytica* (n°20 and 21), *R. mucilaginoso* and *L. elongisporus* rank among the most high-tolerance strains. Indeed, those strains are growing in Gly16SC at a rate allowing tmid values close to the ones achieved in Gly2SC. Finally, even though no strain growth in Gly16SC outperformed the growth at lower concentrations, *C. xylofermentans*, *S. occidentalis* and *T. delbrueckii* appear as the most tolerant strains. The inability to achieve higher cell densities in Gly16SC indicates that all the strains suffer from a certain level of inhibition at high glycerol concentrations. Although the most resistant strains do not escape the inhibition in Gly16SC, their growth in Gly8SC is characterized by carrying capacities 1.5 times higher than the ones in Gly2SC denoting a lesser inhibition in Gly8SC glycerol than the one afflicting the other strains (Table 7-1).

7.3.2 - Evaluating strain affinity with the crudest form of the waste

The dilution of the waste with synthetic media was previously described to improve growth in biodiesel-derived crude glycerol (Ito et al., 2005). Such strategy albeit efficient is impracticable at industrial scale as nutrient addition weight on the production cost

Table 7-2 Investigating sweetwater as a feedstock for yeast growth.

Average carrying capacities (CC) and times needed to reach half of the carrying capacity (t_{mid}) of the 4 technical replicates are summarized in the following table. Averages computed with less than 4 replicates were underlined. The statistical analysis is provided in **Appendix 8**.

Strains	Concentration effect				pH effect			
	SW _{1.5} SC		SW ₃₅ SC		SW ₁₅ pH 6.09		SW ₁₅ pH 3.95	
<i>B. californica</i> (1)	<u>9.45±0</u>	<u>1.4±0</u>	19.75±0.08	1.62±0.01	21.23±0.57	1.05±0.04	22.68±2.85	1.05±0.02
<i>C. xylofermentans</i> (2)	23.7±1.82	0.96±0.08	60.82±4.33	1.91±0.04	39.67±2.92	1.2±0.03	49.2±6.36	0.99±0.11
<i>C. curvatus</i> (3)	21.26±1.15	1.46±0.01	64.98±2.76	1.42±0.01	41.53±0.88	1.24±0.04	93.65±8.76	1.29±0.24
<i>C. saturnus</i> (4)	13.43±2.48	1.45±0.03	26.13±0.73	1.68±0.04	29.25±0.45	1.32±0.02	31.29±0.48	1.29±0.02
<i>L. lipofer</i> (5)	59.51±3.32	0.89±0.09	171.86±27.74	0.69±0.25	79.73±4.38	0.8±0.05	66.07±3.93	0.35±0.06
<i>L. starkeyi</i> (6)	47.11±0.56	0.99±0.02	31.75±11.92	0.08±0.02	66.81±1.62	0.75±0.04	67.86±1.22	0.59±0.06
<i>L. elongisporus</i> (7)	18.34±0.08	1.33±0.03	35.9±3.4	1.85±0.02	21.51±0.95	1.08±0.1	19.88±0.18	1.15±0.02
<i>M. pulcherrima</i> (8)	19.11±0.18	1.38±0.01	38.26±0.28	1.67±0.01	19.83±0.09	1.41±0.01	20.24±0.17	1.42±0.02
<i>P. kudriavzevii</i> (9)	22.13±0.42	1.3±0.02	36.96±2.06	1.54±0.01	27.56±5.35	0.87±0.26	54.06±16.75	0.73±0.11
<i>R. glutinis</i> (10)	24.18±0.27	1.26±0.02	103.3±4.95	1.54±0.03	27.18±0.72	1.29±0.03	54.23±1.99	1.32±0.02
<i>R. mucilaginosa</i> (11)	20.03±0.07	1.3±0.01	22.38±0.1	1.67±0.02	20.6±0.37	1.1±0.03	23.85±0.97	1.02±0.02
<i>R. toruloides</i> (12)	23.03±1.76	1.57±0.05	35.37±0.23	1.73±0.03	21.74±0.35	1.34±0.03	<u>28.07±1.83</u>	<u>1.28±0.02</u>
<i>R. toruloides</i> (13)	21.48±0.62	1.35±0.02	55.84±0.66	1.78±0.03	24.06±1.75	1.15±0.05	42.43±2.47	1.13±0.04
<i>S. stipitis</i> (14)	20.1±0.61	1.45±0	139.16±0.64	1.82±0.05	28.12±6.3	1.3±0.05	21.68±1.7	1.08±0.07
<i>S. stipitis</i> (15)	19.44±1.13	1.35±0.05	37.01±0.43	1.76±0.02	21.27±0.32	1.23±0.03	23.87±2.81	1.09±0.04
<i>S. stipitis</i> (16)	23.76±0.62	1.28±0.02	33.69±1.95	1.73±0	20.71±0.18	1.26±0.03	21.18±0.63	1.17±0.1
<i>S. occidentalis</i> (17)	19.38±0.24	1.12±0.05	24.21±0.09	1.73±0.03	33.83±1.08	1.45±0.01	49.01±1.93	1.24±0.03
<i>T. delbrueckii</i> (18)	18±0.3	0.81±0.06	26.64±0.34	1.65±0.04	21.56±1.06	1.02±0.06	21.09±0.95	0.57±0.07
<i>W. anomalus</i> (19)	13.71±1.88	1.37±0.04	122.47±3.04	1.98±0.03	20.27±0.07	1.41±0.01	21.74±0.35	1.84±0.02
<i>Y. lipolytica</i> (20)	20.43±1.42	1.39±0	19.99±0.12	1.83±0.03	21.59±1.03	1.21±0.04	20.75±0.12	1.32±0.01
<i>Y. lipolytica</i> (21)	<u>17.27±1.52</u>	<u>1.45±0.05</u>	18.31±0.32	1.69±0.01	32.84±10.87	1.33±0.01	19.78±0.3	1.32±0.02
	t_{mid} (h)	CC	t_{mid} (h)	CC	t_{mid} (h)	CC	t_{mid} (h)	CC

could endanger the economic viability of a microbial oil production process. For this reason, the use of SW15, the crudest form sweetwater, without prior dilution could be particularly attractive cost-wise. Of course, SW35 and SW85 have some advantages like their lower impurity contents and lower moisture levels that make them more suitable for storage than SW15. But given strains' tolerance to glycerol, SW35 and SW85 could not be used without prior dilution and/or nutrient addition hence the growth in SW15 was first evaluated. The 21 strains were able to grow in SW15 even *L. lipofer* and *L. starkeyi* exhibited some growth in SW15 (**Table 7-2**) despite their incapacity to grow in Gly16SC (**Table 7-1**). This confirms SW15 suitability for yeast growth without further nutrient addition and positively differentiates SW15 from Gly16SC. Yet the OD₆₀₀ recorded in SW15 (**Table 7-2**) remains lower than the one reached when strains are grown in the reference media SW1.5SC (**Table 7-1**) with an average CC of 1.11 in SW15 against 1.3 in SW1.5SC. Strains with the highest carrying capacities in SW15 are *M. pulcherrima*, *W. anomalus*, the two *Y. lipolytica* strains and *R. glutinis*.

Previous studies showed that pH adjustments could modulate yeast growth in glycerol (Chiruvolu et al., 1998; Razavi & Marc, 2006; Swinnen et al., 2016). *R. mucilaginosa* and *R. toruloides* (12) barely responded to the adjustment of SW15 initial pH. An adjustment of SW15 initial pH from 3.95 (native pH) to 6.09 did not significantly improve the panel growth in the waste. Indeed, the average growth rate of the cohort at both pHs was not found significantly different although some minor individual differences can be mentioned. The two *Y. lipolytica* strains that slightly favors the native pH and *R. glutinis*, *R. toruloides* (13), *C. curvatus*, *T. delbrueckii* and *S. occidentalis* that display a marked preference for the pH 6. Overall, when observable, the pH-related improvements were minor suggesting that nutrient starvation is likely to be the factor limiting strain growth in SW15. Still, SW15 seems to contain enough nutrients to promote cell proliferation. SW15 composition even proved to have a beneficial effect on

the response of certain strains to high glycerol concentration. But the nutrients contained in SW15 might reach a limiting concentration faster than when commercial nutrients are added to the media thus limiting the cell densities attainable with the waste.

Looking only at growth-related parameters, four strains demonstrated a strong predilection for SW15 utilization: *M. pulcherrima*, *Y. lipolytica* (20 and 21) and *W. anomalus*.

7.3.3 - Comparing the lipid production of 21 strains grown in SW₁₅ and SW_{1.5}SC

With the idea that an average cell density could be balanced by a good lipid accumulation, the lipid production in SW15 was then compared across the 21 strains. Lipid accumulation is tightly linked to cell growth and since strains exhibited a wide range of growth rates in sweetwater-based media (**Table 7-2**), lipid content was analyzed after 3 and 10 cultivation days rather than a single time point (**Figure 7-1a**). Looking at average fluorescence signals across the 4 samples, the best lipid accumulators are *L. starkeyi* with an average normalized fluorescence signal of 2863.3 RFUcd, *R. toruloides* (12) (2077.5 RFUcd), *C. saturnus* (1557.7 RFUcd), *L. elongisporus* (1405.4 RFUcd) and *R. glutinis* (1397.9 RFU) (**Figure 7-1**).

When it comes to lipid accumulation in SW15, the same strains are top-ranked but *L. elongisporus* is replaced by *M. pulcherrima* (**Figure 7-1a**). This Nile red-based comparison of lipid contents shows that the strains like *W. anomalus*, *Y. lipolytica* (n°21 and 20) and *M. pulcherrima* that stand out for their growth in SW15 are not the greatest lipid accumulators. On the contrary, *L. starkeyi* one of the strains with the slowest growth in sweetwater ranks among the best lipid-accumulator.

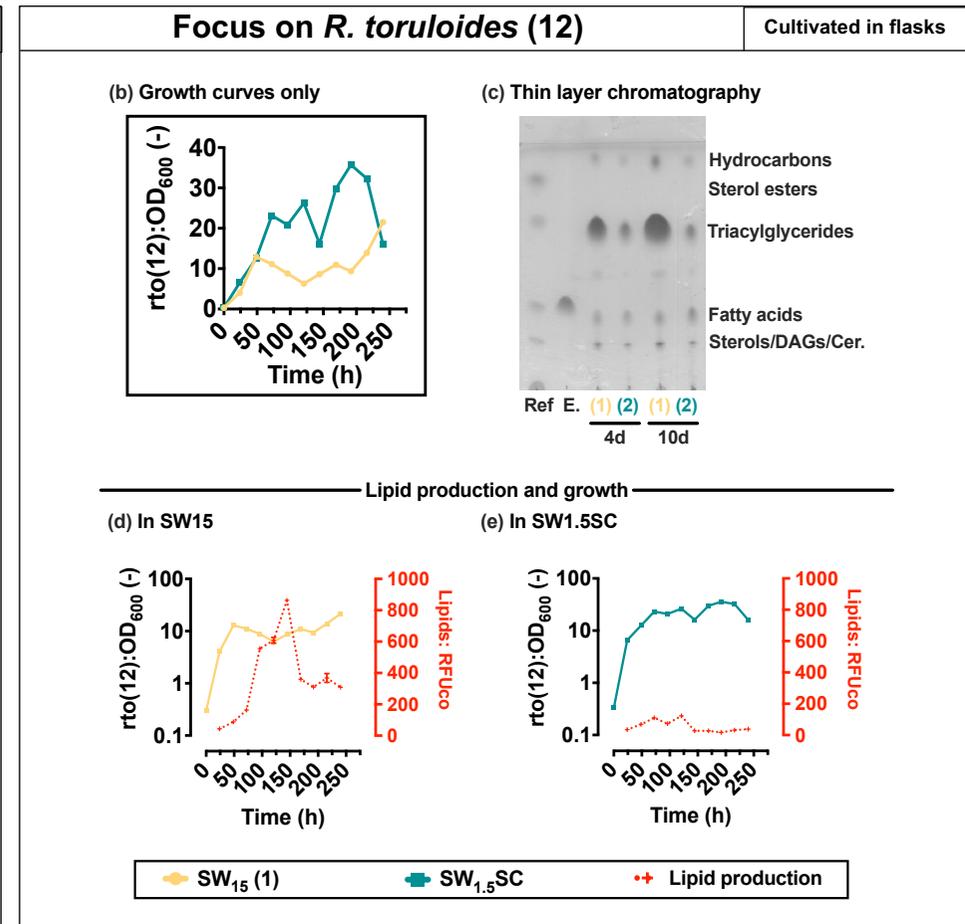
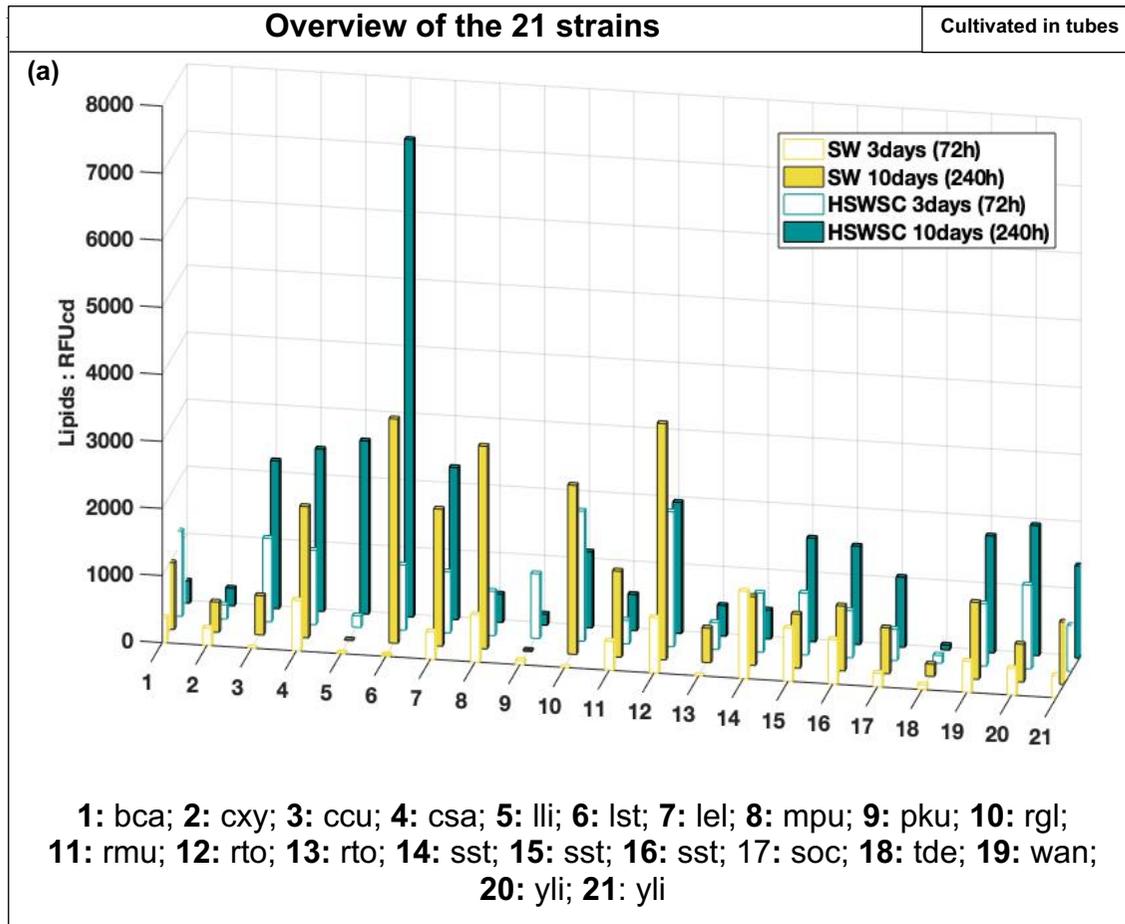


Figure 7-1 Lipid production in sweetwater based media.

The fluorescence of Nile Red stained cells was measured at 625 nm in quadruplicates and normalized by the OD_{600} of the sample (RFUco) or normalized by dilution factor (RFUcd). (a) **Comparison of the 21 investigated strains.** Strain n° and 3-letter codes were used to designate the strains. Please refer to **table 3-1** for full species names. Cells were cultivated in 50-mL centrifuge tubes in sweetwater diluted in synthetic complete nutrient mix [teal bars: SWSC] and in sweetwater with a 15% (*w/v*) glycerol content [yellow bars: SW₁₅]. Samples were taken after 3 cultivation days [empty bars] and after 10 days [filled bars]. For clarity, standard deviations were not plotted but taken into consideration for the comparison and available in **Appendix 9**. (b) **Comparison of *R. toruloides* (12) growth in different media.** Cells were cultivated in shake-flasks in sweetwater diluted in synthetic complete nutrient mix [teal line: SWSC] and in sweetwater with a 15% (*w/v*) glycerol content [yellow line: SW₁₅]. (c) **Thin layer chromatography of lipid extracts from *R. toruloides* (12)** Cells were grown in SW₁₅ (1) or SWSC (2) for 4 or 10 days (abbreviated d). The plate was developed in hexane: diethyl ether: acetic acid (70:30:1) and iodine was used as detection system. (d-e) **Comparison of *R. toruloides* (12) lipid accumulation profiles in different media.** For an improved, visualization of the different growth phases, the OD_{600} were plotted on a logarithmic scale [full lines: growth curves]. After the 24 first cultivation hours, the lipid content was monitored daily by staining samples with Nile red. The resulting values and corresponding standard deviation were then plotted against time [red dotted line: lipid production]. Corresponding microscope images are given in **Appendix 10**.

SW15 seems to promote a lipid accumulation in certain strains compared to its diluted counterpart. This phenomenon will thereafter be discussed as the SW-induced lipid accumulation. Seem to be concerned, *R. mucilaginosa*, *S. stiptitis* (14), *M. pulcherrima* and *C. saturnus* that on average accumulate 1.2 to 2.4-time more in SW15 than in SW1.5SC (**Figure 7-1a**). *R. mucilaginosa*, *S. stiptitis* (14), *M. pulcherrima* and *C. saturnus* are the strains with the most marked difference in accumulation independently of the time point but they are not the only strains potentially susceptible to SW-induced lipid accumulation. Strains accumulating more lipids in SW15 than in SW1.5SC at either time point (i.e. *B. californica*, *M. pulcherrima*, *S. stipitis* (14) and *C. saturnus*, *R. glutinis*, *R. toruloides* (12), *R. mucilaginosa*, *T. delbrueckii*) (**Figure 7-1a**) could also be sensitive to SW-induced lipid accumulation but confirming that hypothesis would require a time-course comparison.

Since *R. toruloides* (12) is the best lipid accumulator in SW15 among these 8 strains, its susceptibility to SW-induced lipid accumulation was verified with time-course monitoring of its lipid accumulation at a larger scale. Shake-flask cultivation confirmed that *R. toruloides* (12) achieves lower OD₆₀₀ than *W. anomalus* grown in the same conditions (data not shown). Furthermore, the comparison of the maximum normalized fluorescence signals recorded for *R. toruloides* (12) in SW1.5SC and SW15 demonstrates the strain sensitivity to SW-induced lipid accumulation since the signal observed in SW15 is 7.1 higher than the maximum signal in SW1.5SC (**Figure 7-1d/e**). Given the morphological differences observed between cells grown in SW15 and the one grown in SW1.5SC (**Appendix 10**), the OD-based normalization of Nile Red fluorescence data can be challenged so the lipid accumulation difference was additionally confirmed on TLC (**Figure 7-1c**). It should be noted that the tendency of *R. toruloides* to form clumps and elongated structures (**Appendix 10**) prevents appropriate cell enumeration in certain conditions explaining why OD-based normalization was preferred here.

The first source of differences in lipid accumulation across all these strains might be a kinetic aspect. Since strains exhibit different growth rates (**Table 7-2**), the kinetic of lipid accumulation might vary from one strain to another and from one media to another for a given strain. Yet, other factors (**Table 7-3**) might contribute to the observed differences.

Table 7-3 Mining for putative acetyl-CoA synthetases (ACS) and ATP:citrate lyases (ACL).

Species	Strains	Acetyl-CoA synthetases (ACS)	ATP:citrate lyases (ACL)
<i>Barnettozyma californica</i>	UCD 09	bca bcg1031 bca bcg4651	-
<i>Clavispora xylofermentans</i>	UFMG-HMD23.3	cxy cxg3577 cxy cxg415	-
<i>Cutaneotrichosporon curvatus</i>	ATCC 20509	ccu ccg1319	ccu ccg3068
<i>Cyberlindnera saturnus</i>	NRRL Y-17396	csa csg2471 csa csg57	-
<i>Lipomyces lipofer</i>	NRRL Y-11555	lli llg3241 lli llg984 (Partial) lli llg985(Partial)	lli llg3288 lli llg3287
<i>Lipomyces starkeyi</i>	NRRL Y-11557	lst ODQ74542.1	lst ODQ70679.1 lst ODQ70680.1
<i>Lodderomyces elongisporus</i>	NRRL YB-4239	lel XP_001525932.1 lel XP_001526839.1	-
<i>Metschnikowia pulcherrima</i>	PRJNA508581	mpu A0A4P6XN54 mpu A0A4P6XXF7	-
<i>Pichia kudriavzevii</i>	CBS 5147	pkrl pkg3201 pkrl pkg3343	-
<i>Rhodotorula glutinis</i>	ATCC 204091	rgl EGU11560.1 (Partial)	rgl EGU12112.1
<i>Rhodotorula mucilaginosa</i>	JGTA-S1	rmu rmg4716	rmu rmg2559 rmu rmg2560
<i>Rhodospiridium toruloides</i>	ATCC 10788	rto rtg931	rto rtg2312
<i>Scheffersomyces stipitis</i>	CBS 6054	sst XP_001385263.1 sst XP_001385819.1	-
<i>Torulaspora delbrueckii</i>	CBS 1146	tde XP_003682513.1 tde XP_003683506.1	-
<i>Wickerhamomyces anomalus</i>	NRRL Y-366	wan XP_019041050.1 wan XP_019041696.1	-
<i>Yarrowia lipolytica</i>	CLIB 89 (W29)	yli AOW06729.1	yli AOW04580.1 yli AOW06401.1

Nitrogen starvation has been described to decrease the intracellular concentration of adenosine monophosphate (AMP)(Yoshino & Murakami, 1982) inhibiting a key enzyme of Krebs cycle, the isocitrate dehydrogenase (ICDH), which triggers a citrate

accumulation in mitochondria. Citrate is then exported to the cytoplasm and can be used to produce acetyl-CoA catalyzed by the ATP: citrate lyase (ACL) in organisms expressing this enzyme (Adrio, 2017). Hence, high C:N ratio can promote lipid accumulation through citrate accumulation in yeast equipped with ACL (Boulton & Ratledge, 1981). No genome was available for *S. occidentalis* but at least one gene putatively encoding ACL was found in the genomes of *R. toruloides*, *R. glutinis* and *C. curvatus*, *R. mucilaginosa*, *Y. lipolytica*, *L. starkeyi* and *L. lipofer* (**Table 7-3**). The other strains must mainly rely on the wide-spread acetyl-CoA synthase (ACS) activity to produce cytosolic acetyl-CoA. In certain yeast, citrate accumulation induced by high C:N ratios can also activate the ACC catalyzing the first step of the fatty acid synthesis (Botham & Ratledge, 1979).

Sensitivity to C:N ratio-induced lipid accumulation varies a lot across strains. Different strains might be sensitive to different C:N thresholds and even for a given strain the C:N ratio required to induce lipid accumulation might vary depending on the nature of the carbon and nitrogen sources (Sitepu et al., 2014a). These differences in sensitivity to nitrogen and carbon sources and concentration might account for some of the lipid accumulation differences observed across strains. Growth in SW15 (**Table 7-2**) suggested the presence of nutrients in the waste. Among those nutrients, there is maybe a nitrogen source that might set a C:N ratio particularly favorable to lipid accumulation in SW-sensitive strains. Dilution of the waste to SW1.5SC with a nutrient mix containing ammonium sulfate will change the C:N ratio explaining some of the lipid accumulation differences observed between cells grown in SW15 and SW1.5SC. The nitrogen content is not the only difference between SW15 and SW1.5SC. Indeed, SW1.5SC also contains number of vitamins that can support yeast lipid accumulation. Therefore, part of the inter-species variations (**Table 7-2a**) can be the reflect of different nutritional requirements. Species accumulating more lipids in SW1.5SC likely have a higher

dependency on exogenous nutrients. The last major difference between SW15 and SW1.5SC is the impurity concentration. Impurity in the form of fatty matter (e.g. free fatty acids, diacylglycerol, triacylglycerol, monoacylglycerol) could support either growth or lipid accumulation by acting as a secondary carbon source or joining the intracellular lipid pool. To the best of our knowledge, only *C. curvatus*, *L. elongisporus*, *S. stipitis*, *P. kudriavzevii*, *R. glutinis*, *R. toruloides*, *Y. lipolytica* (Beopoulos et al., 2009; Fabiszewska et al., 2019; Fukuda, 2013; Gao et al., 2016; Matatkova et al., 2017; Patel et al., 2019) have been described to be able of hydrophobic substrate utilization. But differences in preferences could be observed across species.

Using SW15 as feedstock for lipid oil production without prior dilution of the waste or nutrient supplementation is an attractive option. But it is undeniable that SW1.5SC promotes better growth than SW15. The fact that the undiluted waste seems to promote a better lipid accumulation in a few strains is an encouraging result since the improvement of cell densities in SW15 could be simply achieved by optimizing the feeding strategy or improving strain tolerance to the waste through adaptive evolution. There is a risk that improving growth in SW15 might result in the loss of SW-induced lipid accumulation. A better understanding of the effect of the waste on metabolism might help us identify the most suitable enhancement strategy.

7.3.4 - Investigating the transcriptomic changes prompted by the use sweetwater in *R. toruloides* (12)

Crude glycerol-induced lipid accumulation was previously described with biodiesel-derived crude glycerol after methanol removal and was imputed to fatty acids presence (Razavi & Marc, 2006). Another study also proposed a role of fatty matter in such phenomenon without elucidating the underlying mechanisms (Gao et al., 2016). Here we studied the transcriptomic changes promoted in *R. toruloides* (12) using sweetwater. To decipher the mechanisms of crude glycerol-induced lipid accumulation and identify the

molecular basis of the phenotypic differences observed between SW15, SW1.5SC, Gly15 and Gly1.5SC, the RNA sequencing of *R. toruloides* (12) grown in the four media was performed.

7.3.4.a - The global comparison reveals a predominance of nutrient-related responses

A global comparison including the 4 transcriptomes was conducted to identify the major transcriptomic differences elicited by SW15, SW1.5SC, Gly15 and Gly1.5SC. Hierarchical clustering confirmed that crude and refined glycerol are prompting different transcriptomic responses (Figure 7-2 and Figure 7-3).

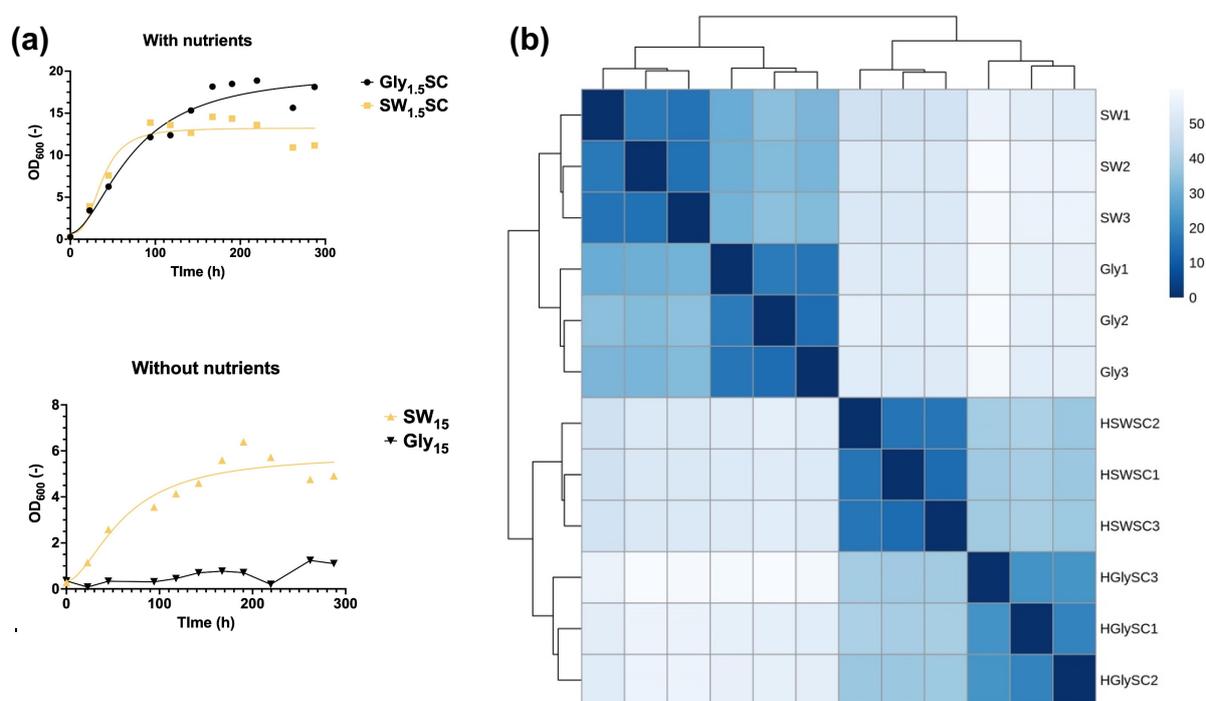


Figure 7-2 *R. toruloides* (12) in SW15, SW1.5SC, Gly15 and Gly1.5SC. (a) *R. toruloides* (12) growth curves in the investigated media (b) Clustergram of sample-to-sample Euclidean distances. The distances were computed based on expression counts normalized by variance stabilizing transformation.

The results of pairwise differential gene expression revealed a larger expression gap between samples grown with and without nutrients. More genes are significantly differentially expressed ($p_{adj} < 0.01$) when comparing samples with and without

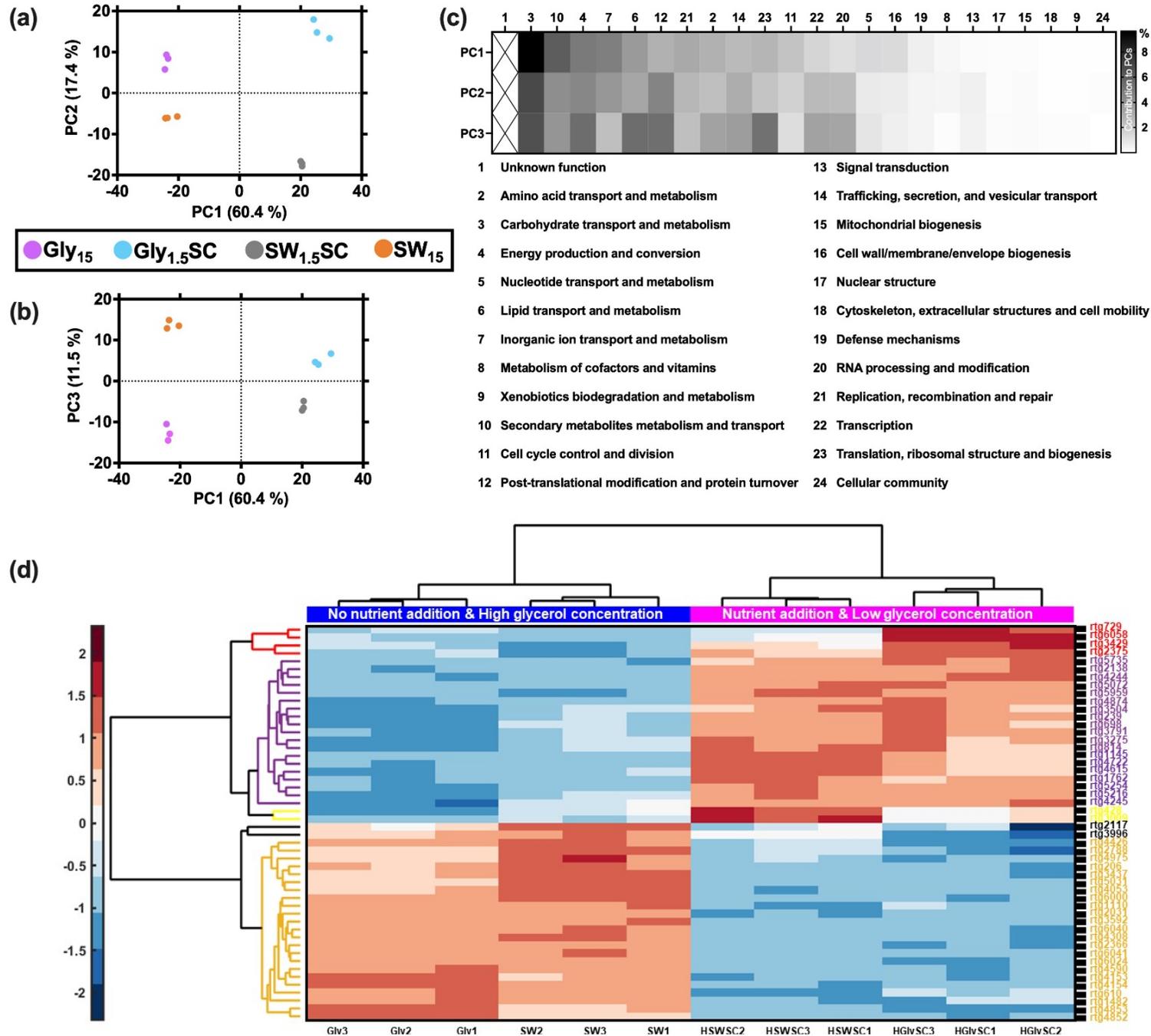


Figure 7-3 Comparison of rto (12) transcriptome in 4 glycerol-based media.

(a-c) Principal component analysis. Functional categories were sorted by descending contribution. Protein of unknown function were excluded from the heatmap for a better visibility but taken into account in the contributions distribution (d) Clustergram of the normalized expression counts of the 50 annotated genes contributing the most to PC1-3. The heatmap represents row-standardized z-scores. A more detailed gene list is available in Appendix 10.

nutrients (SW15 vs SW1.5SC(Ctrl) = 2472 genes; Gly15 vs Gly1.5SC(Ctrl) = 2563 genes) than when looking at the difference between cells cultivated in crude or refined glycerol (SW15 vs Gly15(Ctrl) = 1566 genes; SW1.5SC vs Gly1.5SC(Ctrl) = 1443 genes). Not only the number of regulated genes is higher, but the amplitude of the regulation is also stronger. Samples grown with/without nutrients exhibit log₂ fold changes (LFC) varying between -9.70 to 8.24 for SW15 vs SW1.5SC(Ctrl) and -9.35 to 9.03 for Gly15 vs Gly1.5SC(Ctrl) while the LFC resulting from glycerol type comparison are between -5.98 to 6.09 for SW15 vs Gly15(Ctrl) and -10.59 to 8.03 SW1.5SC vs Gly1.5SC(Ctrl). SW1.5SC vs Gly1.5SC(Ctrl) LFC range is misleading as only 3 genes are associated with those strong responses against 28 genes for differences between Gly15vsGly1.5SC(Ctrl).

A principal component analysis (PCA) analysis was then conducted. The PCA also confirms the predominance of nutrient impact since the first principal component (PC1) accounts for 60.4% of the overall differences while describing quite well the differences between samples grown with and without nutrients (**Figure 7-3a**). The second principal component (PC2) that represents 17.4% of the variations mainly reflects the differences between sweetwater and refined glycerol (**Figure 7-3a**). When projected according to PC2, the greatest distance recorded is between Gly1.5SC and SW1.5SC transcriptomes while a projection according to PC3 distances SW15 and Gly15 transcriptomes (**Figure 7-3a-b**). The third principal component PC3 also represents the differences between refined glycerol and sweetwater but only accounts for 11.5% of the differences (**Figure 7-3b**). Overall, 89.3% of the inter-sample variations are covered by PC 1 to 3 (**Figure 3a-b**). It should be noted that no correlation was found between the strength of the regulation and gene contribution to the principal components. Genes were sorted into 24 functional categories and the contribution of each category to PC 1 to 3 was computed (**Figure 7-3c**). Aware that, to a lesser extent, the PCs also reflect inter-replicate variability, the 50 functionally annotated genes contributing the most to PC 1 to 3 were

analyzed using hierarchical clustering of z-scores of normalized gene expression to visualize genes with high inter-replicate variability (**Figure 7-3d**). For a given condition, genes expression across replicates was found consistent for the 50 genes contributing the most to PC 1 to 3 (TOP50) (**Figure 7-3d**) which is coherent with replicate clusters observed in PCA (**Figure 7-3a-b**).

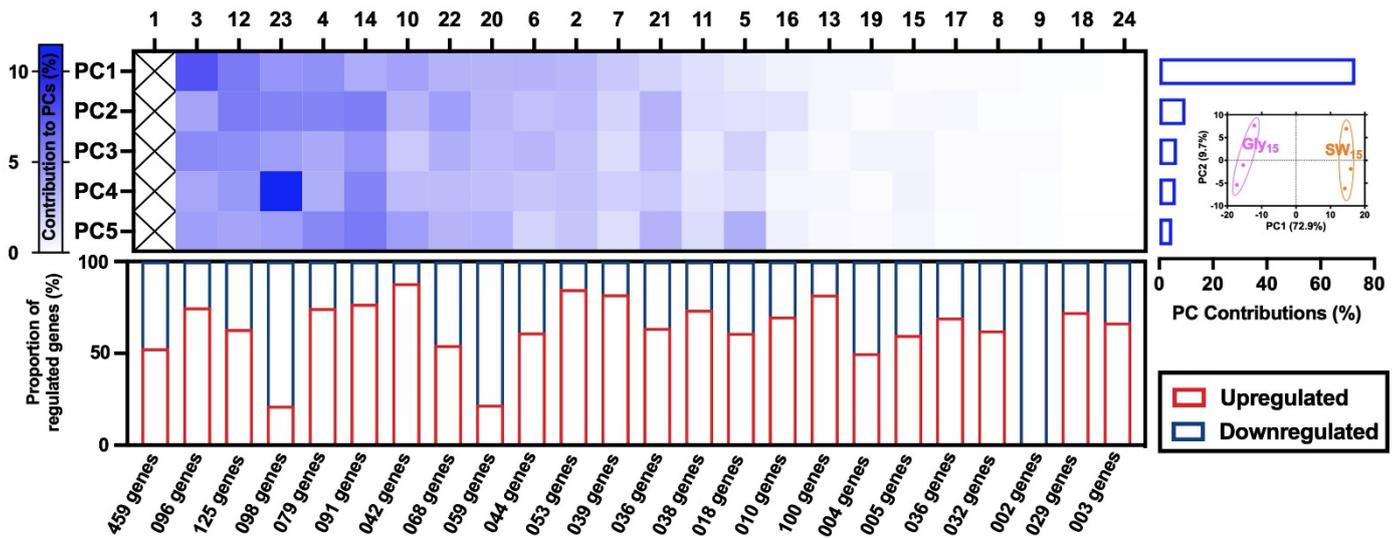
Looking at the transcriptomic differences between SW15, SW1.5SC, Gly15 and Gly1.5SC, it seems that *R. toruloides* (12) addresses the change of condition by adjusting primarily its carbohydrate, inorganic ion, secondary metabolites, lipid, and energy metabolism. The regulation of genes involved in these categories highlighted the importance of nitrogen and carbon catabolite repression. As well as the central role of the protein kinase Snf1 at the crossroad between resources-sensing signaling pathways and stress-response pathways. Even though the nutrient mix contains amino acids, nitrogenous bases and vitamins, these categories have a low contribution to the overall differences between sample but THI5 (a gene initiating thiamin synthesis pathway from pyridoxal phosphate) expression is intriguing as it suggests differences in thiamin or thiamin-precursor content between sweetwater and refined glycerol. A possible difference in nitrogen content between SW15 and Gly15 was also hypothesized and seem to play a major role in recorded phenotypes. While the inorganic ion availability in SW15 and Gly15 seems to induce scavenging transporters. Exception can be made of the Zn^{2+} availability that was found to be a concern mainly in Gly1.5SC compared to the other 3 conditions. An observation that was imputed to faster exhaustion of the ion in Gly1.5SC that can be linked to an increased cell proliferation. Genes involved in key cellular processes like cell cycle, replication, transcription, RNA modifications, translation, post-translational modifications contribute mainly to the difference between SW15 and Gly15 which is consistent with the more important growth phenotype gap observed between the two conditions. And overall, the nutrient availability contributed

more to the global transcriptomic differences than any sign of adaptation to high glycerol concentrations. It is possible that adaptation to high glycerol concentration is mainly mediated through post-translational modification or happens at a much early stage of growth.

7.3.4.b - Focus on the comparison between SW15 and Gly15

Since the absence/presence of nutrients was dominating the global comparison, another comparison was led using only the transcriptomes obtained in SW15 and Gly15. This second analysis should allow the investigation of possible mechanisms of lipid accumulation from sweetwater. Another PCA was performed only and the new PC1 represents 72.92% of the difference observed between cells grown in SW15 and Gly15 when the other PCs mainly describe inter-replicate differences (**Figure 7-4**). An in-depth analysis of the first 50 annotated genes contributing the most to the new PC1 should help link major transcriptomic changes to phenotypic observation recorded between cells grown in SW15 and Gly15. Reference to elements highlighted during the global comparison will sometime be made when they bring additional levels of evidence. As the following discussion focuses on the comparison between SW15 and Gly15 transcriptomes, genes were considered upregulated when they were more expressed in cells grown in SW15 and Gly15. And conversely, genes were considered downregulated when less expressed in cells grown in SW15 than in Gly15. Growth is significantly better in SW15 than in Gly15 (**Figure 7-2**) because SW15 might contain nutrients to support growth compared to Gly15. SW15 unique composition is also enabling changes in lipid accumulation. The global analysis already suggested a possible difference in nitrogen content between SW15 than in Gly15. But the difference in C:N ratio might not be the only contributor to higher strain performances in SW15. As a by-product of fat splitting

(a)



(b)

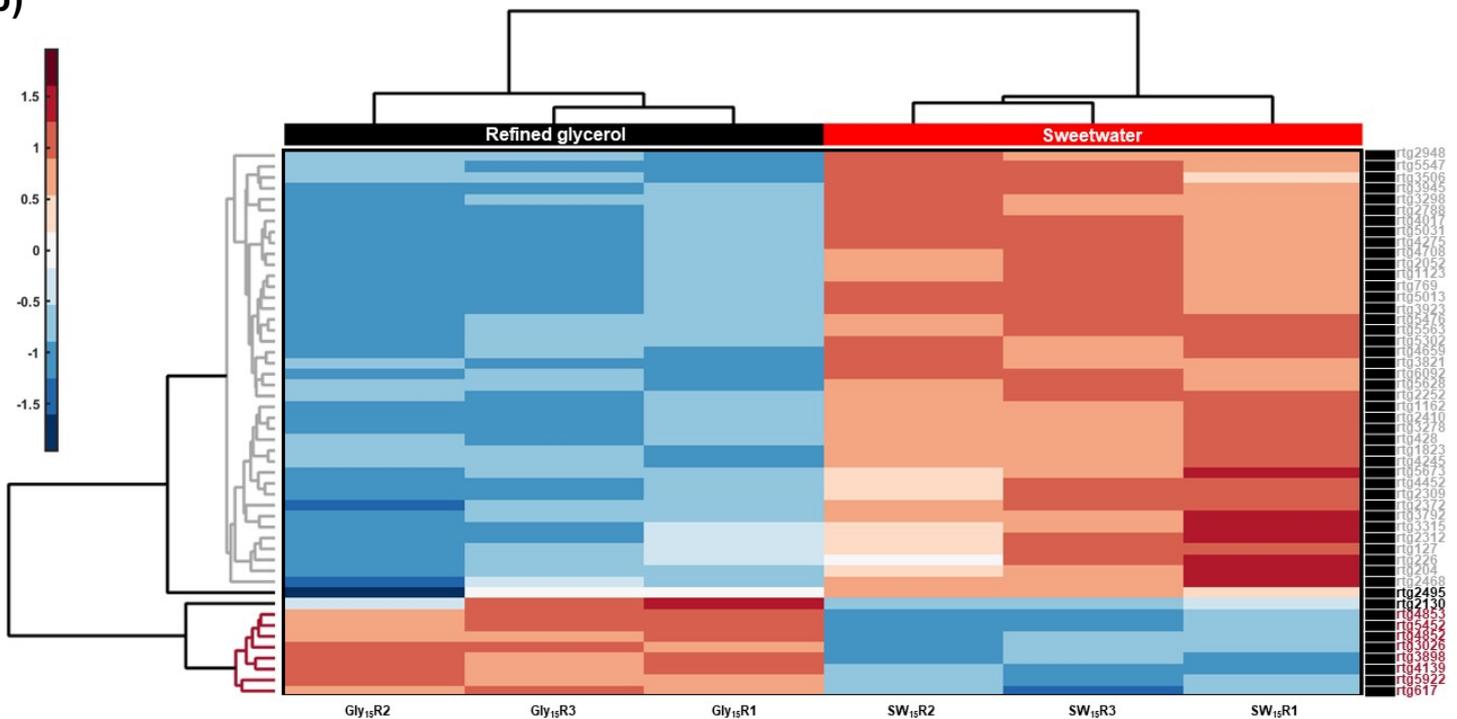


Figure 7-4 Comparison of *R. toruloides* (12) transcriptome in SW15 and Gly15 (Ctrl).

(a) **Principal component analysis and differential expression.** Each column represents a functional category. The categories are described in **Figure 7-3**. The upper part of the diagram is a heatmap illustrating the contribution of each functional category to the principal components (PC1-5) describing SW₁₅ and Gly₁₅ transcriptomes. The functional categories were sorted by descending global contribution to the PCs. The percentage of variance explained by each PC is given as a bar diagram in the upper right corner. And in the lower part of the diagram, the proportion of up-regulated or down-regulated genes compared to the control condition (Gly₁₅) is given for each functional category. (b) **Clustergram of the normalized expression counts of the 50 annotated genes contributing the most to PC1.** The heatmap represents row-standardized z-scores. A more detailed gene list is available in **Appendix 11**.

process, SW15 also contains fatty matter in the form of tri-/di-/mono-glycerides and free fatty acids, these fatty elements might act as a secondary carbon source or be internalized to inflate the lipid content of the strain.

Genes encoding proteins involved in nitrogen management

The global analysis hinted at differences in nitrogen content between SW15 and Gly15 and this second analysis confirms this trend (**Figure 7-3** and **Figure 7-4b**). When cells are grown in refined glycerol Gly15 they express enzymes such as nitrate/nitrite transporter *rtg4852* and the nitrite reductase *rtg4853* (**Figure 7-3** and **Figure 7-4b**). *Rtg4853* is homologous to genes known to be under nitrogen catabolite repression: *umNAR1* from *Ustilago Maydis* and *ncNIT-6* from *Neurospora Crassa*. As such they are induced by the lack of ammonia and/or the presence of nitrate (Banks et al., 1993; Johns et al., 2016). The presence of nitrate in Gly15, is unlikely as ultrapure water was used to dilute a biotechnology grade glycerol (nitrate content not specified by the manufacturer but all the specified minerals are at trace level) hence *rtg4853* is probably induced by lack of ammonium/preferred nitrogen source. When cells are grown in SW15, the expression of *rtg4853* seems to be repressed suggesting that SW15 might contain ammonium or a preferred nitrogen source. In SW15 the preferred nitrogen source might take the form of oligopeptides as suggested by the high expression of oligopeptide transporters *rtg3315* and *rtg5673*. Other hints supporting this hypothesis are: *rtg6000* (ammonium transporter of Amt family-homologous to *scMEP2*) 1.8 times more expressed in SW15 and the upregulation of the amidase *rtg3792* that can enable ammonium production from arginine, tryptophan and phenylalanine could be involved in exogenous oligopeptides digestion and the resulting amino acids processed by *rtg3792* (**Figure 7-5** and **Appendix 10**). But given the expression of the ubiquitin C (*rtg5216*): 5-times more expressed in SW15 than in Gly15. It is also possible that *rtg3792* is acting on amino acids resulting from endogenous protein degradation which then suggests

differences in protein turnover/autophagy between SW15 and Gly15 that would be consistent with the high contribution of genes involved in post-translational modification and protein turnover (**Figure 7-4a**). Yet in both SW15 and Gly15, the nitrogen availability must be low since the global analysis suggested that cells cultivated in both conditions were possibly relying on nitrogen retrieval from guanine that resulted in both cases in an allophanate-induced expression of *rtg4153/4* (DUR3) and allantoin-sensitive expression of *rtg4590* (DAL5).

As mentioned in **Chapter 1 (section 1.2.2.b)**, crude glycerol analysis is often incomplete, and most reports will only describe the composition of crude glycerol in terms of MONG, ash, glycerol and water content (see **Appendix 2**). More comprehensive reports will further detail fatty acid and glyceride compositions but only a few studies dive into the elemental composition of crude glycerols. Hence, data available on nitrogen contents of crude glycerols are limited. Yet a recent comparison of crude glycerols from different sources revealed ([Kumar et al., 2019](#)) that the nitrogen contained in crude glycerol from fat splitting (0.136 % w/w) is 3.3-times higher than the one contained in crude glycerol from the soap industry or 1.7 to 9.7-times higher than nitrogen contents of biodiesel crude glycerols. As a raw material-specific impurity, nitrogen originates from inferior crude oils used as raw material in oleochemical industries (**Chapter 1 - Section 1.2.2.b**). Nitrogen in crude can thus be proteins (Ayoub and Abdullah, 2012) or other nitrogenous compounds (Jungermann & Sontag, 1991). As such, different nitrogen concentrations and forms can be expected depending on the process. So, although nitrogen was previously described in crude glycerols, sweetwater nitrogen content should be further characterized especially since this content seems to be highly variable between sources.

Genes encoding proteins involved the uptake of fatty matter

Intracellular trafficking, secretion, and vesicular transport (**Figure 7-4a**) also play an important role in the differences found between SW15 and Gly15 (**Figure 7-4a**), the upregulation of genes of this category possibly indicates a more active secretion/externalization activities in SW15 which would be consistent with the hypothesis of fatty matter after digestion by secreted proteins. However, the mechanisms of exogenous fatty acids influx remains elusive. Some of it could happen during endocytosis. The upregulation of SLA1 (rtg1123) and EPS15 (rtg2052), two proteins of the PAN1 complexes suggests that SW promotes endocytosis in *R. toruloides* (**Figure 7-3 and Table 7-3**). The PAN1 complex is involved in endosomes internalization during actin-coupled endocytosis (Martin et al., 2007). Endocytosis that could be involved in the incorporation of exogenous elements and/or a quick way to remodel cell membrane in response to stress (López-Hernández et al., 2020). With a β -oxidation less active in SW15 (**Appendix 10**), the hypothesis of strain utilization of the fatty matter as secondary carbon sources seems very unlikely. But the upregulation of the 6 lipases (**Appendix 11**) and especially rtg2252 (**Figure 7-5**) confirms a possible degradation of the fatty matter. Hence, in SW15, at this stage of cell growth, if there is an uptake of exogenous fatty matter, those fatty acids do not undergo β -oxidation but are probably directly joining the cellular fatty acid pool through *ex novo* lipid accumulation participating in the increase of lipid content (**Figure 7-5**). Furthermore, exogenous fatty acids have previously been described to affect $\Delta 9$ desaturase expression in *S. cerevisiae* (Bossie & Martin, 1989) so the upregulation of the $\Delta 9$ desaturase rtg2309 in SW15 could be another evidence in favor of exogenous fatty acids integration (**Figure 7-5**).

SW15 fatty matter doesn't seem to act as a secondary carbon source but instead it was found that the waste might contain sugars. Indeed, a range of glycosidase was

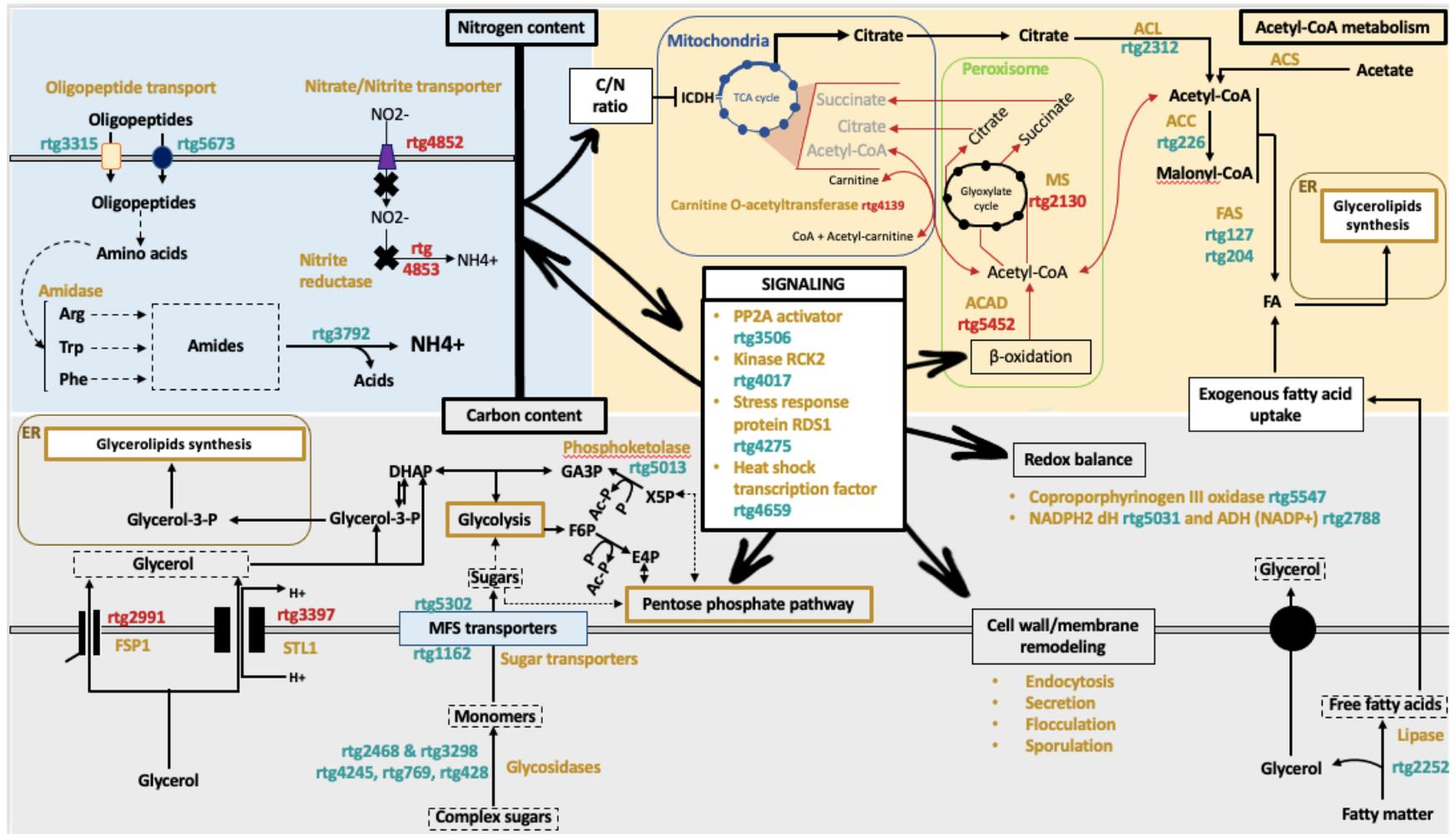


Figure 7-5 Scheme summarizing the main differences between *rto* (12) growth in SW15 compared to the growth in Gly15 (Ctrl).

Genes more expressed in SW15 than in Gly15 are in teal (upregulated) and genes less expressed in SW15 than in Gly15 are in crimson (downregulated). The function or name of relevant proteins is given in gold. A dashed lines indicate a possibility when full lines indicate more certainty. Abbreviations: ACAD=acyl-CoA dehydrogenase; ACC=acetyl-CoA carboxylase; ADH= alcohol dehydrogenase; ACL=ATP:citrate lysase; Ac-P= acetyl-phosphate; CO III= Coproporphyrinogen III oxidase; dH=dehydrogenase; DHAP=dihydroxyacetone; ER=endoplasmic reticulum; E4P=Erythrose-4-phosphate; FAS=Fatty acid synthase; F6P =Fructose-6-phosphate; GA3P= Glyceraldehyde-3-phosphate; ICDH= isocitrate dehydrogenase; MS=malate synthase; MFS=Major facilitator superfamily ;P=phosphate; PPP=Pentose phosphate pathway; PP2A=protein phosphatase 2A; RCK= Radiation sensitivity Complementing Kinase; RDS=Regulator of Drug Sensitivity; X5P= Xylulose-5-Phosphate.

upregulated in SW15 with activities varying from common glucosidase (rtg4245: EC:3.2.1.10; rtg428: EC:3.2.1.21) and β -fructofuranosidase (rtg769: EC:3.2.1.26) to less characterized such as a putative rhamnogalacturaonase (rtg2468) or a member of the glycosyl hydrolase family 88 (rtg3298) (**Figure 7-5**). Among these enzymes, the expression of the oligo-1,6-glucosidase (rtg4245) and of the β -fructofuranosidase (rtg769) contributed the most to the first principal component and were respectively upregulated by 11.1-fold and 6.5-fold when the other glycosidases were only 3.8 to 2.6 times more expressed (**Figure 7-5 and Appendix 10**). Two sugar transporters of MFS family (rtg1162 and rtg5302) were also upregulated and are very likely to contribute to sugar uptake from sweetwater (**Figure 7-5 and Appendix 10**). This upregulation of rtg1162 and rtg5302 comes with a downregulation of genes involved in glycerol uptake: rtg2991 (FSP1) and rtg3397 (STL1) which reminds a carbon catabolite repression suggesting that sweetwater might contain another carbon source than glycerol (Bommareddy et al., 2017).

Unrefined vegetable oils admittedly contain carbohydrates and proteins that given their hydrophilic nature are likely to end up in crude glycerol (Thompson & He, 2006). But sweetwater is derived from refined degummed rapeseed oil so it is very improbable that the initial vegetable oil is the source of sugars in SW15 case. Furthermore, the glucosidases rtg428 (no homolog in sc but identity with ecbglX) and rtg4245 (IMA1) were also highlighted by the global comparison but the two proteins were not clustered together indicating different expression patterns (**Figure 7-3d**). Although rtg428 seems conditioned by the type of glycerol, rtg4245 (IMA1) is predominantly expressed in presence of nutrients at low glycerol concentration. scIMA1 is regulated by the scMALx3 activator that promotes its upregulation in presence of maltose, isomaltose, and α -methylglucopyranoside (Teste et al., 2010). Indirect evidence of scIMA1 regulation by scDAL82 regulator of allophanate inducible genes like scDUR3 (Reimand et al., 2010;

Yoon & Kim, 2013) and during heat shock (Venters et al., 2011) were also described. So rtg4245 (IMA1) upregulation could be related to signals other than sugar presence. The presence of sugar was experimentally confirmed by a phenol-sulfuric assay. These sugars could be of biotic origin which would also explain the nitrogen content and support the thiamine presence hypothesis.

Sugars contained in sweetwater might be directly channeled into glycolysis or enter the pentose phosphate pathway (PPP) depending on their nature (**Figure 7-5**). The PPP seems particularly active in SW15 as suggested by the upregulation of the xylulose-5-phosphate/fructose-6-phosphate phosphoketolase [EC:4.1.2.9 4.1.2.22] (rtg5013). Furthermore, out of the 7 enzymes involved in the pentose phosphate pathway that are significantly differentially expressed only the sedoheptulose-biphosphatase [EC:3.1.3.37] (rtg3728) is less expressed in SW15 than in Gly15 which is consistent with a sedoheptulose-7-phosphate generation through the action of transketolases rather than from sedoheptulose-1,7-phosphate (**Appendix 11**). In addition to glyceraldehyde-3-phosphate (GA3P) generation, the oxidative pentose phosphate pathway has been described to provide NADPH to support fatty acid synthesis in *Y. lipolytica* (Wasylenko et al., 2015). The triose phosphate (GA3P) can be isomerized to DHAP and participate in energy production through glycolysis or DHAP can be converted to glycerol-3-phosphate (G3P) an essential precursor of glycerolipids synthesis taking place in the endoplasmic reticulum (ER) (Qin et al., 2017) (**Figure 7-5**).

Nitrogen-limitation induces a citrate accumulation due to a less active Krebs cycle. The excess of citrate is then channeled into a series of reactions leading to fatty acid synthesis. The first of these reactions is the acetyl-CoA production from citrate through the action of the ACL. A reaction followed by the acetyl-CoA conversion into malonyl-CoA catalyzed by the acetyl-CoA carboxylase (ACC). ACL and ACC are upregulated in

SW15 by 3.5-fold and 2.5-fold (respectively) along with the two fatty acid synthases (FAS) subunits (α : rtg127; β : rtg204) (**Figure 7-5**) and 9 other enzymes involved in glycerolipids, phospholipids and carotenoid production (**Appendix 11**). The upregulation of these enzymes is consistent with an improved lipid accumulation in SW15. Surprisingly, nitrogen-induced lipid accumulation is not also observed in Gly15. A possible explanation lies in the fact that fatty acid degradation through β -oxidation seems more active in Gly15 than in SW15. The β -oxidation can release peroxisomal acetyl-CoA that will then feed the glyoxylate cycle and possibly shunt the ICDH-induced downturn of Krebs cycle preventing thus citrate re-direction towards lipid production. Hence, when cells are grown in refined glycerol Gly15, they redirect acetyl-CoA flux towards TCA cycle for energy production. The downregulation in SW15 of β -oxidation related enzymes like the acyl-CoA dehydrogenase (rtg5452) & the carnitine *o*-acyltransferase (rtg 4139) and the malate synthase (rtg2130) (a key enzyme of the glyoxylate cycle) is supporting the latter hypothesis.

In terms of signaling, the strongest difference between SW15 and Gly15 comes from rtg3506 the activator of the protein phosphatase 2A (PP2A) that is 16.5 times more expressed in SW15. rtg3506 belongs to the phosphotyrosyl phosphatase activator (PTPA) family than can promote PP2A activity. PP2A have been described to be a part yeast nitrogen sensing system in conjunction with TORC pathway but can also modulate other cellular processes (Ariño et al., 2019). TORC pathway might be responsible for the upregulation of the genes involved in nitrogen recovery in response to limited nitrogen availability in both SW15 and Gly15. The TORC-mediated activation of β -oxidation is another well-documented consequence of nitrogen limitation (Singh et al., 2016) even described in *R. toruloides* (Zhu et al., 2012). We have demonstrated that even if β -oxidation is active in Gly15 it is downregulated in SW15. A difference that might contribute to a better lipid accumulation in SW15.

In both SW15 and Gly15, the low nitrogen levels should inhibit the ICDH of the TCA cycle. In Gly15, the active β -oxidation will feed the glyoxylate shunt that can provide the missing intermediaries to the TCA cycle preventing thus citrate accumulation. When in SW15, the β -oxidation cannot compensate the ICDH inhibition which might lead to citrate accumulation and redirection towards fatty acid synthesis (**Figure 7-5**). So, both β -oxidation and glyoxylate activities differ between SW15 and Gly15. The difference in glyoxylate cycle activity might be linked to the presence of sugars in SW15. Indeed, enzymes of the glyoxylate cycle and especially the malate synthase are subjected to carbon catabolite repression so the maintenance of a high C:N ratio, ensures glyoxylate cycle inhibition if the carbon source used can elicit a catabolite repression. In *R. toruloides*, an activation of the glyoxylate cycle was described during the diauxic shift from glucose to glycerol (Bommareddy et al., 2017) confirming that glucose exhaustion derepresses the malate synthase in the species. The result could have been different for a species favoring glycerol over glucose like *Y. lipolytica* (Lubuta et al., 2019). And the difference in β -oxidation could be imputed sugar presence. Indeed, the presence of sugar might indeed also block the utilization of exogenous fatty acids from SW15 forcing this exogenous fatty matter to join the cellular pool of fatty acids. In *S. cerevisiae*, this repression would be mediated by scADR1, a transcription factor involved in the regulation of glucose repressed genes including genes needed for ethanol, glycerol, and fatty acid utilization (Denis & Young, 1983; Simon et al., 1991; Tachibana et al., 2005; Young et al., 2003). To the best of my knowledge, this path was never described in *R. toruloides* and no homolog to scADR1 was found in *R. toruloides* (12) genome suggesting that another regulator might be involved.

Besides TORC pathway and carbon catabolite responses, the high-osmolarity glycerol (HOG) pathway might account for some of the differences observed between SW15 and

Gly15 as suggested by the upregulation of *rtg4017* (RCK2). In *S. cerevisiae*, *scRCK2* is a known target of *scHOG1* and is involved in the response to oxidative and osmotic stress (Bilsland-Marchesan et al., 2000). And *rtg4275* (RDS1) another protein involved in stress response is also upregulated (Ludin et al., 1995). Actually, *rtg4017* and *rtg4275* expression levels are (respectively 2.8 and 3.2 times) higher in SW15 compared to the other 3 conditions which is consistent with the lower biomass production in SW15 compared to Gly/SW1.5SC. As for the proliferation differences between the two conditions, it seems that SW15 promotes more adaptive mechanisms than Gly15. The lack of response to stress in Gly15 might be a sign that the absence of nutrients in Gly15 is such that the cells reached a dormant/quiescent state. Finally, the expression of the heat shock transcription factor *rtg4659* (HSF1) a general stress effector in Gly15 is close to the one achieved in Gly1.5SC. Similarly, the gene expression is not significantly different between SW15 and SW1.5SC. And *rtg4659* expression in sweetwater exceeds the expression in refined glycerol. Suggesting that sweetwater can also trigger stress responses non-nutrient related and that the element(s) promoting HSF1 expression in sweetwater can do so even when the waste is diluted. The differences in growth observed between SW1.5SC and Gly1.5SC might originate from this non-nutrient related stress (**Figure 7-2**).

R. toruloides grown in glycerol has been described to exhibit a lower growth than when it utilizes glucose but accumulates more lipids in glycerol than in glucoses (Bommareddy et al., 2017). Similarly, our strain exhibited an improved lipid production in SW15 associated with a lower growth than in SW1.5SC. This finding is consistent with previous reports showing that cell proliferation is affected by nitrogen limitation sometimes to a point where cells enter a quiescence phase (Klosinska et al., 2011) but nitrogen-limitation can also be beneficial to lipid accumulation in certain species (Dissook et al., 2020; Sitepu et al., 2014a). Differences in sources preferences or in catabolite repression pathways will

result in differences in lipid accumulation explaining the range of lipid contents observed across species.

7.4 - Concluding remarks

The rise of biodiesel has flooded the market with considerable volumes of inexpensive crude glycerol. This drop in crude glycerol value opened up opportunities for new usages including uses as feedstock for biomanufacturing. As the methanol content of biodiesel-derived crude glycerol tends to limit its microbial utilization, other types of crude glycerol could be investigated such as the methanol-free crude glycerol from fat splitting. This waste, also known as sweetwater (SW), albeit more compatible with microbial growth has been overlooked as candidate feedstock for biomanufacturing despite its very low value. Yeasts are very versatile cell factories capable of turning low-value feedstock into valuable chemicals including oleochemicals. Some yeasts are even able to accumulate up to 70% of their dry cell weight as lipid in certain conditions. These lipid-accumulating yeasts are called oleaginous yeasts and are not all close relatives explaining their different nutritional needs.

Here was studied sweetwater ability to sustain growth and lipid production for a panel of oleaginous yeasts. Higher OD₆₀₀ were reached when the waste was used diluted and supplemented with nutrients (SW1.5SC) than in its crude form (SW15). However, better lipid accumulation in SW15 than in SW1.5SC was sometimes observed despite the growth inhibition observed at high glycerol concentration and nutrient-limiting conditions typically being less favorable for growth. *R. toruloides* (12) is one of the strains affected by the SW-induced lipid accumulation. As the strain also turns out to be the one performing best in SW15, its transcriptomic response to the waste was investigated.

The analysis suggested that in *R. toruloides* (12), the nitrogen and carbon content of crude sweetwater (SW15) are promoting a better lipid accumulation than in refined glycerol (Gly15) through a better precursor supply. Indeed, the use of SW15 seems to improve glycerol-3-phosphate (glycolipid precursor), acetyl-CoA (fatty acid precursor), exogenous fatty acids and possibly NADPH (reducing agent in fatty acid synthesis) supply for lipid biosynthesis. From a mechanistic perspective, nitrogen-limitation holds a central role in sweetwater induced-lipid production along with a putative internalization of exogenous fatty acids from SW15 and a inhibition of the β -oxidation likely triggered by a carbon catabolite repression. This type of response reminded the behavior of another strain of this species in presence of glycerol/glucose mixtures described elsewhere (Bommareddy et al., 2017) so sugars presence in SW15 was tested and confirmed. Yet, the nature of the sugars remains to be further characterized. The transcriptomic response of *R. toruloides* (12) also suggested that crude sweetwater and the same concentration of refined glycerol differ in nitrogen and nutrient composition. These differences in composition could be imputed to the presence in SW15 of impurities from the crude oil used for the fat splitting process.

It was also found that SW15 seems to induce both nutrient-related and nutrient-unrelated stresses in *R. toruloides* (12). Both types of stresses might contribute to less efficient growth in SW15 compared to the one achieved when waste is used diluted and supplemented with nutrients (SW1.5SC). Since it has previously been demonstrated that nutrient-related limitations have an impact on lipid accumulation, identifying and alleviating the other sources of stress could improve cell proliferation in the waste. This finding should support future strain engineering endeavors and the design of processes based on sweetwater for sustainable yeast oil production.

When oleaginous yeasts are grown in sweetwater, a very delicate balance between growth and lipid accumulation seems to be at play mediated by metabolites fluxes that can

either be directed toward energy production or lipid production. An optimal balance between cell densities and lipid content is yet to be found but this study demonstrated sweetwater potential as feedstock for yeast oil production.

Part V – Conclusion and prospects

Chapter 8 - Conclusion

8.1 - Highlights

This work addressed two primary objectives: (1) understanding of how oleaginous yeasts - yeasts specialized in lipids accumulation - differ and compare while generating sufficient knowledge to feed GEMs reconstructions (2) confirming the suitability of sweetwater (fat-splitting crude glycerol) as feedstock for the lipid production of a panel oleaginous yeasts while gaining insights into the molecular basis of lipid accumulation from this crude glycerol. These two objectives are aligned with a larger will to solve the ongoing glycerol crisis (described in **Chapter 1**) through the development of new crude glycerol valorization routes that could fit into the circular economy model. The valorization route explored in this study is sweetwater utilization for yeast oil production. Not only yeast oil could be used to regenerate glycerol meeting thus the circular economy requirement, but the targeted oil could also replace or supplement current sources of oils and fats. Nowadays most oils and fats originate either from plants, animals, or petroleum. These sources of oils and fats are increasingly criticized (see **Chapter 1**) but remain more competitive than yeast oil (Abeln, 2021). A better understanding of lipid accumulation mechanisms from glycerol in oleaginous yeasts could help improve their performances and eventually the cost-efficiency of yeast oil production processes. About 5% of the 1600 known yeast species are oleaginous (Garay, 2016) and they are found all over yeast realms so multiple mechanisms might be at play. To avoid characterizing individually every oleaginous yeast, this study identified common and different features in a panel of 17 oleaginous yeast species. These differences were phenotypically evaluated and when possible juxtaposed with genomic data. A combined phenotypic and genomic comparison was chosen here because phenotypes vary with culture conditions when genomic data are supposedly more stable. Yet, given the current level of characterization of unconventional yeast genomes, genomic data could not account for all the differences

observed across strains. Hence, when investigating the molecular basis of lipid accumulation from sweetwater, a transcriptional dimension was added to the analysis.

With the rapid expansion of sequencing technologies, comparative genomic analysis has become more approachable (Mullikin, 2014) and could play a more important role in the exploration of yeast diversity for biomanufacturing. Comparative genomics could unravel patterns that are difficult to identify at the phenotypic level. In this study, comparative genomics was applied to the comparison of yeasts species with good potential for oleochemicals production. Genomes of the investigated species were all functionally annotated following the workflow described in **Chapter 4**. This first step revealed a huge knowledge gap as despite my best effort only 63.8 % of the considered pan-genome were successfully associated with a function. The genes were then sorted by homology into orthologous clusters which allowed the isolation of species-specific genes (unique genes). In **Chapter 6**, the pan-genomic analysis proved to be especially useful for the identification of singular/conserved genomics features and the investigation of the phylogenetic conservation of genes potentially involved in industrially relevant traits. No strong evidence of conservation at genus level was found so two strains of the same genus are not guaranteed to perform equally well in industrial setups. Instead of relying on phylogenetics for guidance when selecting a strain, one should investigate the origin of the strain since environmental adaptation might be responsible for most of the traits industrially appreciated.

After demonstrating in **Chapter 1** the potential of crude glycerol as feedstock for biomanufacturing. Sweetwater - an underrated crude glycerol - was investigated as feedstock for yeast lipid production in **Chapter 7**. The waste was found suitable for yeast growth even in its crudest form but disparities were identified across strains. Interestingly, strains able to achieve high cell densities were not necessarily the best lipid

accumulators and reciprocally. As highlighted in **Chapter 1**, strain selection is a major stake in the development of any processes based on microorganisms. So, considering both growth and lipid production, the strain *R. toruloides* (12) emerged as the host most compatible with sweetwater utilization for lipid production. Furthermore, *R. toruloides* (12) also exhibited good robustness during the evaluation of industrially relevant traits in **Chapter 6**. Results suggest that the unique composition of sweetwater can promote lipid production in some strains [including *R. toruloides* (12)]. Hence, the waste effects were transcriptionally investigated to decipher its molecular impact on the strain lipid production. The transcriptional analysis revealed that the key component affecting *R. toruloides* (12) lipid production is the presence of sugars in sweetwater. These sugars remain to be characterized but they trigger a carbon catabolite repression leading to a chain of metabolic adjustments resulting in an improved lipid accumulation when the strain is grown sweetwater. In addition to sugars, the waste also seems to contain nutrients able to support both growth and lipid accumulation. The transcriptomic response observed in presence of sweetwater resembles the response described elsewhere when another *R. toruloides* strain was grown in glycerol/glucose mixtures (Bommareddy et al., 2017). Even though genetic features explaining part of the lipid accumulation differences (i.e. ACL presence) are identifiable, some of the observed differences might arise from the fact that different species might respond differently to the glycerol/sugar mixtures due to different carbon source preferences.

This work contributed to the expansion of metabolic understanding of sweetwater utilization by delivering a transcriptomic data set that was used for the exploration of lipid metabolism but could also provide insights into the metabolism of other interesting metabolites. **Chapter 7** also included a phenotypic characterization performed on a panel of yeasts which increases the interconnectivity of my work to the pool of knowledge already existing. It can be noted that even though growth-related data might be valid

for any other natural product of these strains, the investigation of product accumulation from sweetwater will require a novel experimental dataset for every type of product investigated when the same transcriptomic dataset could be exploited for the exploration of the effect of sweetwater on the expression of genes involved in the synthesis of any other natural interesting product. This illustrates the importance of combined phenotypic::(gen/transcript)omics analysis. Similarly, the functional annotations generated in **Chapter 4** can serve much more purposes than the one described in this manuscript and should increase the visibility of the used genomic datasets. Finally, all these findings should pave the way toward more integrative methods for unconventional yeasts.

8.2 - Perspectives and future work

8.2.1 - Methodologies for unconventional yeasts (Part II)

This work focused on functional annotation and Nile Red assay when so many unconventional yeasts are in need of suitable genetic engineering toolkits as discussed in **Chapter 1**. But the work described in part II was essential to set this project on the best trajectory. Because inaccurate functional annotations will limit any genomic or transcriptomic-based explorations of unconventional yeasts when these are precisely the key to attaining quality by design for yeast oleochemicals production. Like often when it comes to microorganisms, the medical interest prompted the development of microbial genome-wide association studies to fill the gap between genotype and phenotype on aspects such as pathogenicity or drug resistance (San et al., 2020). It is likely that similar approaches will soon spread to biotechnological applications. And if both phenotypic data and genomic information are needed to create a solid bridge. The difference between **Chapter 4** and **Chapter 5** is a perfect illustration that both disciplines are at very different stages of their development. Where only investigations of the last ambiguities were required for the Nile red method, the current state of yeast functional annotation

called for a whole pipeline development. The biochemical characterization had centuries to mature and has now reached a point where although the data are still missing their acquisition is no longer as limited by technical considerations as it once was. That is to say that the genomic part is the limiting factor in filling the phenotype/genotype gap. To continue this parallel between **Chapter 4** and **Chapter 5**, the Nile red method was used in **Chapter 6** and **Chapter 7** as developed in **Chapter 5** when the functional annotations and pipeline resulting from **Chapter 3** were enriched and curated by the work described in **Chapter 6** and **Chapter 7**. This highlights an essential characteristic of functional annotations: functional annotations and tools for their manipulation and analysis should be flexible entities. The lack of dynamic structures, able to host such flexibility while offering the same traceability as the GitHub platform offers to millions of developers worldwide, prevents genomics from reaching its full potential. Mention of the GitHub platform is only made to illustrate that the traceability and flexibility needed are currently technically accessible. Since there is a growing awareness of that kind of considerations (Marschall et al., 2018), a complexification of the functional genomic toolkit is foreseeable. We can also expect an off-track development similar to the one of Funannotate⁸ - the most complete fungal annotation pipeline - that has been downloaded 7.7k times since its first release in 2016 but is still not described in any publications. And it is understandable considering that in 5 years, the pipeline benefitted from 74 upgrades. As complete as Funannotate is, it does not include an integrated data management system that would enable the simultaneous update of databases, annotations, and analysis results (**Figure 8-1**). Three entities that are for now completely disconnected when the whole process could benefit from an improved interconnection.

⁸ <https://github.com/nextgenusfs/funannotate>

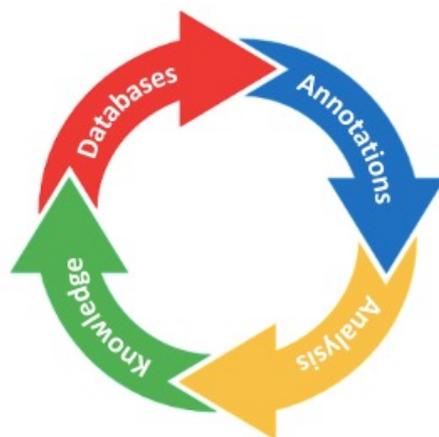


Figure 8-1 Anticipated future of functional annotations.

8.2.2 - New insights into oleaginous yeast metabolism (Part III)

The same way pan-genome might become the new reference genomes (Marschall et al., 2018) it is time to move away from a *Saccharomyces*-centred knowledge to a knowledge more representative of yeast actual diversity. With the huge amount of phenotypic data available in the literature and the increasing availability of omics data it is time for dots to be connected and gaps to be filled so we can one combined study at the time reach another level of understanding of cellular processes and increase the accuracy of GEMs and other predictive tools. In this knowledge expansion considering several species would be preferable to increase the work interconnectivity and because it might facilitate the detection of anomalies and set knowledge acquisition on a continuous quality control track.

The development of microbial oil technology can be decomposed in 3 stages (Ratledge, 2013; Sitepu et al., 2014a; Yan & Pflieger, 2020): (1) a phase of pathway analysis and characterization spanning from 1950 to 2011 (2) a phase of demonstration of microbial oil potential (i.e. proof of concept) spanning from 1989 to 2012 (3) and in the past decade, we have observed an acceleration of performance enhancement and industrialization. This evolution can be correlated with the development of biology that transitioned from classic biology to molecular biology before entering the era of genomics

and synthetic & system biology (Lachance et al., 2019). In the past twenty years, tools that can support the development of biology as an engineering discipline and fundamental science have emerged (Lachance et al., 2019). Taking advantage of these new methods to investigate oleaginous species might seem like a step back since we will be re-entering a characterization phase, but it should be regarded as momentum gathering. Especially since this knowledge acquisition will be a synonym for performance enhancement and hopefully industrialization. Sure, the current state of these tools and methods still suffers from the number of limitations (some highlighted in this manuscript). But since practice makes perfect, it is only by multiplying the type of approaches described in this manuscript that we will reach the full potential of system biology and omics technologies.

8.2.3 - Toward an enhanced utilization of sweetwater (Part IV)

The transcriptomic dataset will be used to support the reconstruction of condition-specific models to predict genetic engineering targets specific to the improvement of sweetwater utilization that respects the balance between growth and product accumulation. The current information might not be enough to properly model sweetwater composition so an analysis of sweetwater composition could be required and will only make the dataset stronger and improve its interconnectivity and reusability.

Albeit not mentioned in **Chapitre 1**, co-product valorization is also a possible strategy to improve the profitability of microbial oils (Chen et al., 2021). Using conventional rational design approaches to identify genetic engineering targets respecting the balance between growth and product accumulation is already quite challenging (**section 1.3.4.a**). It is very likely that the efficient optimization of the accumulation of 2 products will only be achieved with the assistance of computational methods. *R. toruloides* GEM could first be used to predict the best product/co-products combinations accessible from glycerol before starting working toward a process. Since *R. toruloides* has a quite

developed genetic engineering toolkit, metabolic engineering would be an option in the early stages of the development. But these means should be used wisely to attain pre-defined target productivities.

Unfortunately, access to target productivities is difficult and too often engineering attempts are disconnected from the industrial reality. A way to bypass that difficulty would be to pick a co-product among the top value-added chemicals from biomass reported by the U.S. Department of Energy Office of Scientific and Technical Information (Holladay et al., 2007; Werpy & Petersen, 2004). The two reports give valuable insights into the applications and target productivities for candidate products and remain up to this day references (the first report was cited more than 2100 times) when it comes to value-added products accessible from biomanufacturing. But these reports focus on replacement of current common chemical building blocks (small molecules) and neglect more complex structures. Hopefully, more studies considering the incredible natural metabolite diversity of cell factories will be released in the coming years.

The mechanistic understanding of how crude glycerol(s) composition affects microbial hosts will help design processes able to accommodate compositional variability of the waste. Once this remaining bottleneck is addressed, the identification of the right opportunity (i.e. timing and product(s)) will be the key to the establishment of bioprocesses based on crude glycerol, as discussed in **Chapter 1**. Comparing feedstocks used to produce bioethanol (the traditional emblematic biomanufacturing product) and biodiesel ⁹ (the recent development) gives hope that an increasing proportion of industrial processes will incorporate the use of wastes as raw material. Indeed only 2% of waste feedstocks (molasses) are used for bioethanol production while 20% of biodiesel

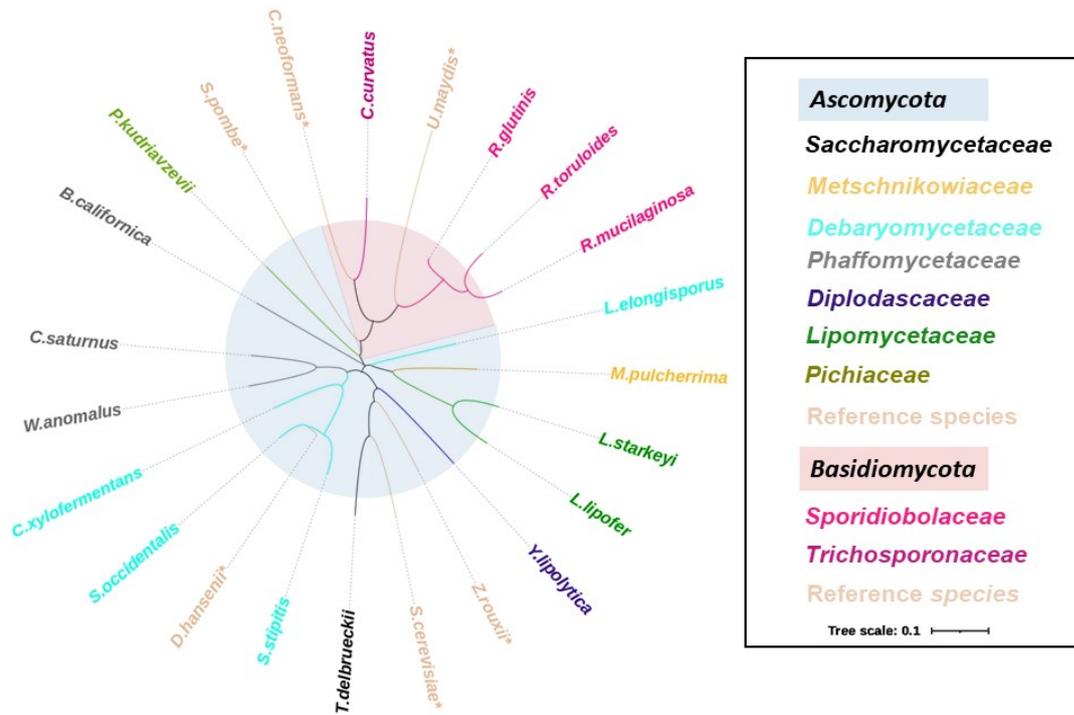
⁹ In 2021, biodiesel production remains mainly a chemical process.

production originates from used cooking oils (OECD-FAO, 2020). As the development of biofuels was fostered by policies and economic incentives, the current quest for sustainability should open new opportunities for yeast oleochemical production from crude glycerol, a convenient and economically attractive feedstock.

8.3 - Concluding thoughts

The upcoming complexification of genomics will widen the gap between computational biologists and experimentalists when their interdependency couldn't be higher. I humbly hope my work could be a reminder of the power of cooperation. A cooperation needed not only between experimentalists and bioinformaticians but also between academia and industry because it is only together that will prevail over the challenges of the century.

Appendices



Appendix 1 Phylogenetic tree showing the relationship between studied yeasts and reference yeasts.

Names followed by an asterisk correspond to reference species not included in this study. A phylogenetic tree was reconstructed using the 5.8S rRNA sequences of the 17 studied species and 5 reference species. Sequences were downloaded from NCBI are accessible using the following accession iD: MH595307.1 (bca), JN099268.1 (cxy), MK973015.1 (ccu), KY037834.1 (csa), KY104044.1 (lli), AB614134.1 (lst), MH059652.1 (lel), KY104206.1 (mpu), MT539201.1 (pkr), MT635318.1 (rgl), LT604871.1 (rmu), LC506275.1 (rto), LC438666.1 (sst), AJ229075.1 (tde), MT573404.1 (wan), KM115153.1 (yli), HQ026732.1 (soc). The phylogeny was then computed using EMBL-EBI simple phylogeny tool (Goujon et al., 2010; Larkin et al., 2007) that aligns the sequences using CLUSTALW (Larkin et al., 2007) and performs a clustering using the neighbor-joining algorithm. The tree was the formatted using iTOL (Letunic and Bork, 2019).

Appendix 2 Crude glycerol compositions reported in the literature.

Raw material	Catalyst (w/w)	Chemical reaction	pH	Glycerol (w/w)	MeOH (w/w)	Soap (w/w)	MONG (w/w)	Ash (w/w)	Moisture (w/w)	Reference
Soybean oil	KOH	Transesterification	n.r.	62.0%	12.8%	25.2%	n.r.	n.r.	n.r.	(Pyle <i>et al.</i> , 2008)
Chicken fat/soy bean oil (50/50, w/w)	KOH	Transesterification	n.r.	62.3%	14.4%	23.2%	n.r.	n.r.	n.r.	(Pyle <i>et al.</i> , 2008)
Canola oil	NaOH	Transesterification	n.r.	56.5%	28.3%	15.3%	n.r.	n.r.	n.r.	(Pyle <i>et al.</i> , 2008)
Sunflower oil	0.5% NaOH	Transesterification	n.r.	75.0%	1.0%	9.3%	n.r.	n.r.	n.r.	(Rehman <i>et al.</i> , 2008)
Sunflower oil	0.5% NaOH	Transesterification	n.r.	90.0%	<1.0%	n.r.	n.r.	n.r.	n.r.	(Rehman <i>et al.</i> , 2008)
Jatropha oil	n.r.	Transesterification	n.r.	18.0-22.0%	14.5%	29.0%	11.0-21.0%	n.r.	n.r.	(Hiremath <i>et al.</i> , 2011)
Palm oil	n.r.	Transesterification	n.r.	80.5%	0.5%	n.r.	<2.0%	n.r.	n.r.	(Liu <i>et al.</i> , 2013)
IdaGold oil	NaOCH ₃	Transesterification	n.r.	62.9%	n.r.	n.r.	n.r.	n.r.	n.r.	(Thompson and He, 2006)
PacGold oil	NaOCH ₃	Transesterification	n.r.	62.9%	n.r.	n.r.	n.r.	n.r.	n.r.	(Thompson and He, 2006)
Rapeseed oil	NaOCH ₃	Transesterification	n.r.	65.7%	n.r.	n.r.	n.r.	n.r.	n.r.	(Thompson and He, 2006)
Canola oil	NaOCH ₃	Transesterification	n.r.	67.8%	n.r.	n.r.	n.r.	n.r.	n.r.	(Thompson and He, 2006)
Soybean oil	NaOCH ₃	Transesterification	n.r.	67.8%	n.r.	n.r.	n.r.	n.r.	n.r.	(Thompson and He, 2006)
Crambe oil	NaOCH ₃	Transesterification	n.r.	62.5%	n.r.	n.r.	n.r.	n.r.	n.r.	(Thompson and He, 2006)
Waste vegetable oil	NaOCH ₃	Transesterification	n.r.	76.6%	n.r.	n.r.	n.r.	n.r.	n.r.	(Thompson and He, 2006)
Waste vegetable oil	n.r.	Transesterification	n.r.	63.4%	4.4%	n.r.	9.6%	7.4%	13.9%	(Chanjula <i>et al.</i> , 2016)

Raw material	Catalyst (w/w)	Chemical reaction	pH	Glycerol (w/w)	MeOH (w/w)	Soap (w/w)	MONG (w/w)	Ash (w/w)	Moisture (w/w)	Reference
Soybean oil	n.r.	Transesterification	n.r.	87.0%	0.0%	n.r.	5.7%	3.2%	9.2%	(Lammers <i>et al.</i> , 2008)
Vegetable oil	n.r.	Transesterification	n.r.	87.4%	0.1%	n.r.	n.r.	5.9%	8.0%	(Orengo <i>et al.</i> , 2014)
Palm oil	n.r.	Transesterification	n.r.	88.9%	0.5%	n.r.	0.0%	3.5%	5.6%	(Chanjula <i>et al.</i> , 2016)
Rapeseed oil	n.r.	Transesterification	n.r.	80.0%	0.5%	n.r.	n.r.	8.0%	n.r.	(Bartoň <i>et al.</i> , 2013)
Soybean oil	n.r.	Transesterification	n.r.	80.3%	n.r.	n.r.	1.6%	5.0%	12.0%	(Lage <i>et al.</i> , 2014a)
Soybean oil	n.r.	Transesterification	n.r.	80.3%	0.4%	n.r.	n.r.	n.r.	12.4%	(Shin <i>et al.</i> , 2012)
Pork fat	n.r.	Transesterification	n.r.	74.7%	0.0%	n.r.	0.9%	n.r.	10.3%	(Silveira <i>et al.</i> , 2015)
Castor/soybean/sunflower/cotton seed oils	n.r.	Transesterification	n.r.	36.2%	8.7%	n.r.	0.4%	2.0%	6.2%	(Lage <i>et al.</i> , 2014b)
Soybean/sunflower oils	n.r.	Transesterification	n.r.	83.0%	0.0%	n.r.	n.r.	6.0%	11.0%	(Carvalho <i>et al.</i> , 2015)
n.r.	n.r.	Transesterification	n.r.	82.6%	0.4%	n.r.	0.6%	9.3%	7.1%	(Kass <i>et al.</i> , 2013)
Soybean oil	n.r.	Transesterification	n.r.	80.3%	0.0%	n.r.	3.0%	5.7%	12.0%	(San Vito <i>et al.</i> , 2015)
n.r.	n.r.	Transesterification	n.r.	86.6%	0.0%	n.r.	n.r.	5.9%	7.5%	(Egea <i>et al.</i> , 2016)
n.r.	n.r.	Transesterification	n.r.	86.6%	0.0%	n.r.	0.5%	3.2%	9.2%	(Lammers <i>et al.</i> , 2008)
n.r.	n.r.	Transesterification	n.r.	84.4%	0.0%	n.r.	1.0%	3.2%	10.7%	(Kim <i>et al.</i> , 2013)
n.r.	n.r.	Transesterification	3.3	77.0%	0.0%	n.r.	4.6%	2.3%	16.1%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	5.4	94.8%	0.0%	n.r.	3.2%	0.0%	2.0%	(Hansen <i>et al.</i> , 2009)

Raw material	Catalyst (w/w)	Chemical reaction	pH	Glycerol (w/w)	MeOH (w/w)	Soap (w/w)	MONG (w/w)	Ash (w/w)	Moisture (w/w)	Reference
n.r.	n.r.	Transesterification	7.6	96.5%	0.0%	n.r.	1.0%	0.0%	1.3%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	9.0	38.4%	0.1%	n.r.	57.0%	4.2%	0.3%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	2.3	61.1%	0.2%	n.r.	5.6%	29.4%	2.5%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	10.6	66.7%	11.4%	n.r.	18.8%	2.9%	0.2%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	10.8	64.5%	13.9%	n.r.	18.1%	3.4%	0.0%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	2.7	83.4%	0.2%	n.r.	4.2%	1.5%	10.7%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	2.0	76.1%	1.8%	n.r.	6.9%	3.5%	11.7%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	2.4	74.5%	0.6%	n.r.	6.7%	4.6%	14.3%	(Hansen <i>et al.</i> , 2009)
n.r.	n.r.	Transesterification	8.6	63.4%	4.7%	n.r.	25.3%	5.6%	1.0%	(Hansen <i>et al.</i> , 2009)
Soybean oil	n.r.	Transesterification	6.9	63.0%	6.2%	0.0%	0.0%	2.7%	28.7%	(Hu <i>et al.</i> , 2012)
Soybean oil	n.r.	Transesterification	9.7	22.9%	10.9%	26.2%	23.5%	3.0%	18.2%	(Hu <i>et al.</i> , 2012)
Soybean oil	n.r.	Transesterification	9.5	33.3%	12.6%	26.1%	22.3%	2.8%	6.5%	(Hu <i>et al.</i> , 2012)
Waste vegetable oil	n.r.	Transesterification	9.4	27.8%	8.6%	20.5%	38.8%	2.7%	4.1%	(Hu <i>et al.</i> , 2012)
Soybean/waste vegetable oils	n.r.	Transesterification	10.0	57.1%	11.3%	31.4%	0.9%	5.7%	1.0%	(Hu <i>et al.</i> , 2012)
n.r.	n.r.	Transesterification	n.r.	75.0%	n.r.	n.r.	6.0%	10.0%	10.0%	(Tan <i>et al.</i> , 2013)
n.r.	n.r.	Hydrolysis	n.r.	88.0-90.0%	n.r.	n.r.	3.1-4.1%	0.7-1.0%	8.0-9.0%	(Tan <i>et al.</i> , 2013)
n.r.	n.r.	Saponification	n.r.	83.0-84.0%	n.r.	n.r.	0.9-1.2%	8.5-9.5%	6.0-7.0%	(Tan <i>et al.</i> , 2013)

MONG includes entrained fatty matters (*e.g.*, FFAs, FAMES, DAGs, MAGs and unprocessed oil) and other organic matters (*e.g.*, proteins), but excludes methanol and soap.

(a)

Query	BLAST KOALA		KOBAS sc		KOBAS kp		eggNOGmapper			OrthoMCL	
iD	Definition	KO	Gene ID Gene name Hyperlink	Gene ID Gene name Hyperlink	Definition	COG	KO	Orthogroups	Type		
bcg5222	(RefSeq) serine/threonine pr	K12766	sce:YMR139W RIM11, GSK3, MDS1 http://www	ppa:PAS_chr1-4_0368 http://www.gend	S_TKc	G	K03083	cluster2018	Core		

(b)

None

(blank)

COG S

(RefSeq) uncharacterized protein<

(RefSeq) hypothetical protein<

Inherit from ascNOG: Conserved hypothetical protein

(RefSeq) Cytoplasmic protein of unknown function<

VID27 cytoplasmic protein

(RefSeq) Putative protein of unknown function<

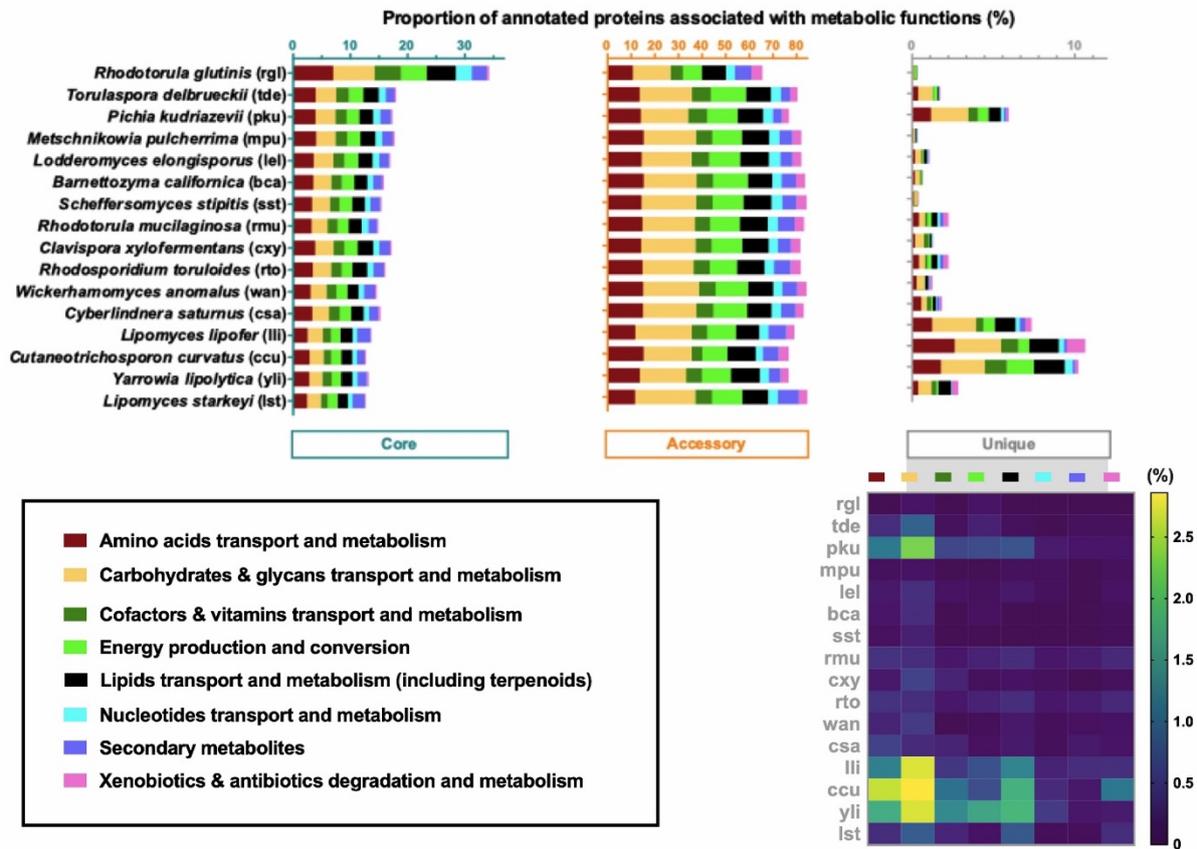
Appendix 3 Illustrations of Chapter 4

(a) Example of juxtaposed annotations as mentioned in Figure 4-1 Graphical illustration of the annotation pipeline principle Figure 4-1. Orthogroups is an abbreviation for Orthologous groups.

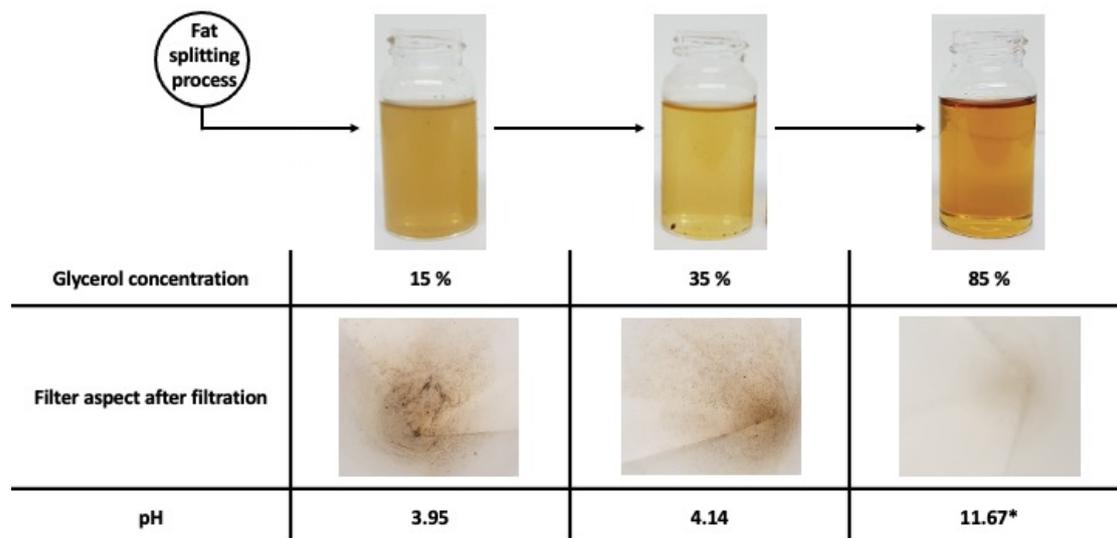
(b) Example of Unknown markers

Appendix 4 Contributions of the principal components to the PCA presented in Figure 6-4

Principal component	Contribution (%)	Cumulative contribution (%)
PC1	32.7800033	32.7800033
PC2	20.30755653	53.08755983
PC3	11.36748304	64.45504287
PC4	8.711225842	73.16626871
PC5	6.980696311	80.14696503
PC6	4.446474207	84.59343923
PC7	3.333052894	87.92649213
PC8	2.82587191	90.75236404
PC9	2.423871312	93.17623535
PC10	1.775821377	94.95205672
PC11	1.462624222	96.41468095
PC12	1.39231758	97.80699853
PC13	0.915780357	98.72277888
PC14	0.877602582	99.60038147
PC15	0.399618535	100



Appendix 5 Proportion of core, accessory, and unique genes per strain for genes associated with metabolic functions. Percentages given represent parts of the number of genes associated with a metabolic function. One gene can be associated with multiple metabolic sub-categories. Core genes are shared across all the studied strains. Accessory genes are shared across a few strains or specific to one strain (paralogs groups). Unique genes are specific to one strain. As the proportion of unique genes is very low for some strains, the data was also plotted as a heat-map to facilitate the analysis of the distribution of these genes among metabolic subcategories.



Appendix 6 Aspect and properties of the three types of sweetwater used in this study.

The asterisk indicates that SW85 was diluted before pH measurement. SW15 is released by the fat splitting process and usually concentrated using an evaporator to reach a glycerol concentration of 35 %. On this SW35, a fat flocculation is then performed to remove traces of fat using alum as coagulant and lime as flocculant. The addition of lime drastically increases the pH. The precipitates formed tend to settle down while sweeping out compounds in suspension. After the precipitate removal, the sweetwater is then concentrated in continuous multi-stage water evaporation to achieve glycerol concentration of 85% which correspond to SW85 often marketed as “semi-crude glycerol”. If higher purity is required sweetwater can then distilled to attain glycerol concentration of 95 and higher (corresponding to the different grades of refined glycerol not shown on the figure).

	Gly2SC vs. Gly8SC		Gly2SC vs. Gly16SC		Gly8SC vs. Gly16SC	
	tmid	cc	tmid	cc	tmid	cc
<i>B. californica</i> (1)	0.9992	0.0079	0.0111	0.0227	0.0099	0.9308
<i>C. xylofermentans</i> (2)	<0.0001	<0.0001	0.0051	<0.0001	0.1119	0.9571
<i>C. curvatus</i> (3)	0.7731	0.0116	0.0011	0.8461	0.0105	0.002
<i>C. saturnus</i> (4)	0.9174	<0.0001	0.0907	<0.0001	0.2027	<0.0001
<i>L. lipofer</i> (5)	<0.0001	0.0074	<0.0001	<0.0001	<0.0001	<0.0001
<i>L. starkeyi</i> (6)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>L. elongisporus</i> (7)	<0.0001	0.2166	0.0025	0.0248	<0.0001	<0.0001
<i>M. pulcherrima</i> (8)	0.5924	0.0002	0.7017	0.0009	0.9835	0.9062
<i>P. kudriavzevii</i> (9)	0.0077	<0.0001	<0.0001	<0.0001	<0.0001	0.7756
<i>R. glutinis</i> (10)	0.8231	0.0012	0.0013	<0.0001	0.0002	<0.0001
<i>R. mucilaginosa</i> (11)	0.6806	<0.0001	0.1579	<0.0001	0.5727	0.0693
<i>R. toruloides</i> (12)	>0.9999	<0.0001	<0.0001	<0.0001	<0.0001	>0.9999
<i>R. toruloides</i> (13)	0.827	<0.0001	<0.0001	<0.0001	<0.0001	0.0056
<i>S. stipitis</i> (14)	0.9907	<0.0001	<0.0001	0.0003	<0.0001	0.0326
<i>S. stipitis</i> (15)	0.392	<0.0001	0.392	0.0356	>0.9999	0.0002
<i>S. stipitis</i> (16)	0.0025	<0.0001	<0.0001	<0.0001	<0.0001	0.5349
<i>S. occidentalis</i> (17)	0.9976	<0.0001	<0.0001	<0.0001	<0.0001	0.9921
<i>T. delbrueckii</i> (18)	0.9945	<0.0001	<0.0001	<0.0001	<0.0001	0.8842
<i>W. anomalus</i> (19)	0.8889	<0.0001	0.5679	<0.0001	0.8449	0.707
<i>Y. lipolytica</i> (20)	0.8807	<0.0001	0.6694	<0.0001	0.3775	0.3034
<i>Y. lipolytica</i> (21)	0.9971	0.0001	0.9795	0.1259	0.9616	0.0601

Appendix 7 Results of ANOVA and Tukey's Post Hoc Multiple Comparisons for parameters described in Table 7.1. The table summarises p values probability of non-significant -ns - effects) for cc (carrying capacities) and tmid (times needed to reach half of the carrying capacities).

	SW_npH vs. SW_apH		SW_npH vs. SW35SC		SW_npH vs. SW1.5SC		SW_apH vs. SW35SC		SW_apH vs. SW1.5SC		SW35SC vs. SW1.5SC	
<i>B. californica</i> (1)	0.9695	0.9994	0.0488	<0.0001	0.7995	<0.0001	0.0983	<0.0001	0.9677	<0.0001	0.1832	0.0136
<i>C. xylofermentans</i> (2)	0.018	<0.0001	<0.0001	0.8613	0.0022	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>C. curvatus</i> (3)	<0.0001	0.5475	<0.0001	0.0021	<0.0001	0.0354	<0.0001	<0.0001	<0.0001	0.0004	<0.0001	0.8093
<i>C. saturnus</i> (4)	0.9208	0.87	<0.0001	0.0013	0.3795	<0.0001	<0.0001	0.0173	0.7681	<0.0001	0.0006	<0.0001
<i>L. lipofer</i> (5)	0.0002	<0.0001	0.1791	<0.0001	<0.0001	<0.0001	<0.0001	0.1946	<0.0001	0.0861	<0.0001	0.0001
<i>L. starkeyi</i> (6)	0.9879	0.002	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>L. elongisporus</i> (7)	0.9575	0.3953	0.964	0.0006	<0.0001	<0.0001	0.7587	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>M. pulcherrima</i> (8)	0.9993	0.998	0.9854	0.8519	<0.0001	<0.0001	0.996	0.9233	<0.0001	<0.0001	<0.0001	<0.0001
<i>P. kudriavzevii</i> (9)	<0.0001	0.0109	<0.0001	<0.0001	<0.0001	<0.0001	0.3342	<0.0001	0.0203	<0.0001	<0.0001	<0.0001
<i>R. glutinis</i> (10)	<0.0001	0.9006	<0.0001	0.5032	<0.0001	<0.0001	0.7881	0.895	<0.0001	<0.0001	<0.0001	<0.0001
<i>R. mucilaginosus</i> (11)	0.7446	0.3435	0.6386	<0.0001	0.9686	<0.0001	0.9981	<0.0001	0.9456	<0.0001	0.8861	<0.0001
<i>R. toruloides</i> (12)	0.2679	0.5528	0.4716	<0.0001	0.1573	<0.0001	0.9783	<0.0001	0.0002	<0.0001	0.0009	0.0035
<i>R. toruloides</i> (13)	<0.0001	0.9777	<0.0001	<0.0001	0.0003	<0.0001	0.8548	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>S. stipitis</i> (14)	0.191	<0.0001	0.9617	<0.0001	<0.0001	<0.0001	0.0644	0.0047	<0.0001	<0.0001	<0.0001	<0.0001
<i>S. stipitis</i> (15)	0.8526	0.0102	0.5185	<0.0001	0.0004	<0.0001	0.9416	0.0316	<0.0001	<0.0001	<0.0001	<0.0001
<i>S. stipitis</i> (16)	0.9989	0.2117	0.8548	0.0542	0.0008	<0.0001	0.7809	0.929	0.0004	<0.0001	0.0122	<0.0001
<i>S. occidentalis</i> (17)	<0.0001	<0.0001	<0.0001	0.0301	<0.0001	<0.0001	<0.0001	<0.0001	0.0165	<0.0001	0.4415	<0.0001
<i>T. delbrueckii</i> (18)	0.9989	<0.0001	0.7741	<0.0001	0.3156	<0.0001	0.688	<0.0001	0.3958	<0.0001	0.0394	<0.0001
<i>W. anomalus</i> (19)	0.9687	<0.0001	0.0639	<0.0001	<0.0001	0.0155	0.1778	0.7802	<0.0001	<0.0001	<0.0001	<0.0001
<i>Y. lipolytica</i> (20)	0.9936	0.0762	0.9997	0.4353	0.9954	<0.0001	0.9841	0.0006	0.9596	<0.0001	0.9991	<0.0001
<i>Y. lipolytica</i> (21)	0.0004	0.9937	0.9207	0.0882	0.9686	<0.0001	0.0006	0.1425	<0.0001	<0.0001	0.9936	<0.0001
	tmid	cc	tmid	cc	tmid	cc	tmid	cc	tmid	cc	tmid	cc

Appendix 8 Results of ANOVA and Tukey's Post Hoc Multiple Comparisons for parameters described in Table 7.2. The table summarises p values for cc (carrying capacities) and tmid (times needed to reach half of the carrying capacities).

(a)

Strain	SW15 (3 days)	SW15 (10 days)	SW1.5SC (3 days)	SW1.5SC (10 days)
<i>B. californica</i> (1)	376,4 ± 15,8	996,4 ± 15,6	1270,6 ± 44,2	330,8 ± 9
<i>C. xylofermentans</i> (2)	257,2 ± 3,3	450,4 ± 42,4	212,7 ± 3,9	261,4 ± 18,7
<i>C. curvatus</i> (3)	-	584,9 ± 9,9	1241,8 ± 44	2200,6 ± 103,3
<i>C. saturnus</i> (4)	752,6 ± 13,1	1955,8 ± 112,2	1104,7 ± 6,5	2417,6 ± 94,4
<i>L. lipofer</i> (5)	9,1 ± 2,7	9 ± 1,4	168,7 ± 8	2580,9 ± 47,4
<i>L. starkeyi</i> (6)	21,9 ± 1,4	3344,9 ± 48,1	971,2 ± 26,9	7115,5 ± 172,2
<i>L. elongisporus</i> (7)	406,4 ± 4,4	2040,4 ± 54,8	907,7 ± 92,8	2266,9 ± 29,3
<i>M. pulcherrima</i> (8)	699,5 ± 20,4	3020,8 ± 38,5	660,8 ± 48,9	420,7 ± 32,4
<i>P. kudriavzevii</i> (9)	60 ± 1,2	19,6 ± 0,4	961,9 ± 4,9	164,3 ± 4,5
<i>R. glutinis</i> (10)	-	2519,8 ± 47,5	1938,6 ± 29,1	1133,3 ± 28,4
<i>R. mucilaginosa</i> (11)	432,1 ± 27,8	1281,1 ± 19,1	351,2 ± 22,2	537,9 ± 6,4
<i>R. toruloides</i> (12)	823,4 ± 49,8	3522,1 ± 136,5	2011,2 ± 46,6	1953,4 ± 48,7
<i>R. toruloides</i> (13)	-	514 ± 12,2	398,2 ± 16,2	457,2 ± 17
<i>S. stipitis</i> (14)	1303,2 ± 29,7	1025,3 ± 72,2	881,4 ± 31,7	432,3 ± 2,8
<i>S. stipitis</i> (15)	793 ± 14,4	805,6 ± 39,5	921,3 ± 12,7	1543,1 ± 54,2
<i>S. stipitis</i> (16)	662,7 ± 7,2	968,6 ± 9,3	700,1 ± 4,1	1465,4 ± 29,2
<i>S. occidentalis</i> (17)	200,3 ± 1,2	682 ± 51,3	463,4 ± 10,1	1043,2 ± 11,3
<i>T. delbrueckii</i> (18)	70,6 ± 1,9	182 ± 4	117,2 ± 4,7	72,8 ± 2,1
<i>W. anomalus</i> (19)	462,2 ± 28,4	1145 ± 143	926,6 ± 22,4	1740,2 ± 30,3
<i>Y. lipolytica</i> (20)	398 ± 87,5	562,5 ± 12	1248,2 ± 40,8	1937,7 ± 23,7
<i>Y. lipolytica</i> (21)	331,7 ± 8,6	933 ± 50	682,8 ± 32,3	1375,3 ± 149,2

(b)

iD	SW15 (3 days) vs. SW15 (10 days)	SW15 (3 days) vs. SW1.5SC (3 days)	SW15 (3 days) vs. SW1.5SC (10 days)	SW15 (10 days) vs. SW1.5SC (3 days)	SW15 (10 days) vs. SW1.5SC (10 days)	SW1.5SC (3 days) vs. SW1.5SC (10 days)
1	<0.0001	<0.0001	0.5467	<0.0001	<0.0001	<0.0001
2	<0.0001	0.567	0.9993	<0.0001	<0.0001	0.4898
3	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
4	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
5	>0.9999	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
6	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
7	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
8	<0.0001	0.6738	<0.0001	<0.0001	<0.0001	<0.0001
9	0.6428	<0.0001	0.0141	<0.0001	0.0002	<0.0001
10	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
11	<0.0001	0.0889	0.0123	<0.0001	<0.0001	<0.0001
12	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.3354
13	<0.0001	<0.0001	<0.0001	0.0048	0.3511	0.3171
14	0.9831	0.0013	<0.0001	0.0048	<0.0001	<0.0001
15	<0.0001	0.697	<0.0001	<0.0001	<0.0001	<0.0001
16	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
17	0.0073	0.5282	>0.9999	0.2368	0.009	0.5689
18	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
19	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
20	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
21	<0.0001	<0.0001	0.5467	<0.0001	<0.0001	<0.0001

Appendix 9 (a) Raw data plotted in Figure 7-1a and associated standard deviations. (b) Results of the ANOVA and Tuckey's post hoc Multiple Comparisons. The table summarises p-values.

(a)



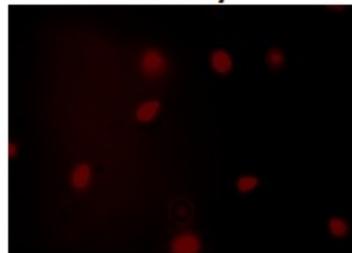
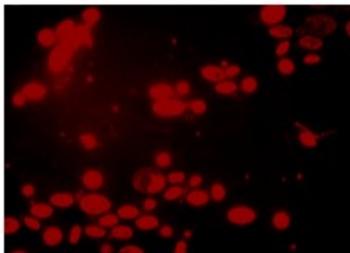
(b)

4th day

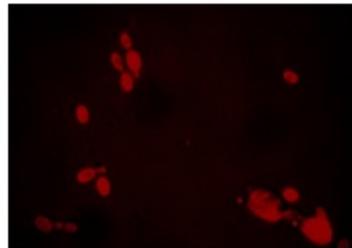
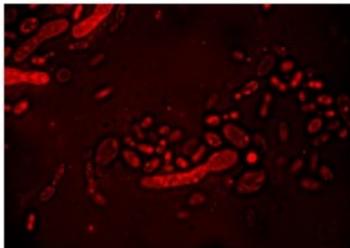
10th day

(c)

SW15



SW1.5SC



R. toruloides (12)

Appendix 10 Microscopic observation of Nile red stained *R. toruloides* (12) cells and macroscopic cell clump. (a) Bright field microscopy of *R. toruloides* (12) stained with Nile red (magnification x1000) and grown 4 days in SW15 (b) Fluorescence microscopy of the *R. toruloides* (12) stained with Nile red (magnification x1000) and grown 4 or 10 days in SW15 or SW1.5SC (magnification x1000) (c) 1 cm diameter clump of *R. toruloides* (12) cells.

Appendix 11 Annotation of genes represented in Figure 7-3

iD	KO	COG	Definition	Interclass
rtg2117	K01580	E	glutamate decarboxylase [EC:4.1.1.15]	Amino acids transport and metabolism
rtg4590	K08192	G	MFS transporter, ACS family, allantoin permease	Carbohydrates transport and metabolism
rtg428	K05349	G	beta-glucosidase [EC:3.2.1.21]	Carbohydrates transport and metabolism
rtg1145	-	G	Major Facilitator Superfamily	Carbohydrates transport and metabolism
rtg1762	K08139	G	MFS transporter, SP family, sugar:H ⁺ symporter	Carbohydrates transport and metabolism
rtg3791	K01183	G	chitinase [EC:3.2.1.14]	Carbohydrates transport and metabolism
rtg4244	K08141	G	MFS transporter, SP family, general alpha glucoside:H ⁺ symporter	Carbohydrates transport and metabolism
rtg4245	K01182	G	oligo-1,6-glucosidase [EC:3.2.1.10]	Carbohydrates transport and metabolism
rtg4615	K02429	G	MFS transporter, FHS family, L-fucose permease	Carbohydrates transport and metabolism
rtg4722	K08178	G	MFS transporter, SHS family, lactate transporter	Carbohydrates transport and metabolism
rtg5254	K08139	G	MFS transporter, SP family, sugar:H ⁺ symporter	Carbohydrates transport and metabolism
rtg1110	K02365	D	separase [EC:3.4.22.49]	Cell cycle control, cell division, chromosome partitioning
rtg5072	K11547	D	kinetochore protein NDC80	Cell cycle control, cell division, chromosome partitioning
rtg3275	K03843	M	alpha-1,3/alpha-1,6-mannosyltransferase [EC:2.4.1.132 2.4.1.257]	Cell wall/membrane/envelope biogenesis
rtg239	K11253	B	histone H3	Chromatin structure and dynamics
rtg698	K11251	B	histone H2A	Chromatin structure and dynamics
rtg3504	K11253	-	histone H3	Chromatin structure and dynamics Intracellular trafficking, secretion, and vesicular transport
rtg610	K20520	Z	drebrin-like protein	Cytoskeleton
rtg3437	-	V	Male sterility protein	Defense mechanisms
rtg2031	K00467	C	lactate 2-monooxygenase [EC:1.13.12.4]	Energy production and conversion
rtg4853	K00326;K17877	C	cytochrome-b5 reductase [EC:1.6.2.2] nitrite reductase (NAD(P)H) [EC:1.7.1.4]	Energy production and conversion
rtg5031	K00354	C	NADPH2 dehydrogenase [EC:1.6.99.1]	Energy production and conversion
rtg1482	K19792	P	low-affinity ferrous iron transport protein	Inorganic ion transport and metabolism
rtg2366	K03549;K17871	P	KUP system potassium uptake protein NADH:ubiquinone reductase (non-electrogenic) [EC:1.6.5.9]	Inorganic ion transport and metabolism

iD	KO	COG	Definition	Interclass
rtg4153	K20989	P	urea-proton symporter	Inorganic ion transport and metabolism
rtg4852	K02575	P	MFS transporter, NNP family, nitrate/nitrite transporter	Inorganic ion transport and metabolism
rtg6000	K03320	P	ammonium transporter, Amt family	Inorganic ion transport and metabolism
rtg3429	K18278	-	Thiamine precursor biosynthesis enzyme	Metabolism of cofactors and vitamins
rtg6058	K14709	P	solute carrier family 39 (zinc transporter), member 1/2/3	Inorganic ion transport and metabolism
rtg4154	K20989		urea-proton symporter	Intracellular trafficking, secretion, and vesicular transport
rtg3592	K23516	I	xanthine dioxygenase [EC:1.14.11.48]	Lipid transport and metabolism
rtg2138	-	I	MaoC like domain (peroxisomal dehydratase - ncbiconfrimed)	Lipid transport and metabolism
rtg5959	-	I	Acyl-CoA dehydrogenase, N-terminal domain	Lipid transport and metabolism
rtg4308	K01487	F	guanine deaminase [EC:3.5.4.3]	Nucleotide transport and metabolism
rtg4975	K06910		phosphatidylethanolamine-binding protein	Peptidases and inhibitors
rtg4874	K03564	O	peroxiredoxin Q/BCP [EC:1.11.1.15]	Post-translational modification and protein turnover
rtg5216	K08770	L	ubiquitin C	Replication, recombination and repair
rtg729	-	A	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	RNA processing and modification
rtg206	K13237	Q	peroxisomal 2,4-dienoyl-CoA reductase [EC:1.3.1.34]	Secondary metabolites metabolism and transport
rtg2788	K00002	Q	alcohol dehydrogenase (NADP+) [EC:1.1.1.2]	Secondary metabolites metabolism and transport
rtg4053	-	Q	Retinal pigment epithelial membrane protein	Secondary metabolites metabolism and transport
rtg4426	K16216	Q	benzil reductase ((S)-benzoin forming) [EC:1.1.1.320]	Secondary metabolites metabolism and transport
rtg6024	-	Q	2OG-Fe(II) oxygenase superfamily	Secondary metabolites metabolism and transport
rtg6040	-	Q	short chain dehydrogenase reductase family protein	Secondary metabolites metabolism and transport
rtg6041	-	Q	short chain dehydrogenase reductase family protein	Secondary metabolites metabolism and transport
rtg3996	K00121	Q	S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase [EC:1.1.1.284 1.1.1.1]	Secondary metabolites metabolism and transport
rtg3009	K13953	Q	alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]	Secondary metabolites metabolism and transport
rtg2375	-	Q	Animal haem peroxidase	Secondary metabolites metabolism and transport
rtg814	K06686;K08282	T	cell cycle protein kinase DBF20 [EC:2.7.11.-] non-specific serine/threonine protein kinase [EC:2.7.11.1]	Signal transduction mechanisms

Appendix 12 Top50 of annotated genes contributing the most to PC1 in the differential expression analysis of SW₁₅ vs Gly₁₅

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
Nitrogen scavenging/Nitrogen recycling						
rtg4852	10	-2.19279	<0.01	K02575	P	MFS transporter, NNP family, nitrate/nitrite transporter
rtg3792	33	2.216161	<0.01	K01426	E	Amidase [EC:3.5.1.4]
rtg3315	38	1.47717	<0.01		T	OPT oligopeptide transporter protein
rtg4853	44	-1.43723	<0.01	K00326;K17877	C	Cytochrome-b5 reductase [EC:1.6.2.2] nitrite reductase (NAD(P)H) [EC:1.7.1.4]
rtg5673	45	1.560374	<0.01	K03305	E	Proton-dependent oligopeptide transporter, POT family
Signaling						
rtg3506	7	4.04169	<0.01	K17605	D;T	Serine/threonine-protein phosphatase 2A activator
rtg4017	24	1.723819	<0.01	K23785	T	Serine/threonine-protein kinase RCK2 [EC:2.7.11.1]
rtg4275	39	1.50702	<0.01		S	Stress response protein Rds1
rtg4659	59	1.451221	<0.01	K09419	K	Heat shock transcription factor, other eukaryote
rtg4708	26	1.621092	<0.01		T	Cytoplasm protein (RGC2)
Membrane remodelling						
rtg3898	8	-5.3022	<0.01			Rare lipoprotein A (rlpa)-like double-psi beta-barrel/expansin
rtg6092	9	3.854252	<0.01		O	Fasciclin domain (involved in cell adhesion)/Beta-ig-h3 fasciclin
rtg2052	43	1.484737	<0.01	K12472	T;U	Epidermal growth factor receptor substrate 15
rtg1123	60	1.359099	<0.01	K20046	Z	Actin cytoskeleton-regulatory complex protein SLA1
Protein modification and turnover						
rtg5563	15	2.083641	<0.01	K01875	J	Seryl-trna synthetase [EC:6.1.1.11]
rtg3923	42	1.524287	<0.01	K03695	O	ATP-dependent Clp protease ATP-binding subunit clpb
rtg2372	63	1.347649	<0.01		O	ATP-dependent protease La (LON) domain
Energy and redox balance						
rtg5547	6	3.209785	<0.01	K00228	H	Coproporphyrinogen III oxidase [EC:1.3.3.3]
rtg5031	32	1.59539	<0.01	K00354	C	NADPH2 dehydrogenase [EC:1.6.99.1]
rtg2788	37	1.509143	<0.01	K00002	Q	Alcohol dehydrogenase (NADP+) [EC:1.1.1.2]
Carbon catabolite repression						
rtg4245	1	3.477264	<0.01	K01182	G	Oligo-1,6-glucosidase [EC:3.2.1.10]
rtg5013	13	1.949483	<0.01	K01621	G	Xylulose-5-phosphate/fructose-6-phosphate phosphoketolase [EC:4.1.2.9 4.1.2.22]
rtg769	14	1.93709	<0.01	K01193	G	Beta-fructofuranosidase [EC:3.2.1.26]

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
rtg5302	19	1.801282	<0.01	K08139	G	MFS transporter, SP family, sugar:H ⁺ symporter (HXT family)
rtg428	25	2.701482	<0.01	K05349	G	Beta-glucosidase [EC:3.2.1.21]
rtg2468	27	1.702245	<0.01			Pbh1 domain (= tertiary structures of pectate lyases and rhamnogalacturonase A)
rtg1823	29	1.788605	<0.01		E;G	Triose-phosphate Transporter family
rtg1162	34	1.699774	<0.01		G	Major facilitator superfamily
rtg3298	56	1.396897	<0.01		S	Glycosyl hydrolase family 88
Fatty acid uptake						
rtg2252	31	2.678024	<0.01			Lipase B precursor
rtg2309	46	1.418252	<0.01	K00507	I	Stearoyl-coa desaturase (Delta-9 desaturase) [EC:1.14.19.1]
Beta-oxidation and Glyoxylate cycle related						
rtg2130	3	-5.98348	n.a.	K01638	C	Malate synthase [EC:2.3.3.9]
rtg5452	41	-1.53047	<0.01		I	Acyl-coa dehydrogenase, C-terminal domain
rtg4139	64	-1.74603	<0.01	K00624	I	Carnitine O-acetyltransferase [EC:2.3.1.7]
FA synthesis						
rtg2312	17	1.797549	<0.01	K01648	C	ATP citrate (pro-S)-lyase [EC:2.3.3.8]
rtg127	50	1.355341	<0.01	K00667	I	Fatty acid synthase subunit alpha, fungi type [EC:2.3.1.86]
rtg226	57	1.329832	<0.01	K11262	I	Acetyl-coa carboxylase / biotin carboxylase 1 [EC:6.4.1.2 6.3.4.14 2.1.3.15]
rtg204	58	1.376327	<0.01	K00667;K00668	I	Fatty acid synthase subunit beta, fungi type [EC:2.3.1.86]
Unclassified						
rtg5628	2	2.609292	<0.01			Stigma-specific protein, Stig1
rtg3278	4	3.410903	<0.01		Q	ABC-2 type transporter
rtg4452	5	4.020687	<0.01		S	Inherit from basnog: delayed-type hypersensitivity antigen
rtg3026	21	-2.24292	<0.01		S	Protein similar to cwj C-terminus 2
rtg2410	28	2.059587	<0.01		G	Fungal trichothecene (sesquiterpenoids) efflux pump (MFS family)
rtg2948	30	2.616607	<0.01		S	Inherit from funog: CCCH zinc finger DNA binding protein
rtg2495	35	1.401479	ns		S	Predicted membrane protein (DUF2231)
rtg617	40	-2.3205	<0.01	K00622	Q	Arylamine N-acetyltransferase [EC:2.3.1.5]
rtg5922	48	-2.1786	<0.01		S	SCP (Cystein-rich secretory protein)
rtg3821	53	1.752634	<0.01		S	FAD dependent oxidoreductase
rtg3945	54	1.42675	<0.01		S	Zinc-binding dehydrogenase

iD	Rank #	LFC	p_{adj}	KO	COG	Definition
rtg5476	61	1.64211	<0.01		Q	ABC transporter transmembrane region

Gene ranks in terms of contribution to PC1 were given in the column Rank # and genes were sorted correspond sorted by descending contribution in each category. The smallest the rank is the highest the contribution will be. LFC (Log2 fold) gives the estimated change between gene expression level of in SW₁₅ compared to Gly₁₅ (Ctrl). This change is associated with a FDR adjusted p-value either non-significant (ns), non-computed (na), lower than 0.01 (<0.01), between 0.01 and 0.05 (<0.05) or between 0.05 and 0.1 (<0.1). Only padj <0.01 were considered significant in the main manuscript.

Appendix 13 Differential expression results of SW₁₅ vs Gly₁₅ of relevant genes associated to lipid metabolism

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
Acetyl-CoA metabolism						
Acetyl-CoA synthetase						
rtg931	1711	0.463515	<0.05	K01895	H	Acetyl-coa synthetase [EC:6.2.1.1]
ATP:Citrate lyase						
rtg2312	17	1.797549	<0.01	K01648	C	ATP citrate (pro-S)-lyase [EC:2.3.3.8]
Acetyl-CoA C-acetyltransferase (ACAT)						
rtg1529	3125	-0.22239	ns	K07513	I	Acetyl-coa acyltransferase 1 [EC:2.3.1.16]
rtg3157	4703	-0.11376	ns	K07513	I	Acetyl-coa acyltransferase 1 [EC:2.3.1.16]
rtg4298	1766	-0.36792	<0.01	K07513	I	Acetyl-coa acyltransferase 1 [EC:2.3.1.16]
Citrate synthase						
rtg1700	1590	0.529734	<0.05	K01647	C	Citrate synthase [EC:2.3.3.1]
rtg2745	3093	0.528872	ns	K01647		Citrate synthase [EC:2.3.3.1]
rtg2747	251	0.908961	<0.01	K01647	C	Citrate synthase [EC:2.3.3.1]
Malate dehydrogenases						
rtg1361	1394	0.429893	<0.01	K00026	C	Malate dehydrogenase [EC:1.1.1.37]
rtg1596	2036	0.43835	ns	K00027	C	Malate dehydrogenase (oxaloacetate-decarboxylating) [EC:1.1.1.38]
rtg3452	359	0.777535	<0.01	K00026	C	Malate dehydrogenase [EC:1.1.1.37]
rtg6104	649	-1.02196	<0.01	K00027;K00029	C	Malate dehydrogenase (oxaloacetate-decarboxylating) [EC:1.1.1.38] malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+) [EC:1.1.1.40]
Fumarase						
rtg5411	1021	0.559453	<0.01	K01679	C	Fumarate hydratase, class II [EC:4.2.1.2]
Succinate dH						
rtg1051	663	0.683249	<0.01	K18168	S	Succinate dehydrogenase assembly factor 2
rtg1256	286	0.828923	<0.01	K00234	C	Succinate dehydrogenase (ubiquinone) flavoprotein subunit [EC:1.3.5.1]
rtg1876	2552	-0.45795	ns	K18167	O	Succinate dehydrogenase assembly factor 1
rtg4479	1105	0.505001	<0.01	K00236	C	Succinate dehydrogenase (ubiquinone) cytochrome b560 subunit

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
rtg5261	1913	0.340104	<0.05	K00235	C	Succinate dehydrogenase (ubiquinone) iron-sulfur subunit [EC:1.3.5.1]
rtg603	987	0.600242	<0.01	K00237	C	Succinate dehydrogenase (ubiquinone) membrane anchor subunit
Succinyl-CoA synthase						
rtg5615	1600	0.385984	<0.01	K01900	C	Succinyl-coa synthetase beta subunit [EC:6.2.1.4 6.2.1.5]
rtg5749	2564	0.255147	<0.05	K01899	C	Succinyl-coa synthetase alpha subunit [EC:6.2.1.4 6.2.1.5]
alpha-ketoglutarate dH						
rtg1066	324	0.83069	<0.01	K00658	C	2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase) [EC:2.3.1.61]
rtg2808	180	0.972097	<0.01	K00164	G	2-oxoglutarate dehydrogenase E1 component [EC:1.2.4.2]
rtg3479	4498	0.145618	ns	K00164;K15791	G	2-oxoglutarate dehydrogenase E1 component [EC:1.2.4.2] probable 2-oxoglutarate dehydrogenase E1 component DHKTD1 [EC:1.2.4.2]
Isocitrate dH						
rtg3415	3893	0.145291	ns	K00031	C	Isocitrate dehydrogenase [EC:1.1.1.42]
rtg5809	1036	0.495925	<0.01	K00030	E	Isocitrate dehydrogenase (NAD+) [EC:1.1.1.41]
rtg5810	418	0.772541	<0.01	K00030	E	Isocitrate dehydrogenase (NAD+) [EC:1.1.1.41]
Aconitase						
rtg4484	252	0.846646	<0.01	K01681	C	Aconitate hydratase [EC:4.2.1.3]
Isocitrate lyase						
rtg5517	3932	0.126558	ns	K01637	C	Isocitrate lyase [EC:4.1.3.1]
Malate synthase						
rtg2130	3	-5.98348	n.a	K01638	C	Malate synthase [EC:2.3.3.9]
rtg3633	136	1.127702	<0.01	K01638	C	Malate synthase [EC:2.3.3.9]
De novo lipid synthesis						
Acetyl-CoA carboxylase (ACC)						
rtg226	57	1.329832	<0.01	K11262	I	Acetyl-coa carboxylase / biotin carboxylase 1 [EC:6.4.1.2 6.3.4.14 2.1.3.15]
Fatty acid synthase (FAS)						

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
rtg127	50	1.355341	<0.01	K00667	I	Fatty acid synthase subunit alpha, fungi type [EC:2.3.1.86]
rtg204	58	1.376327	<0.01	K00667; K00668	I	Fatty acid synthase subunit beta, fungi type [EC:2.3.1.86]
1-acylglycerol-3-phosphate O-acyltransferase (AGPAT)						
rtg5664	2596	0.266996	<0.1	K13519	I	Lysophospholipid acyltransferase [EC:2.3.1.51 2.3.1.23 2.3.1.-]
rtg2173	5298	-0.02788	ns		I	Acyltransferase (lysophospholipid acyltransferase)
rtg5355	5810	-0.00527	ns		I	Acyltransferase (lysophosphatidic acid acyltransferase / lysophosphatidylinositol acyltransferase)
Glycerol-3-phosphate acyltransferase (GPAT)						
rtg1283	3535	-0.25377	ns	K13507	I	Glycerol-3-phosphate O-acyltransferase / dihydroxyacetone phosphate acyltransferase [EC:2.3.1.15 2.3.1.42]
rtg3012	801	0.567038	<0.01	K13507	I	Glycerol-3-phosphate O-acyltransferase / dihydroxyacetone phosphate acyltransferase [EC:2.3.1.15 2.3.1.42]
rtg3414	4623	-0.1619	ns	K13507	I	Glycerol-3-phosphate O-acyltransferase / dihydroxyacetone phosphate acyltransferase [EC:2.3.1.15 2.3.1.42]
Phosphatidic acid phosphatase (PAP)						
rtg1218	3731	0.174715	ns	K18693	I	Diacylglycerol diphosphate phosphatase / phosphatidate phosphatase [EC:3.1.3.81 3.1.3.4]
rtg3174	2980	0.923743	ns			(refseq) bifunctional diacylglycerol diphosphate phosphatase/phosphatidate phosphatase< acidppc
rtg3837	4851	0.398036	ns			
Diacylglycerol O-acyltransferase (DGAT) and associated						
rtg257	4851	0.530466	<0.1	K15728	I	Phosphatidate phosphatase LPIN [EC:3.1.3.4]
rtg267	1385	0.478952	<0.05	K00679	I	Phospholipid:diacylglycerol acyltransferase [EC:2.3.1.158]
alpha, omega-oxidation						
Aldehyde dehydrogenase						
rtg1234	3133	0.237564	ns	K00128;K00129	C	Aldehyde dehydrogenase (NAD(P)+) [EC:1.2.1.5] aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
rtg1331	1672	-0.61742	<0.05		C	Aldehyde dehydrogenase family
rtg1455	1741	0.41167	<0.05		V	D-lactaldehyde dehydrogenase
rtg3413	4758	0.105004	ns	K00128	C	Aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
rtg4159	1901	0.360954	<0.01	K00128;K14085	C	Aldehyde dehydrogenase (NAD+) [EC:1.2.1.3] aldehyde dehydrogenase family 7 member A1 [EC:1.2.1.31 1.2.1.8 1.2.1.3]
rtg4562	976	-0.64903	<0.01	K00128	C	Aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
rtg4984	668	-0.64935	<0.01	K00128	C	Aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
rtg513	4741	0.101235	ns	K00128	C	Aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
rtg5244	2097	-0.84413	<0.05	K00128	C	Aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
rtg738	1018	-0.50214	<0.01	K00128	C	Aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]
Alcohol dehydrogenase						
rtg1625	1196	0.602372	<0.01	K00002	Q	Alcohol dehydrogenase (NADP+) [EC:1.1.1.2]
rtg2078	144	1.105919	<0.01	K00002	Q	Alcohol dehydrogenase (NADP+) [EC:1.1.1.2]
rtg2788	37	1.509143	<0.01	K00002	Q	Alcohol dehydrogenase (NADP+) [EC:1.1.1.2]
rtg289		0.795027	<0.01	K13953	Q	Alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]
rtg2985	4952	-0.17527	ns	K13953	Q	Alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]
rtg2999	114	1.40059	<0.01		C	Aryl-alcohol dehydrogenase
rtg3009	844	1.106993	<0.01	K13953	Q	Alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]
rtg3974	2441	-0.45701	<0.05	K13954	C	Alcohol dehydrogenase [EC:1.1.1.1]
rtg3996	776	0.554074	<0.01	K00121	Q	S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase [EC:1.1.1.284 1.1.1.1]
rtg4272	4812	-0.14568	ns	K13954		Alcohol dehydrogenase [EC:1.1.1.1]
rtg4461	5004	-0.12369	ns	K13953	Q	Alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]
rtg4464	4973	0.163424	ns	K13953	Q	Alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]
rtg5549	2598	0.34862	ns		C	Alcohol dehydrogenase groes-like domain
rtg6061	5183	-0.07196	ns	K13953;K13955	C	Alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1] zinc-binding alcohol dehydrogenase/oxidoreductase
rtg6101	5370	-0.2791	ns		Q	Alcohol dehydrogenase groes-like domain
Cytochrome P450 alkane monooxygenase/cytochrome P450 reductase complex						
rtg3373	2175	-0.77424	<0.05		Q	Cytochrome P450
rtg3536	1733	-0.5131	<0.1		Q	Cytochrome p450
rtg4089	380	0.369566	ns		Q	Cytochrome p450
rtg5028	2606	0.276924	ns		Q	Cytochrome P450

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
Fatty alcohol oxidase (FAO)				K17756		None found
Lipases						
rtg2110	3745	0.321266	ns		O	Secretory lipase (family LIP)
rtg2248	4610	-0.09527	ns	K06130;K06999		Lysophospholipase II [EC:3.1.1.5] phospholipase/carboxylesterase
rtg2252	31	2.678024	<0.01			(refseq) putative Lipase B precursor<
rtg2257	4026	0.806034	ns		I	Inherit from funog: GDSL-like Lipase/Acylhydrolase
rtg2311	399	1.074245	<0.01	K01054	I	Acylglycerol lipase [EC:3.1.1.23]
rtg2505	1516	0.555816	<0.05	K01046	I	Triacylglycerol lipase [EC:3.1.1.3]
rtg2599	1517	0.555816	<0.05	K01046	I	Triacylglycerol lipase [EC:3.1.1.3]
rtg3022	2988	0.242132	ns	K14018	I	Phospholipase A-2-activating protein
rtg3516	4122	0.146469	ns		I;O;T	Lipase (class 3)
rtg3673	538	1.268124	<0.01		O	Secretory lipase
rtg3917	101	1.41606	<0.01	K13333	I	Lysophospholipase [EC:3.1.1.5]
rtg409	3139	0.459692	ns		T	Extracellular triacylglycerol lipase precursor (EC 3.1.1.3)
rtg434	2261	-0.49125	<0.1	K13985	S	N-acyl-phosphatidylethanolamine-hydrolysing phospholipase D [EC:3.1.4.54]
rtg4451	84	1.900333	<0.01	K13278	E	60kda lysophospholipase [EC:3.1.1.5 3.1.1.47 3.5.1.1]
rtg4582	2328	0.254256	ns	K01046;K15979	K	Staphylococcal nuclease domain-containing protein 1 triacylglycerol lipase [EC:3.1.1.3]
rtg4806	1948	0.372918	<0.1			(refseq) triglyceride lipase<
rtg4919	4041	0.187818	ns		I	GDSL-like Lipase/Acylhydrolase
rtg498	2099	-0.50138	<0.1	K01115;K15429		Phospholipase D1/2 [EC:3.1.4.4] trna (guanine37-N1)- methyltransferase [EC:2.1.1.228]
rtg511	5730	0.109941	ns	K06130;K06999	I	Lysophospholipase II [EC:3.1.1.5] phospholipase/carboxylesterase
rtg5423	2591	0.237573	ns	K01115	I	Phospholipase D1/2 [EC:3.1.4.4]
rtg5514	2729	-0.37738	ns	K01054		Acylglycerol lipase [EC:3.1.1.23]
rtg5578	831	0.670736	<0.01	K01052	I	Lysosomal acid lipase/cholesteryl ester hydrolase [EC:3.1.1.13]
rtg6015	5646	-0.0113	ns		I	Phospholipase/carboxylesterase

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
rtg6067	5302	0.022295	ns	K00997;K13333;K16342		Cytosolic phospholipase A2 [EC:3.1.1.4] holo-[acyl-carrier protein] synthase [EC:2.7.8.7] lysophospholipase [EC:3.1.1.5]
rtg913	3449	-0.18737	ns			Partial alpha/beta-hydrolase lipase region
β-oxidation						
Acyl-Coa dH						
rtg2045	1273	-1.29835	<0.01			(refseq) bifunctional hydroxyacyl-coa dehydrogenase/enoyl-coa hydratase FOX2< short chain dehydrogenase
rtg23	2267	-0.86759	<0.05			Acyl-coa dehydrogenase
rtg2805	4708	-0.12365	ns	K00253		Isovaleryl-coa dehydrogenase [EC:1.3.8.4]
rtg4074	5070	-0.08105	ns			Acyl-coa dehydrogenase, C-terminal domain
rtg4571	2264	0.513542	<0.05	K00249		Acyl-coa dehydrogenase [EC:1.3.8.7]
rtg5452	41	-1.53047	<0.01			Acyl-coa dehydrogenase, C-terminal domain
rtg5959	2923	-0.23662	ns			Acyl-coa dehydrogenase, N-terminal domain
rtg701	669	-0.70349	<0.01			Acyl-coa dehydrogenase, C-terminal domain
enoyl-CoA hydratase						
rtg2045	1273	-1.29835	<0.01			(refseq) bifunctional hydroxyacyl-coa dehydrogenase/enoyl-coa hydratase FOX2< short chain dehydrogenase
rtg3404	863	-0.60749	<0.01			Enoyl-coa hydratase/isomerase family
rtg5742	2945	-0.73283	ns	K05605;K05607		3-hydroxyisobutyryl-coa hydrolase [EC:3.1.2.4] methylglutaconyl-coa hydratase [EC:4.2.1.18]
rtg781	4023	-0.15761	ns	K05605;K07511		3-hydroxyisobutyryl-coa hydrolase [EC:3.1.2.4] enoyl-coa hydratase [EC:4.2.1.17]
3L-hydroxyacyl-coa dehydrogenases						
rtg3194	848	-1.04956	<0.01	K00074		3-hydroxybutyryl-coa dehydrogenase [EC:1.1.1.157]
rtg1254	5159	0.077437	ns	K08683		3-hydroxyacyl-coa dehydrogenase / 3-hydroxy-2-methylbutyryl-coa dehydrogenase [EC:1.1.1.35 1.1.1.178]
β-ketothiolase						
rtg1529	3125	-0.22239	ns	K07513		Acetyl-coa acyltransferase 1 [EC:2.3.1.16]
rtg3157	4703	-0.11376	ns	K07513		Acetyl-coa acyltransferase 1 [EC:2.3.1.16]
rtg4298	1766	-0.36792	<0.01	K07513		Acetyl-coa acyltransferase 1 [EC:2.3.1.16]

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
Carotenoids synthesis						
rtg281	1189	0.55891	<0.01	K00804	H	Geranylgeranyl diphosphate synthase, type III [EC:2.5.1.1 2.5.1.10 2.5.1.29]
rtg4052	1768	0.356349	<0.1	K17841	I	15-cis-phytoene synthase / lycopene beta-cyclase [EC:2.5.1.32 5.5.1.19]
rtg4054	337	0.801796	<0.01	K15745	H	Phytoene desaturase (3,4-didehydrolycopene-forming) [EC:1.3.99.30]
rtg1071	2788	0.252253	ns	K01641	I	Hydroxymethylglutaryl-coa synthase [EC:2.3.3.10]
rtg2210	1378	0.499395	<0.1	K00021	I	Hydroxymethylglutaryl-coa reductase (NADPH) [EC:1.1.1.34]
rtg64	2091	-0.42366	<0.05	K01823	P	Isopentenyl-diphosphate Delta-isomerase [EC:5.3.3.2]
rtg55	6078	0.072335	ns	K01823	Q	Isopentenyl-diphosphate Delta-isomerase [EC:5.3.3.2]
rtg281	1189	0.55891	<0.01	K00804	H	Geranylgeranyl diphosphate synthase, type III [EC:2.5.1.1 2.5.1.10 2.5.1.29]
rtg5984	2166	-0.35186	ns	K00787	H	Farnesyl diphosphate synthase [EC:2.5.1.1 2.5.1.10]
Pentose phosphate pathway						
Oxidative branch						
rtg5491	2656	0.297504692	<0.1	K07404	G	6-phosphogluconolactonase [EC:3.1.1.31]
rtg994	695	0.586741311	<0.01	K01057	G	6-phosphogluconolactonase [EC:3.1.1.31]
rtg3203	190	0.93934475	<0.01	K00033	G	6-phosphogluconate dehydrogenase [EC:1.1.1.44 1.1.1.343]
Non-oxidative branch						
rtg2879	3444	0.363202447	ns	K00615		Transketolase [EC:2.2.1.1]
rtg2881	307	0.806224211	<0.01	K00615	G	Transketolase [EC:2.2.1.1]
rtg5013	13	1.94948339	<0.01	K01621	G	Xylulose-5-phosphate/fructose-6-phosphate phosphoketolase [EC:4.1.2.9 4.1.2.22]
rtg5082	2674	0.27808639	<0.1	K01807		Ribose 5-phosphate isomerase A [EC:5.3.1.6]
rtg11	5616	0.011283002	ns	K01783	G	Ribulose-phosphate 3-epimerase [EC:5.1.3.1]
rtg4731	1157	0.451830958	<0.01	K00616	G	Transaldolase [EC:2.2.1.2]
rtg3728	564	-	<0.01	K15634;K22315		Sedoheptulose-bisphosphatase [EC:3.1.3.37]
		0.769286911				

iD	Rank #	LFC	p _{adj}	KO	COG	Definition
rtg1810	1269	-	<0.1	K00852	G	Ribokinase [EC:2.7.1.15]
		0.554187819				
rtg1972	1920	-	<0.05	K01835	G	Phosphoglucomutase [EC:5.4.2.2]
		0.458514409				
rtg2834	2125	0.426807962	<0.01	K01835	G	Phosphoglucomutase [EC:5.4.2.2]
rtg3427	3269	-	ns	K00948	F	Ribose-phosphate pyrophosphokinase [EC:2.7.6.1]
		0.381710813				

Gene ranks in terms of contribution to PC1 were given in the column Rank #. The smallest the rank is the highest the contribution will be. LFC (Log2 fold) gives the estimated change between gene expression level of in SW₁₅ compared to Gly₁₅ (Ctrl). This change is associated with a FDR adjusted p-value either non-significant (ns), non-computed (na), lower than 0.01 (<0.01), between 0.01 and 0.05 (<0.05) or between 0.05 and 0.1 (<0.1). Only p_{adj} <0.01 were considered significant in the main manuscript

References

- Abeln, F., & Chuck, C. J. (2021). The history, state of the art and future prospects for oleaginous yeast research. *Microbial Cell Factories*, 20(1), 1-31.
<https://doi.org/10.1186/s12934-021-01712-1>
- Adrio, J. L. (2017). Oleaginous yeasts: Promising platforms for the production of oleochemicals and biofuels. *Biotechnology and Bioengineering*, 114(9), 1915-1920.
<https://doi.org/10.1002/bit.26337>
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10), R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
- Angers, P., Cadoret, N., Estaque, L., Garro, J. M., Jollez, P., & Lemieux, A. (2003). Production of high grade and high concentration of free fatty acids from residual oils, fats and greases (European Union Patent N° EP1280875A1).
<https://patents.google.com/patent/EP1280875A1/en>
- Antón, J. (2011). Halotolerance. In M. Gargaud, R. Amils, J. C. Quintanilla, H. J. (Jim) Cleaves, W. M. Irvine, D. L. Pinti, & M. Viso (Éds.), *Encyclopedia of Astrobiology* (p. 727-727). Springer. https://doi.org/10.1007/978-3-642-11274-4_695
- Ardi, M. S., Aroua, M. K., & Hashim, N. A. (2015). Progress, prospect and challenges in glycerol purification process: A review. *Renewable and Sustainable Energy Reviews*, 42, 1164-1173. <https://doi.org/10.1016/j.rser.2014.10.091>
- Arhar, S., Gogg-Fassolter, G., Ogrizović, M., Pačnik, K., Schwaiger, K., Žganjar, M., Petrovič, U., & Natter, K. (2021). Engineering of *Saccharomyces cerevisiae* for the accumulation of high amounts of triacylglycerol. *Microbial Cell Factories*, 20(1), 1-15.
<https://doi.org/10.1186/s12934-021-01640-0>
- Ariño, J., Velázquez, D., & Casamayor, A. (2019). Ser/Thr protein phosphatases in fungi: Structure, regulation and function. *Microbial Cell*, 6(5), 217-256.
<https://doi.org/10.15698/mic2019.05.677>
- Arthur, H., & Watson, K. (1976). Thermal adaptation in yeast: Growth temperatures, membrane lipid, and cytochrome composition of psychrophilic, mesophilic, and thermophilic yeasts. *Journal of Bacteriology*, 128(1), 56-68.

Athalye, S. K., Garcia, R. A., & Wen, Z. (2009). Use of Biodiesel-Derived Crude Glycerol for Producing Eicosapentaenoic Acid (EPA) by the Fungus *Pythium irregulare*. *Journal of Agricultural and Food Chemistry*, 57(7), 2739-2744. <https://doi.org/10.1021/jf803922w>

Avato, P., & Tava, A. (2021). Rare fatty acids and lipids in plant oilseeds : Occurrence and bioactivity. *Phytochemistry Reviews*. <https://doi.org/10.1007/s11101-021-09770-4>

Ayoub, M., & Abdullah, A. Z. (2012). Critical review on the current scenario and significance of crude glycerol resulting from biodiesel industry towards more sustainable renewable energy industry. *Renewable and Sustainable Energy Reviews*, 16(5), 2671-2686. <https://doi.org/10.1016/j.rser.2012.01.054>

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formisano, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., ... Zagnitko, O. (2008). The RAST Server : Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9, 75. <https://doi.org/10.1186/1471-2164-9-75>

Aznar-Moreno, J. A., & Durrett, T. P. (2017). Review : Metabolic engineering of unusual lipids in the synthetic biology era. *Plant Science: An International Journal of Experimental Plant Biology*, 263, 126-131. <https://doi.org/10.1016/j.plantsci.2017.07.007>

Banks, G. R., Shelton, P. A., Kanuga, N., Holden, D. W., & Spanos, A. (1993). The *Ustilago maydis nar1* gene encoding nitrate reductase activity: Sequence and transcriptional regulation. *Gene*, 131(1), 69-78. [https://doi.org/10.1016/0378-1119\(93\)90670-x](https://doi.org/10.1016/0378-1119(93)90670-x)

Barman, A., & Tamuli, R. (2015). Multiple cellular roles of *Neurospora crassa plc-1*, *splA2*, and *cpe-1* in regulation of cytosolic free calcium, carotenoid accumulation, stress responses, and acquisition of thermotolerance. *Journal of Microbiology (Seoul, Korea)*, 53(4), 226-235. <https://doi.org/10.1007/s12275-015-4465-1>

Barman, A., Gohain, D., Bora, U., & Tamuli, R. (2018). Phospholipases play multiple cellular roles including growth, stress tolerance, sexual development, and virulence in fungi. *Microbiological Research*, 209, 55-69. <https://doi.org/10.1016/j.micres.2017.12.012>

Bauer, F., & Hulteberg, C. (2013). Is there a future in glycerol as a feedstock in the production of biofuels and biochemicals? *Biofuels, Bioproducts and Biorefining*, 7(1), 43-51. <https://doi.org/10.1002/bbb.1370>

Becker, J., Lange, A., Fabarius, J., & Wittmann, C. (2015). Top value platform chemicals : Bio-based production of organic acids. *Current Opinion in Biotechnology*, 36, 168-175. <https://doi.org/10.1016/j.copbio.2015.08.022>

Bekatorou, A., Psarianos, C., & Koutinas, A. (2006). Production of Food Grade Yeasts. *Food Technology and Biotechnology*, 44.

Beltran, G., Novo, M., Guillamón, J. M., Mas, A., & Rozès, N. (2008). Effect of fermentation temperature and culture media on the yeast lipid composition and wine volatile compounds. *International Journal of Food Microbiology*, 121(2), 169-177. <https://doi.org/10.1016/j.ijfoodmicro.2007.11.030>

Benini, S. (2020). *Carbohydrate-Active Enzymes : Structure, Activity and Reaction Products*. MDPI.

Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M.-T., Hapala, I., Papanikolaou, S., Chardot, T., & Nicaud, J.-M. (2008). Control of Lipid Accumulation in the Yeast *Yarrowia lipolytica*. *Applied and Environmental Microbiology*, 74(24), 7779-7789. <https://doi.org/10.1128/AEM.01412-08>

Beopoulos, A., Chardot, T., & Nicaud, J.-M. (2009). *Yarrowia lipolytica* : A model and a tool to understand the mechanisms implicated in lipid accumulation. *Biochimie*, 91(6), 692-696. <https://doi.org/10.1016/j.biochi.2009.02.004>

Beopoulos, A., Nicaud, J.-M., & Gaillardin, C. (2011). An overview of lipid metabolism in yeasts and its impact on biotechnological processes. *Applied Microbiology and Biotechnology*, 90(4), 1193-1206. <https://doi.org/10.1007/s00253-011-3212-8>

Beopoulos, A., Haddouche, R., Kabran, P., Dulermo, T., Chardot, T., & Nicaud, J.-M. (2012). Identification and characterization of DGA2, an acyltransferase of the DGAT1 acyl-CoA:diacylglycerol acyltransferase family in the oleaginous yeast *Yarrowia lipolytica*. New insights into the storage lipid metabolism of oleaginous yeasts. *Applied Microbiology and Biotechnology*, 93(4), 1523-1537. <https://doi.org/10.1007/s00253-011-3506-x>

Bernauer, L., Radkohl, A., Lehmayr, L., & Emmerstorfer-Augustin, A. (2021). Komagataella phaffii as Emerging Model Organism in Fundamental Research. *Frontiers in Microbiology*, 11. <https://doi.org/10.3389/fmicb.2020.607028>

Beroeinc. (2018). Lipids Market Outlook, Share, Size, Forecast, Trends, Report. <https://www.beroeinc.com/category-intelligence/lipids-market/>

Bilsland-Marchesan, E., Ariño, J., Saito, H., Sunnerhagen, P., & Posas, F. (2000). Rck2 Kinase Is a Substrate for the Osmotic Stress-Activated Mitogen-Activated Protein Kinase Hog1. *Molecular and Cellular Biology*, 20(11), 3887-3895. <https://doi.org/10.1128/MCB.20.11.3887-3895.2000>

Black, C. S. (2013). Bioconversion of Glycerol to Dihydroxyacetone by immobilized *Gluconacetobacter xylinus* cells [Thesis, University of Waikato]. <https://researchcommons.waikato.ac.nz/handle/10289/7955>

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). antiSMASH 5.0 : Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(W1), W81-W87. <https://doi.org/10.1093/nar/gkz310>

Blitzblau, H. G., Consiglio, A. L., Teixeira, P., Crabtree, D. V., Chen, S., Konzock, O., Chifamba, G., Su, A., Kaminen, A., MacEwen, K., Hamilton, M., Tsakraklides, V., Nielsen, J., Siewers, V., & Shaw, A. J. (2021). Production of 10-methyl branched fatty acids in yeast. *Biotechnology for Biofuels*, 14(1), 12. <https://doi.org/10.1186/s13068-020-01863-0>

Blötz, C., & Stülke, J. (2017). Glycerol metabolism and its implication in virulence in *Mycoplasma*. *FEMS Microbiology Reviews*, 41(5), 640-652. <https://doi.org/10.1093/femsre/fux033>

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

Borneman, A. R., Gianoulis, T. A., Zhang, Z. D., Yu, H., Rozowsky, J., Seringhaus, M. R., Wang, L. Y., Gerstein, M., & Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species. *Science (New York, N.Y.)*, 317(5839), 815-819. <https://doi.org/10.1126/science.1140748>

Bornscheuer, U. T. (2018). *Lipid Modification by Enzymes and Engineered Microbes*. Elsevier.

Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., & Palsson, B. Ø. (2016). Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, 113(26), E3801-3809. <https://doi.org/10.1073/pnas.1523199113>

Bossie, M. A., & Martin, C. E. (1989). Nutritional regulation of yeast delta-9 fatty acid desaturase activity. *Journal of Bacteriology*, 171(12), 6409-6413. <https://doi.org/10.1128/jb.171.12.6409-6413.1989>

Botham, P. A., & Ratledge, C. (1979). A biochemical explanation for lipid accumulation in *Candida* 107 and other oleaginous micro-organisms. *Journal of General Microbiology*, 114(2), 361-375. <https://doi.org/10.1099/00221287-114-2-361>

Botstein, D., & Fink, G. R. (2011). *Yeast : An Experimental Organism for 21st Century Biology*. *Genetics*, 189(3), 695-704. <https://doi.org/10.1534/genetics.111.130765>

Bouchez, C., & Devin, A. (2019). Mitochondrial Biogenesis and Mitochondrial Reactive Oxygen Species (ROS): A Complex Relationship Regulated by the cAMP/PKA Signaling Pathway. *Cells*, 8(4), Article 4. <https://doi.org/10.3390/cells8040287>

Boulton, C. A., & Ratledge, C. (1981). Correlation of lipid accumulation in yeasts with possession of ATP: Citrate lyase. *Journal of General Microbiology*. <http://agris.fao.org/agris-search/search.do?recordID=US201301963520>

Boundy-Mills, K. (2012). Yeast culture collections of the world: Meeting the needs of industrial researchers. *Journal of Industrial Microbiology & Biotechnology*, 39(5), 673-680. <https://doi.org/10.1007/s10295-011-1078-5>

Braunwald, T., Schwemmlin, L., Graeff-Hönninger, S., French, W. T., Hernandez, R., Holmes, W. E., & Claupein, W. (2013). Effect of different C/N ratios on carotenoid and lipid production by *Rhodotorula glutinis*. *Applied Microbiology and Biotechnology*, 97(14), 6581-6588. <https://doi.org/10.1007/s00253-013-5005-8>

Brzica, H., Breljak, D., Burckhardt, B., Burckhardt, G., & Sabolic, I. (2013). Oxalate: From the Environment to Kidney Stones. *Arhiv za Higijenu Rada i Toksikologiju*, 64, 609-630. <https://doi.org/10.2478/10004-1254-64-2013-2428>

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59-60. <https://doi.org/10.1038/nmeth.3176>

Buck, J., & Andrews, J. (1999). Attachment of the Yeast *Rhodospordium toruloides* Is Mediated by Adhesives Localized at Sites of Bud Cell Development. *Applied and environmental microbiology*, 65, 465-471. <https://doi.org/10.1128/AEM.65.2.465-471.1999>

Buck, J., & Andrews, J. (1999). Attachment of the Yeast *Rhodospordium toruloides* Is Mediated by Adhesives Localized at Sites of Bud Cell Development. *Applied and environmental microbiology*, 65, 465-471. <https://doi.org/10.1128/AEM.65.2.465-471.1999>

Butler, G., Rasmussen, M. D., Lin, M. F., Santos, M. A. S., Sakthikumar, S., Munro, C. A., Rheinbay, E., Grabherr, M., Forche, A., Reedy, J. L., Agrafioti, I., Arnaud, M. B., Bates, S., Brown, A. J. P., Brunke, S., Costanzo, M. C., Fitzpatrick, D. A., de Groot, P. W. J., Harris, D., ... Cuomo, C. A. (2009). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature*, 459(7247), 657-662. <https://doi.org/10.1038/nature08064>

Buzzini, P., Innocenti, M., Turchetti, B., Libkind, D., van Broock, M., & Mulinacci, N. (2007). Carotenoid profiles of yeasts belonging to the genera *Rhodotorula*, *Rhodospordium*, *Sporobolomyces*, and *Sporidiobolus*. *Canadian Journal of Microbiology*, 53(8), 1024-1031. <https://doi.org/10.1139/W07-068>

Cai, S., Hu, C., & Du, S. (2007). Comparisons of Growth and Biochemical Composition between Mixed Culture of Alga and Yeast and Monocultures. *Journal of Bioscience and Bioengineering*, 104(5), 391-397. <https://doi.org/10.1263/jbb.104.391>

Caporusso, A., Capece, A., & De Bari, I. (2021). Oleaginous Yeasts as Cell Factories for the Sustainable Production of Microbial Lipids by the Valorization of Agri-Food Wastes. *Fermentation*, 7(2), 50. <https://doi.org/10.3390/fermentation7020050>

Casey, E., Sedlak, M., Ho, N. W. Y., & Mosier, N. S. (2010). Effect of acetic acid and pH on the cofermentation of glucose and xylose to ethanol by a genetically engineered strain of *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 10(4), 385-393. <https://doi.org/10.1111/j.1567-1364.2010.00623.x>

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

Cavallo, E., Charreau, H., Cerrutti, P., & Foresti, M. L. (2017). *Yarrowia lipolytica*: A model yeast for citric acid production. *FEMS Yeast Research*, 17(8). <https://doi.org/10.1093/femsyr/fox084>

Chang, A., Scheer, M., Grote, A., Schomburg, I., & Schomburg, D. (2009). BRENDA, AMENDA and FRENDA the enzyme information system: New content and tools in 2009. *Nucleic Acids Research*, 37(suppl_1), D588-D592. <https://doi.org/10.1093/nar/gkn820>

Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., & Schomburg, D. (2021). BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Research*, 49(D1), D498-D508. <https://doi.org/10.1093/nar/gkaa1025>

Chatzifragkou, A., & Papanikolaou, S. (2012). Effect of impurities in biodiesel-derived waste glycerol on the performance and feasibility of biotechnological processes. *Applied Microbiology and Biotechnology*, 95(1), 13-27. <https://doi.org/10.1007/s00253-012-4111-3>

Chen, W., Sommerfeld, M., & Hu, Q. (2011). Microwave-assisted Nile red method for *in vivo* quantification of neutral lipids in microalgae. *Bioresource Technology*, 102(1), 135-141. <https://doi.org/10.1016/j.biortech.2010.06.076>

Chen, X.-S., Ren, X.-D., Dong, N., Li, S., Li, F., Zhao, F.-L., Tang, L., Zhang, J.-H., & Mao, Z.-G. (2012). Culture medium containing glucose and glycerol as a mixed carbon source improves ϵ -poly-L-lysine production by *Streptomyces* sp. M-Z18. *Bioprocess and Biosystems Engineering*, 35(3), 469-475. <https://doi.org/10.1007/s00449-011-0586-z>

Chen, I.-M. A., Markowitz, V. M., Chu, K., Anderson, I., Mavromatis, K., Kyrpides, N. C., & Ivanova, N. N. (2013). Improving Microbial Genome Annotations in an Integrated Database Context. *PLoS ONE*, 8(2), Article 2. <https://doi.org/10.1371/journal.pone.0054859>

Chen, Z., & Liu, D. (2016). Toward glycerol biorefinery: Metabolic engineering for the production of biofuels and chemicals from glycerol. *Biotechnology for Biofuels*, 9. <https://doi.org/10.1186/s13068-016-0625-8>

Chen, L., Qian, X., Zhang, X., Zhou, X., Zhou, J., Dong, W., Xin, F., Zhang, W., Jiang, M., & Ochsenreither, K. (2021). Co-production of microbial lipids with valuable

chemicals. *Biofuels, Bioproducts and Biorefining*, 15(3), 945-954. <https://doi.org/10.1002/bbb.2209>

Chilakamarry, C. R., Sakinah, A. M. M., Zularisam, A. W., & Pandey, A. (2021). Glycerol waste to value added products and its potential applications. *Systems Microbiology and Biomanufacturing*, 1(4), 378-396. <https://doi.org/10.1007/s43393-021-00036-w>

Chiruvolu, V., Eskridge, K., Cregg, J., & Meagher, M. (1998). Effects of glycerol concentration and pH on growth of recombinant *Pichia pastoris* yeast. *Applied Biochemistry and Biotechnology*, 75(2), 163-173. <https://doi.org/10.1007/BF02787771>

Chun, B. H., Kim, K. H., Jeon, H. H., Lee, S. H., & Jeon, C. O. (2017). Pan-genomic and transcriptomic analyses of *Leuconostoc mesenteroides* provide insights into its genomic and metabolic features and roles in kimchi fermentation. *Scientific Reports*, 7(1), 1-16. <https://doi.org/10.1038/s41598-017-12016-z>

Ciriminna, R., Pina, C. D., Rossi, M., & Pagliaro, M. (2014). Understanding the glycerol market. *European Journal of Lipid Science and Technology*, 116(10), 1432-1439. <https://doi.org/10.1002/ejlt.201400229>

Cirulis, J. T., Strasser, B. C., Scott, J. A., & Ross, G. M. (2012). Optimization of staining conditions for microalgae with three lipophilic dyes to reduce precipitation and fluorescence variability. *Cytometry Part A*, 81A(7), 618-626. <https://doi.org/10.1002/cyto.a.22066>

Claret da Cunha, A., Gomes, L. S., Godoy-Santos, F., Faria-Oliveira, F., Teixeira, J. A., Sampaio, G. M. S., Trópia, M. J. M., Castro, I. M., Lucas, C., & Brandão, R. L. (2019). High-affinity transport, cyanide-resistant respiration, and ethanol production under aerobiosis underlying efficient high glycerol consumption by *Wickerhamomyces anomalus*. *Journal of Industrial Microbiology and Biotechnology*, 46(5), 709-723. <https://doi.org/10.1007/s10295-018-02119-5>

Clomburg, J. M., Crumbley, A. M., & Gonzalez, R. (2017). Industrial biomanufacturing : The future of chemical production. *Science*, 355(6320), aag0804. <https://doi.org/10.1126/science.aag0804>

Cofré, O., Ramírez, M., Gómez, J. M., & Cantero, D. (2016). Pilot scale fed-batch fermentation in a closed loop mixed reactor for the biotransformation of crude glycerol into ethanol and hydrogen by *Escherichia coli* MG1655. *Biomass and Bioenergy*, 91, 37-47. <https://doi.org/10.1016/j.biombioe.2016.04.015>

- Cohen, Z., & Ratledge, C. (2015). *Single Cell Oils : Microbial and Algal Oils*. Elsevier.
- Colin, T., Bories, A., Lavigne, C., & Moulin, G. (2001). Effects of acetate and butyrate during glycerol fermentation by *Clostridium butyricum*. *Current Microbiology*, 43(4), 238-243. <https://doi.org/10.1007/s002840010294>
- Colin, T., Bories, A., Lavigne, C., & Moulin, G. (2001). Effects of acetate and butyrate during glycerol fermentation by *Clostridium butyricum*. *Current Microbiology*, 43(4), 238-243. <https://doi.org/10.1007/s002840010294>
- Cooksey, K. E., Guckert, J. B., Williams, S. A., & Callis, P. R. (1987). Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red. *Journal of Microbiological Methods*, 6(6), 333-345. [https://doi.org/10.1016/0167-7012\(87\)90019-4](https://doi.org/10.1016/0167-7012(87)90019-4)
- 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 *Rhodospiridium toruloides*. *ELife*, 7, e32110. <https://doi.org/10.7554/eLife.32110>
- Coronado, C. J. R. (2013). Glycerol : Production, consumption, prices, characterization and new trends in combustion. *Renewable and Sustainable Energy Reviews*.
- Correia, K., & Mahadevan, R. (2018). Pan-genome-scale network reconstruction : A framework to increase the quantity and quality of metabolic network reconstructions throughout the tree of life. *BioRxiv*, 412593. <https://doi.org/10.1101/412593>
- Correia, K., Yu, S. M., & Mahadevan, R. (2019). AYbRAH : A curated ortholog database for yeasts and fungi spanning 600 million years of evolution. *Database*, 2019(baz022). <https://doi.org/10.1093/database/baz022>
- Correia, K., & Mahadevan, R. (2020). Pan-Genome-Scale Network Reconstruction : Harnessing Phylogenomics Increases the Quantity and Quality of Metabolic Models. *Biotechnology Journal*, 15(10), 1900519. <https://doi.org/10.1002/biot.201900519>
- De la Hoz Siegler, H., Ayidzoe, W., Ben-Zvi, A., Burrell, R. E., & McCaffrey, W. C. (2012). Improving the reliability of fluorescence-based neutral lipid content measurements in microalgal cultures. *Algal Research*, 1(2), 176-184. <https://doi.org/10.1016/j.algal.2012.07.004>

Denis, C. L., & Young, E. T. (1983). Isolation and characterization of the positive regulatory gene *ADR1* from *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 3(3), 360-370. <https://doi.org/10.1128/mcb.3.3.360-370.1983>

Díaz, T., Fillet, S., Campoy, S., Vázquez, R., Viña, J., Murillo, J., & Adrio, J. L. (2018). Combining evolutionary and metabolic engineering in *Rhodospiridium 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>

Dien, B. S., Slininger, P. J., Kurtzman, C. P., Moser, B. R., O'Bryan, P. J., Dien, B. S., Slininger, P. J., Kurtzman, C. P., Moser, B. R., & O'Bryan, P. J. (2016). Identification of superior lipid producing *Lipomyces* and *Myxozyma* yeasts. *AIMS Environmental Science*, 3(1), 1-20. <https://doi.org/10.3934/environsci.2016.1.1>

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 *Rhodospiridium toruloides* IFO0880 accounting for functional genomics and phenotypic data. *Metabolic Engineering Communications*, 9. <https://doi.org/10.1016/j.mec.2019.e00101>

Dissook, S., Putri, S. P., & Fukusaki, E. (2020). Metabolomic Analysis of Response to Nitrogen-Limiting Conditions in *Yarrowia* spp. *Metabolites*, 11(1), Article 1. <https://doi.org/10.3390/metabo11010016>

Djordjevic, J. T. (2010). Role of Phospholipases in Fungal Fitness, Pathogenicity, and Drug Development – Lessons from *Cryptococcus Neoformans*. *Frontiers in Microbiology*, 1. <https://doi.org/10.3389/fmicb.2010.00125>

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (Oxford, England), 29(1), 15-21. <https://doi.org/10.1093/bioinformatics/bts635>

Dos Santos, S. C., Teixeira, M. C., Dias, P. J., & Sá-Correia, I. (2014). MFS transporters required for multidrug/multixenobiotic (MD/MX) resistance in the model yeast: Understanding their physiological function through post-genomic approaches. *Frontiers in Physiology*, 5. <https://doi.org/10.3389/fphys.2014.00180>

Drăgan, M.-A., Moghul, I., Priyam, A., Bustos, C., & Wurm, Y. (2016). GeneValidator: Identify problems with protein-coding gene predictions. *Bioinformatics*, 32(10), 1559-1561. <https://doi.org/10.1093/bioinformatics/btw015>

Du, C. (2014). Fermentative Itaconic Acid Production. *Journal of Biodiversity, Bioprospecting and Development*, 01. <https://doi.org/10.4172/2376-0214.1000119>

Dujon, B. (2015). Basic principles of yeast genomics, a personal recollection. *FEMS Yeast Research*, 15(fov047), Article fov047. <https://doi.org/10.1093/femsyr/fov047>

Dulermo, T., Lazar, Z., Dulermo, R., Rakicka, M., Haddouche, R., & Nicaud, J.-M. (2015). Analysis of ATP-citrate lyase and malic enzyme mutants of *Yarrowia lipolytica* points out the importance of mannitol metabolism in fatty acid synthesis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1851(9), 1107-1117. <https://doi.org/10.1016/j.bbailip.2015.04.007>

Dutta, A. K., Kamada, K., & Ohta, K. (1996). Spectroscopic studies of Nile red in organic solvents and polymers. *Journal of Photochemistry and Photobiology A: Chemistry*, 93(1), 57-64. [https://doi.org/10.1016/1010-6030\(95\)04140-0](https://doi.org/10.1016/1010-6030(95)04140-0)

Eklom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7(9), 1026-1042. <https://doi.org/10.1111/eva.12178>

Eneva, R., Engibarov, S., Abrashev, R., Krumova, E., & Angelova, M. (2021). Sialic acids, sialoconjugates and enzymes of their metabolism in fungi. *Biotechnology & Biotechnological Equipment*, 35(1), 364-375. <https://doi.org/10.1080/13102818.2021.1879678>

Fabiszewska, A., Misiukiewicz-Stępień, P., Papińska-Goryca, M., Zieniuk, B., & Białecka-Florjańczyk, E. (2019). An Insight into Storage Lipid Synthesis by *Yarrowia lipolytica* Yeast Relating to Lipid and Sugar Substrates Metabolism. *Biomolecules*, 9(11), Article 11. <https://doi.org/10.3390/biom9110685>

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. <https://doi.org/10.3390/jof7040320>

Fakas, S., Papanikolaou, S., Galiotou-Panayotou, M., Komaitis, M., & Aggelis, G. (2008). Organic nitrogen of tomato waste hydrolysate enhances glucose uptake and lipid accumulation in *Cunninghamella echinulata*. *Journal of Applied Microbiology*, 105(4), 1062-1070. <https://doi.org/10.1111/j.1365-2672.2008.03839.x>

Fei, Q., O'Brien, M., Nelson, R., Chen, X., Lowell, A., & Dowe, N. (2016). Enhanced lipid production by *Rhodospiridium toruloides* using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source. *Biotechnology for Biofuels*, 9. <https://doi.org/10.1186/s13068-016-0542-x>

Fickers, P., Benetti, P.-H., Waché, Y., Marty, A., Mauersberger, S., Smit, M. S., & Nicaud, J.-M. (2005). Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *FEMS Yeast Research*, 5(6-7), 527-543. <https://doi.org/10.1016/j.femsyr.2004.09.004>

Fickers, P., Marty, A., & Nicaud, J. M. (2011). The lipases from *Yarrowia lipolytica*: Genetics, production, regulation, biochemical characterization and biotechnological applications. *Biotechnology Advances*, 29(6), 632-644. <https://doi.org/10.1016/j.biotechadv.2011.04.005>

Fillet, S., & Adrio, J. L. (2016). Microbial production of fatty alcohols. *World Journal of Microbiology & Biotechnology*, 32(9), 152. <https://doi.org/10.1007/s11274-016-2099-z>

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39(suppl_2), W29-W37. <https://doi.org/10.1093/nar/gkr367>

Foster, J., & Nakata, P. A. (2014). An oxalyl-CoA synthetase is important for oxalate metabolism in *Saccharomyces cerevisiae*. *FEBS Letters*, 588(1), 160-166. <https://doi.org/10.1016/j.febslet.2013.11.026>

Friedlander, J., Tsakraklides, V., Kamineni, A., Greenhagen, E. H., Consiglio, A. L., MacEwen, K., Crabtree, D. V., Afshar, J., Nugent, R. L., Hamilton, M. A., Joe Shaw, A., South, C. R., Stephanopoulos, G., & Brevnova, E. E. (2016). Engineering of a high lipid producing *Yarrowia lipolytica* strain. *Biotechnology for Biofuels*, 9(1), 77. <https://doi.org/10.1186/s13068-016-0492-3>

Fukuda, R. (2013). Metabolism of hydrophobic carbon sources and regulation of it in n-alkane-assimilating yeast *Yarrowia lipolytica*. *Bioscience, Biotechnology, and Biochemistry*, 77(6), 1149-1154. <https://doi.org/10.1271/bbb.130164>

Funk, I. (2020). *Biotechnological Production of Dicarboxylic Acids as Building Blocks for Bio-based Polymers*. Universitätsbibliothek der TU München.

Gadd, G. M., Bahri-Esfahani, J., Li, Q., Rhee, Y. J., Wei, Z., Fomina, M., & Liang, X. (2014). Oxalate production by fungi: Significance in geomycology, biodeterioration and bioremediation. *Fungal Biology Reviews*, 28(2), 36-55. <https://doi.org/10.1016/j.fbr.2014.05.001>

Galdieri, L., Chang, J., Mehrotra, S., & Vancura, A. (2013). Yeast Phospholipase C Is Required for Normal Acetyl-CoA Homeostasis and Global Histone Acetylation*. *Journal of Biological Chemistry*, 288(39), 27986-27998. <https://doi.org/10.1074/jbc.M113.492348>

Gao, Z., Ma, Y., Wang, Q., Zhang, M., Wang, J., & Liu, Y. (2016). Effect of crude glycerol impurities on lipid preparation by *Rhodospiridium toruloides* yeast 32489. *Bioresource Technology*, 218(Supplement C), 373-379. <https://doi.org/10.1016/j.biortech.2016.06.088>

Garay, L. A., Boundy-Mills, K. L., & German, J. B. (2014). Accumulation of High-Value Lipids in Single-Cell Microorganisms: A Mechanistic Approach and Future Perspectives. *Journal of Agricultural and Food Chemistry*, 62(13), 2709-2727. <https://doi.org/10.1021/jf4042134>

Garay, L. A., Sitepu, I. R., Cajka, T., Chandra, I., Shi, S., Lin, T., German, J. B., Fiehn, O., & Boundy-Mills, K. L. (2016). Eighteen new oleaginous yeast species. *Journal of Industrial Microbiology & Biotechnology*, 43(7), 887-900. <https://doi.org/10.1007/s10295-016-1765-3>

Garcia, E. J., Vevea, J. D., & Pon, L. A. (2018). Lipid droplet autophagy during energy mobilization, lipid homeostasis and protein quality control. *Frontiers in bioscience (Landmark edition)*, 23, 1552-1563.

Garlapati, V. K., Shankar, U., & Budhiraja, A. (2015). Bioconversion technologies of crude glycerol to value added industrial products. *Biotechnology Reports*, 9, 9-14. <https://doi.org/10.1016/j.btre.2015.11.002>

Gerde, J. A., Montalbo-Lomboy, M., Yao, L., Grewell, D., & Wang, T. (2012). Evaluation of microalgae cell disruption by ultrasonic treatment. *Bioresource Technology*, 125, 175-181. <https://doi.org/10.1016/j.biortech.2012.08.110>

Gilman, J., Walls, L., Bandiera, L., & Menolascina, F. (2021). Statistical Design of Experiments for Synthetic Biology. *ACS Synthetic Biology*, 10(1), 1-18. <https://doi.org/10.1021/acssynbio.0c00385>

Gordon, J. L., Armisen, D., Proux-Wéra, E., ÓhÉigeartaigh, S. S., Byrne, K. P., & Wolfe, K. H. (2011). Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), 20024-20029. <https://doi.org/10.1073/pnas.1112808108>

Gore-Lloyd, D., Sumann, I., Brachmann, A. O., Schneeberger, K., Ortiz-Merino, R. A., Moreno-Beltrán, M., Schläfli, M., Kirner, P., Santos Kron, A., Rueda-Mejia, M. P., Somerville, V., Wolfe, K. H., Piel, J., Ahrens, C. H., Henk, D., & Freimoser, F. M. (2019). Snf2 controls pulcherriminic acid biosynthesis and antifungal activity of the biocontrol yeast *Metschnikowia pulcherrima*. *Molecular Microbiology*, 112(1), 317-332. <https://doi.org/10.1111/mmi.14272>

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, 38(Web Server issue), W695-699. <https://doi.org/10.1093/nar/gkq313>

Greenea. (2015). Glycerine market: Lack of interdependence between supply and demand. <http://www.greenea.com/publication/glycerine-market-lack-of-interdependence-between-supply-and-demand/>

Green, M. R., & Sambrook, J. (2016). Precipitation of DNA with Ethanol. *Cold Spring Harbor Protocols*, 2016(12), Article 12. <https://doi.org/10.1101/pdb.prot093377>

Greenspan, P., Mayer, E. P., & Fowler, S. D. (1985). Nile red: A selective fluorescent stain for intracellular lipid droplets. *Journal of Cell Biology*, 100(3), 965-973. <https://doi.org/10.1083/jcb.100.3.965>

Griesemer, M., Kimbrel, J. A., Zhou, C. E., Navid, A., & D'haeseleer, P. (2018). Combining multiple functional annotation tools increases coverage of metabolic annotation. *BMC Genomics*, 19(1), 948. <https://doi.org/10.1186/s12864-018-5221-9>

Groguenin, A., Waché, Y., Garcia, E. E., Aguedo, M., Husson, F., LeDall, M.-T., Nicaud, J.-M., & Belin, J.-M. (2004). Genetic engineering of the β -oxidation pathway in the yeast *Yarrowia lipolytica* to increase the production of aroma compounds. *Journal of Molecular Catalysis B: Enzymatic*, 28(2), 75-79. <https://doi.org/10.1016/j.molcatb.2004.01.006>

Gross, H., & König, G. (2006). Terpenoids from Marine Organisms: Unique Structures and their Pharmacological Potential. *Phytochemistry Rev*, 5, 115-141. <https://doi.org/10.1007/s11101-005-5464-3>

Guimarães, L. C., Florczak-Wyspianska, J., de Jesus, L. B., Viana, M. V. C., Silva, A., Ramos, R. T. J., Soares, S. de C., & Soares, S. de C. (2015). Inside the Pan-genome—Methods and Software Overview. *Current Genomics*, 16(4), 245-252. <https://doi.org/10.2174/1389202916666150423002311>

Guirimand, G., Kulagina, N., Papon, N., Hasunuma, T., & Courdavault, V. (2021). Innovative Tools and Strategies for Optimizing Yeast Cell Factories. *Trends in Biotechnology*, 39(5), 488-504. <https://doi.org/10.1016/j.tibtech.2020.08.010>

Gunstone, F., & Heming, M. (2004). Glycerol—An important product of the oleochemical industry. Undefined. <https://www.semanticscholar.org/paper/Glycerol-%E2%80%94an-important-product-of-the-oleochemical-Gunstone-Heming/da6d215cdc0e3712028d194d2e6acd274ad23c89>

Gurr, M. I., Harwood, J. L., & Frayn, K. N. (2008). *Lipid Biochemistry: An Introduction*. John Wiley & Sons.

Haapalainen, A. M., Meriläinen, G., Pirilä, P. L., Kondo, N., Fukao, T., & Wierenga, R. K. (2007). Crystallographic and kinetic studies of human mitochondrial acetoacetyl-CoA thiolase: The importance of potassium and chloride ions for its structure and function. *Biochemistry*, 46(14), 4305-4321. <https://doi.org/10.1021/bi6026192>

Hafez, R. M., Abdel-Rahman, T. M., & Naguib, R. M. (2017). Uric acid in plants and microorganisms: Biological applications and genetics - A review. *Journal of Advanced Research*, 8(5), 475-486. <https://doi.org/10.1016/j.jare.2017.05.003>

Hansen, C. F., Hernandez, A., Mullan, B. P., Moore, K., Trezona-Murray, M., King, R. H., Pluske, J. R., Hansen, C. F., Hernandez, A., Mullan, B. P., Moore, K., Trezona-Murray, M., King, R. H., & Pluske, J. R. (2009). A chemical analysis of samples of crude glycerol from the production of biodiesel in Australia, and the effects of feeding crude glycerol to growing-finishing pigs on performance, plasma metabolites and meat quality at slaughter. *Animal Production Science*, 49(2), 154-161. <https://doi.org/10.1071/EA08210>

Hao, P., Zheng, H., Yu, Y., Ding, G., Gu, W., Chen, S., Yu, Z., Ren, S., Oda, M., Konno, T., Wang, S., Li, X., Ji, Z.-S., & Zhao, G. (2011). Complete Sequencing and Pan-Genomic Analysis of *Lactobacillus delbrueckii* subsp. *Bulgaricus* Reveal Its Genetic Basis for Industrial Yogurt Production. *PLOS ONE*, 6(1), e15964. <https://doi.org/10.1371/journal.pone.0015964>

Hazimah, H., Ooi, T. L., & Salmiah, A. (2003). Recovery of glycerol and diglycerol from glycerol pitch. *Journal of Oil Palm Research*, 15, 1-5.

Hein, E.-M., & Hayen, H. (2012). Comparative Lipidomic Profiling of *S. cerevisiae* and Four Other Hemiascomycetous Yeasts. *Metabolites*, 2(1), 254-267. <https://doi.org/10.3390/metabo2010254>

Hicks, R. H., Chuck, C. J., Scott, R. J., Leak, D. J., & Henk, D. A. (2019). Comparison of Nile Red and Cell Size Analysis for High-Throughput Lipid Estimation Within Oleaginous Yeast. *European Journal of Lipid Science and Technology*, 121(11), 1800355. <https://doi.org/10.1002/ejlt.201800355>

Hill, K. (2001). Fats and Oils as Oleochemical Raw Materials. *Journal of Oleo Science*, 50(5), 433-444. <https://doi.org/10.5650/jos.50.433>

Hiltunen, J. K., Schonauer, M. S., Autio, K. J., Mittelmeier, T. M., Kastaniotis, A. J., & Dieckmann, C. L. (2009). Mitochondrial Fatty Acid Synthesis Type II: More than Just Fatty Acids. *The Journal of Biological Chemistry*, 284(14), 9011-9015. <https://doi.org/10.1074/jbc.R800068200>

Hoff, K. J., & Stanke, M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, 65(1), e57. <https://doi.org/10.1002/cpbi.57>

Hoja, U., Marthol, S., Hofmann, J., Stegner, S., Schulz, R., Meier, S., Greiner, E., & Schweizer, E. (2004). HFA1 encoding an organelle-specific acetyl-CoA carboxylase controls mitochondrial fatty acid synthesis in *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*, 279(21), 21779-21786. <https://doi.org/10.1074/jbc.M401071200>

Holdsworth, J. E., Veenhuis, M., & Ratledge, C. (1988). Enzyme activities in oleaginous yeasts accumulating and utilizing exogenous or endogenous lipids. *Journal of General Microbiology*, 134(11), 2907-2915. <https://doi.org/10.1099/00221287-134-11-2907>

Holladay, J. E., White, J. F., Bozell, J. J., & Johnson, D. (2007). Top value-added chemicals from biomass-Volume II—Results of screening for potential candidates from biorefinery lignin. Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

Holm, A. K. (2013). Investigation of glycerol assimilation and cofactor metabolism in *Lactococcus lactis*. Technical University of Denmark (DTU).

Hounslow, E., Noirel, J., Gilmour, D. J., & Wright, P. C. (2017). Lipid quantification techniques for screening oleaginous species of microalgae for biofuel production.

European Journal of Lipid Science and Technology, 119(2), 1500469.
<https://doi.org/10.1002/ejlt.201500469>

Hu, C., Zhao, X., Zhao, J., Wu, S., & Zhao, Z. K. (2009). Effects of biomass hydrolysis by-products on oleaginous yeast *Rhodospiridium toruloides*. *Bioresource Technology*, 100(20), 4843-4847. <https://doi.org/10.1016/j.biortech.2009.04.041>

Hu, S., Luo, X., Wan, C., & Li, Y. (2012). Characterization of Crude Glycerol from Biodiesel Plants. *Journal of Agricultural and Food Chemistry*, 60(23), 5915-5921. <https://doi.org/10.1021/jf3008629>

Huang, G.-H., Chen, G., & Chen, F. (2009). Rapid screening method for lipid production in alga based on Nile red fluorescence. *Biomass and Bioenergy*, 33(10), 1386-1392. <https://doi.org/10.1016/j.biombioe.2009.05.022>

Huang, X.-F., Liu, J.-N., Lu, L.-J., Peng, K.-M., Yang, G.-X., & Liu, J. (2016). Culture strategies for lipid production using acetic acid as sole carbon source by *Rhodospiridium toruloides*. *Bioresource Technology*, 206, 141-149. <https://doi.org/10.1016/j.biortech.2016.01.073>

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., von Mering, C., & Bork, P. (2016). eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44(D1), D286-D293. <https://doi.org/10.1093/nar/gkv1248>

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309-D314. <https://doi.org/10.1093/nar/gky1085>

Huf, S., Krügener, S., Hirth, T., Rupp, S., & Zibek, S. (2011). Biotechnological synthesis of long-chain dicarboxylic acids as building blocks for polymers. *European Journal of Lipid Science and Technology*, 113(5), 548-561. <https://doi.org/10.1002/ejlt.201000112>

Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Palaniappan, K., Szeto, E., Pillay, M., Chen, I.-M. A., Pati, A., Nielsen, T., Markowitz, V. M., & Kyrpides, N. C. (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Standards in Genomic Sciences*, 10. <https://doi.org/10.1186/s40793-015-0077-y>

IHS Markit. (2018). Glycerin—Chemical Economics Handbook (CEH). <https://ihsmarkit.com/products/glycerin-chemical-economics-handbook.html>

Ilic, S., Konstantinović, S., Ćirić, J., Savic, D., Gojgic-Cvijovic, G., & Veljković, V. (2016). Crude glycerol and whey as carbon and nitrogen sources for the production of antibiotics. *Advanced technologies*, 5, 5-9. <https://doi.org/10.5937/savteh1601005I>

Iqbal, H., Kyazze, G., & Keshavarz, T. (2013). Advances in the Valorization of Lignocellulosic Materials by Biotechnology: An Overview. *BioResources*, 8. <https://doi.org/10.15376/biores.8.2.3157-3176>

Irvan, Trisakti, B., Hasibuan, R., & Joli, M. (2018). Fermentative utilization of glycerol residue for the production of acetic acid. *IOP Conference Series: Materials Science and Engineering*, 309, 012126. <https://doi.org/10.1088/1757-899X/309/1/012126>

Isahak, W. N. R. W., Ramli, Z. A. C., Ismail, M., Jahim, J. M., & Yarmo, M. A. (2015). Recovery and Purification of Crude Glycerol from Vegetable Oil Transesterification. *Separation & Purification Reviews*, 44(3), 250-267. <https://doi.org/10.1080/15422119.2013.851696>

Ito, T., Nakashimada, Y., Senba, K., Matsui, T., & Nishio, N. (2005). Hydrogen and ethanol production from glycerol-containing wastes discharged after biodiesel manufacturing process. *Journal of Bioscience and Bioengineering*, 100(3), 260-265. <https://doi.org/10.1263/jbb.100.260>

Jacobs, M. N., Covaci, A., Gheorghe, A., & Schepens, P. (2004). Time Trend Investigation of PCBs, PBDEs, and Organochlorine Pesticides in Selected n–3 Polyunsaturated Fatty Acid Rich Dietary Fish Oil and Vegetable Oil Supplements; Nutritional Relevance for Human Essential n–3 Fatty Acid Requirements. *Journal of Agricultural and Food Chemistry*, 52(6), 1780-1788. <https://doi.org/10.1021/jf035310q>

Jantzen, S. G., Sutherland, B. J., Minkley, D. R., & Koop, B. F. (2011). GO Trimming : Systematically reducing redundancy in large Gene Ontology datasets. *BMC Research Notes*, 4(1), 267. <https://doi.org/10.1186/1756-0500-4-267>

Jarry, A., & Seraudie, Y. (1994). Process for the production of itaconic acid by fermentation (European Union Patent No EP0684314B1). <https://patents.google.com/patent/EP0684314B1/en>

Jeffries, T. W., Grigoriev, I. V., Grimwood, J., Laplaza, J. M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., Jin, Y.-S., Passoth, V., & Richardson, P. M. (2007). Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nature Biotechnology*, 25(3), 319-326. <https://doi.org/10.1038/nbt1290>

Jiang, Y., Liu, W., Zou, H., Cheng, T., Tian, N., & Xian, M. (2014). Microbial production of short chain diols. *Microbial Cell Factories*, 13(1), 165. <https://doi.org/10.1186/s12934-014-0165-5>

Johns, A. M. B., Love, J., & Aves, S. J. (2016). Four Inducible Promoters for Controlled Gene Expression in the Oleaginous Yeast *Rhodotorula toruloides*. *Frontiers in Microbiology*, 7, 1666. <https://doi.org/10.3389/fmicb.2016.01666>

Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, 11(1), 431. <https://doi.org/10.1186/1471-2105-11-431>

Jordan, M., & Stettler, M. (2014). Tools for high-throughput process and medium optimization. *Methods in Molecular Biology* (Clifton, N.J.), 1104, 77-88. https://doi.org/10.1007/978-1-62703-733-4_6

Jungermann, E., & Sontag, N. O. V. (1991). *Glycerine: A key cosmetic ingredient*. New York: Marcel Dekker. <https://trove.nla.gov.au/version/24787025>

Jönsson, L. J., & Martín, C. (2016). Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresource Technology*, 199, 103-112. <https://doi.org/10.1016/j.biortech.2015.10.009>

Redhu, A., Shah, A. H., & Prasad, R. (2016). MFS transporters of *Candida* species and their role in clinical drug resistance. *FEMS Yeast Research*, 16(fow043), Article fow043. <https://doi.org/10.1093/femsyr/fow043>

Kamisaka, Y., Tomita, N., Kimura, K., Kainou, K., & Uemura, H. (2007). DGA1 (diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the Δ snf2 disruptant of *Saccharomyces cerevisiae*. *Biochemical Journal*, 408(1), 61-68. <https://doi.org/10.1042/BJ20070449>

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, 42(Database issue), D199-D205. <https://doi.org/10.1093/nar/gkt1076>

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, 428(4), 726-731. <https://doi.org/10.1016/j.jmb.2015.11.006>

Karamerou, E. E., Theodoropoulos, C., & Webb, C. (2017). Evaluating feeding strategies for microbial oil production from glycerol by *Rhodotorula glutinis*. *Engineering in Life Sciences*, 17(3), 314-324. <https://doi.org/10.1002/elsc.201600073>

Kaur, J., Sarma, A. K., Jha, M. K., & Gera, P. (2020). Valorisation of crude glycerol to value-added products: Perspectives of process technology, economics and environmental issues. *Biotechnology Reports*, 27, e00487. <https://doi.org/10.1016/j.btre.2020.e00487>

Kavšček, M., Bhutada, G., Madl, T., & Natter, K. (2015). Optimization of lipid production with a genome-scale model of *Yarrowia lipolytica*. *BMC Systems Biology*, 9. <https://doi.org/10.1186/s12918-015-0217-4>

Kazamia, E., Czesnick, H., Nguyen, T. T. V., Croft, M. T., Sherwood, E., Sasso, S., Hodson, S. J., Warren, M. J., & Smith, A. G. (2012). Mutualistic interactions between vitamin B12-dependent algae and heterotrophic bacteria exhibit regulation. *Environmental Microbiology*, 14(6), 1466-1476. <https://doi.org/10.1111/j.1462-2920.2012.02733.x>

Khersonsky, O., & Tawfik, D. S. (2010). 8.03—Enzyme Promiscuity – Evolutionary and Mechanistic Aspects. In H.-W. (Ben) Liu & L. Mander (Éds.), *Comprehensive Natural Products II* (p. 47-88). Elsevier. <https://doi.org/10.1016/B978-008045382-8.00155-6>

Kim, M., Park, B. G., Kim, E.-J., Kim, J., & Kim, B.-G. (2019). In silico identification of metabolic engineering strategies for improved lipid production in *Yarrowia lipolytica* by genome-scale metabolic modelling. *Biotechnology for Biofuels*, 12(1), 187. <https://doi.org/10.1186/s13068-019-1518-4>

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 *Rhodospiridium toruloides*. *Frontiers in Bioengineering and Biotechnology*, 8, 1484. <https://doi.org/10.3389/fbioe.2020.612832>

Kimura, K., Yamaoka, M., & Kamisaka, Y. (2004). Rapid estimation of lipids in oleaginous fungi and yeasts using Nile red fluorescence. *Journal of Microbiological Methods*, 56(3), 331-338. <https://doi.org/10.1016/j.mimet.2003.10.018>

Kitcha, S., & Cheirsilp, B. (2011). Screening of Oleaginous Yeasts and Optimization for Lipid Production Using Crude Glycerol as a Carbon Source. *Energy Procedia*, 9, 274-282. <https://doi.org/10.1016/j.egypro.2011.09.029>

Klein, M., Islam, Z., Knudsen, P. B., Carrillo, M., Swinnen, S., Workman, M., & Nevoigt, E. (2016). The expression of glycerol facilitators from various yeast species improves growth on glycerol of *Saccharomyces cerevisiae*. *Metabolic Engineering Communications*, 3, 252-257. <https://doi.org/10.1016/j.meteno.2016.09.001>

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

Klosinska, M. M., Crutchfield, C. A., Bradley, P. H., Rabinowitz, J. D., & Broach, J. R. (2011). Yeast cells can access distinct quiescent states. *Genes & Development*, 25(4), 336-349. <https://doi.org/10.1101/gad.2011311>

Klug, L., & Daum, G. (2014). Yeast lipid metabolism at a glance. *FEMS Yeast Research*, 14(3), 369-388. <https://doi.org/10.1111/1567-1364.12141>

Knothe, G. (2010). Biodiesel and renewable diesel: A comparison. *Progress in Energy and Combustion Science*, 36(3), 364-373. <https://doi.org/10.1016/j.pecs.2009.11.004>

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 *Rhodospiridium toruloides*. *BMC Microbiology*, 14, 50. <https://doi.org/10.1186/1471-2180-14-50>

Kolouchová, I., Maťátková, O., Sigler, K., Masák, J., & Řezanka, T. (2016). Lipid accumulation by oleaginous and non-oleaginous yeast strains in nitrogen and phosphate limitation. *Folia Microbiologica*, 61(5), 431-438. <https://doi.org/10.1007/s12223-016-0454-y>

Komesu, A., Oliveira, J., Martins, L., Maciel, M., & Filho, R. (2017). Lactic acid production to purification: A review. *BioResources*, 12, 4364-4383. <https://doi.org/10.15376/biores.12.2.4364-4383>

Kosaric, N., & Sukan, F. V. (2010). *Biosurfactants: Production: Properties: Applications*. CRC Press.

Kruyssen, F. J., de Boer, W. R., & Wouters, J. T. (1980). Effects of carbon source and growth rate on cell wall composition of *Bacillus subtilis* subsp. *Niger*. *Journal of Bacteriology*, 144(1), 238-246.

Kregiel, D., Pawlikowska, E., & Antolak, H. (2017). Non-Conventional Yeasts in Fermentation Processes: Potentialities and Limitations. *Old Yeasts - New Questions*. <https://doi.org/10.5772/intechopen.70404>

Kumar, N., & Skolnick, J. (2012). EFICAz2.5: Application of a high-precision enzyme function predictor to 396 proteomes. *Bioinformatics*, 28(20), 2687-2688. <https://doi.org/10.1093/bioinformatics/bts510>

Kumar, L. R., Yellapu, S. K., Tyagi, R. D., & Zhang, X. (2019). A review on variation in crude glycerol composition, bio-valorization of crude and purified glycerol as carbon source for lipid production. *Bioresource Technology*, 293, 122155. <https://doi.org/10.1016/j.biortech.2019.122155>

Kumar, B., Bhardwaj, N., Agrawal, K., Chaturvedi, V., & Verma, P. (2020). Current perspective on pretreatment technologies using lignocellulosic biomass: An emerging biorefinery concept. *Fuel Processing Technology*, 199, 106244. <https://doi.org/10.1016/j.fuproc.2019.106244>

Lachance, J.-C., Rodrigue, S., & Palsson, B. O. (2019). Minimal cells, maximal knowledge. *eLife*, 8, e45379. <https://doi.org/10.7554/eLife.45379>

Lamers, D., van Biezen, N., Martens, D., Peters, L., van de Zilver, E., Jacobs-van Dreumel, N., Wijffels, R. H., & Lokman, C. (2016). Selection of oleaginous yeasts for fatty acid production. *BMC Biotechnology*, 16, 45. <https://doi.org/10.1186/s12896-016-0276-7>

Lamprecht, A., & Benoit, J.-P. (2003). Simple liquid-chromatographic method for Nile Red quantification in cell culture in spite of photobleaching. *Journal of Chromatography B*, 787(2), 415-419. [https://doi.org/10.1016/S1570-0232\(02\)00962-5](https://doi.org/10.1016/S1570-0232(02)00962-5)

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0.

Bioinformatics (Oxford, England), 23(21), 2947-2948.
<https://doi.org/10.1093/bioinformatics/btm404>

Larroude, M., Rossignol, T., Nicaud, J.-M., & Ledesma-Amaro, R. (2018). Synthetic biology tools for engineering *Yarrowia lipolytica*. *Biotechnology Advances*, 36(8), 2150-2164. <https://doi.org/10.1016/j.biotechadv.2018.10.004>

Ledesma-Amaro, R., & Nicaud, J.-M. (2016). *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. *Progress in Lipid Research*, 61, 40-50. <https://doi.org/10.1016/j.plipres.2015.12.001>

Lee, R. A., & Lavoie, J.-M. (2013). From first- to third-generation biofuels : Challenges of producing a commodity from a biomass of increasing complexity. *Animal Frontiers*, 3(2), 6-11. <https://doi.org/10.2527/af.2013-0010>

Lee, J. J. L., Chen, L., Cao, B., & Chen, W. N. (2016). Engineering *Rhodospiridium toruloides* with a membrane transporter facilitates production and separation of carotenoids and lipids in a bi-phasic culture. *Applied Microbiology and Biotechnology*, 100(2), 869-877. <https://doi.org/10.1007/s00253-015-7102-3>

Leman, J. (1997). Oleaginous Microorganisms : An Assessment of the Potential. In S. L. Neidleman & A. I. Laskin (Éds.), *Advances in Applied Microbiology* (Vol. 43, p. 195-243). Academic Press. [https://doi.org/10.1016/S0065-2164\(08\)70226-0](https://doi.org/10.1016/S0065-2164(08)70226-0)

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4 : Recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256-W259. <https://doi.org/10.1093/nar/gkz239>

Levering, J., Broddrick, J., & Zengler, K. (2015). Engineering of oleaginous organisms for lipid production. *Current Opinion in Biotechnology*, 36(Supplement C), 32-39. <https://doi.org/10.1016/j.copbio.2015.08.001>

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL : Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*, 13(9), 2178-2189. <https://doi.org/10.1101/gr.1224503>

Li, Y., Zhao, Z. (Kent), & Bai, F. (2007). High-density cultivation of oleaginous yeast *Rhodospiridium toruloides* Y4 in fed-batch culture. *Enzyme and Microbial Technology*, 41(3), 312-317. <https://doi.org/10.1016/j.enzmictec.2007.02.008>

Li, Q., Du, W., & Liu, D. (2008). Perspectives of microbial oils for biodiesel production. *Applied Microbiology and Biotechnology*, 80(5), 749-756. <https://doi.org/10.1007/s00253-008-1625-9>

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), 323. <https://doi.org/10.1186/1471-2105-12-323>

Li, Z., Sun, H., Mo, X., Li, X., Xu, B., & Tian, P. (2013). Overexpression of malic enzyme (ME) of *Mucor circinelloides* improved lipid accumulation in engineered *Rhodotorula glutinis*. *Applied Microbiology and Biotechnology*, 97(11), 4927-4936. <https://doi.org/10.1007/s00253-012-4571-5>

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>

Li-Beisson, Y., & Peltier, G. (2013). Third-generation biofuels: Current and future research on microalgal lipid biotechnology. *OCL*, 20(6), D606. <https://doi.org/10.1051/ocl/2013031>

Lin, E. C. (1976). Glycerol dissimilation and its regulation in bacteria. *Annual Review of Microbiology*, 30, 535-578. <https://doi.org/10.1146/annurev.mi.30.100176.002535>

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 *Rhodospiridium toruloides*. *FEMS Yeast Research*, 14(4), 547-555. <https://doi.org/10.1111/1567-1364.12140>

Liu, H., Zhao, X., Wang, F., Li, Y., Jiang, X., Ye, M., Zhao, Z. K., & Zou, H. (2009). Comparative proteomic analysis of *Rhodospiridium toruloides* during lipid accumulation. *Yeast*, 26(10), 553-566. <https://doi.org/10.1002/yea.1706>

Liu, Y., Koh, C. M. J., Ngoh, S. T., & Ji, L. (2015). Engineering an efficient and tight d-amino acid-inducible gene expression system in *Rhodospiridium/Rhodotorula* species. *Microbial Cell Factories*, 14, 170. <https://doi.org/10.1186/s12934-015-0357-7>

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* (*Rhodospiridium*)

yeast species. *Microbial Cell Factories*, 15(1), 200. <https://doi.org/10.1186/s12934-016-0600-x>

Liu, W., Wu, S., Lin, Q., Gao, S., Ding, F., Zhang, X., Aljohi, H. A., Yu, J., & Hu, S. (2018). RGAAT: A Reference-based Genome Assembly and Annotation Tool for New Genomes and Upgrade of Known Genomes. *Genomics, Proteomics & Bioinformatics*, 16(5), 373-381. <https://doi.org/10.1016/j.gpb.2018.03.006>

Liu, Q., Yu, T., Li, X., Chen, Y., Campbell, K., Nielsen, J., & Chen, Y. (2019). Rewiring carbon metabolism in yeast for high level production of aromatic chemicals. *Nature Communications*, 10(1), 4976. <https://doi.org/10.1038/s41467-019-12961-5>

Loewenstein, Y., Raimondo, D., Redfern, O. C., Watson, J., Frishman, D., Linial, M., Orengo, C., Thornton, J., & Tramontano, A. (2009). Protein function annotation by homology-based inference. *Genome Biology*, 10(2), 207. <https://doi.org/10.1186/gb-2009-10-2-207>

Lopes, H., & Rocha, I. (2017). Genome-scale modelling of yeast: Chronology, applications and critical perspectives. *FEMS Yeast Research*, 17(5). <https://doi.org/10.1093/femsyr/fox050>

Lu, H., Chen, H., Tang, X., Yang, Q., Zhang, H., Chen, Y. Q., & Chen, W. (2020). Time-resolved multi-omics analysis reveals the role of nutrient stress-induced resource reallocation for TAG accumulation in oleaginous fungus *Mortierella alpina*. *Biotechnology for Biofuels*, 13(1), 116. <https://doi.org/10.1186/s13068-020-01757-1>

Lubuta, P., Workman, M., Kerkhoven, E. J., & Workman, C. T. (2019). Investigating the Influence of Glycerol on the Utilization of Glucose in *Yarrowia lipolytica* Using RNA-Seq-Based Transcriptomics. *G3 (Bethesda, Md.)*, 9(12), 4059-4071. <https://doi.org/10.1534/g3.119.400469>

Ludin, K. M., Hilti, N., & Schweingruber, M. E. (1995). *Schizosaccharomyces pombe* rds1, an adenine-repressible gene regulated by glucose, ammonium, phosphate, carbon dioxide and temperature. *Molecular & General Genetics: MGG*, 248(4), 439-445. <https://doi.org/10.1007/BF02191644>

Lynen, F., & Ochoa, S. (1953). Enzymes of fatty acid metabolism. *Biochimica et Biophysica Acta*, 12(1), 299-314. [https://doi.org/10.1016/0006-3002\(53\)90149-8](https://doi.org/10.1016/0006-3002(53)90149-8)

López-Hernández, T., Haucke, V., & Maritzen, T. (2020). Endocytosis in the adaptation to cellular stress. *Cell Stress*, 4(10), 230-247. <https://doi.org/10.15698/cst2020.10.232>

Löbs, A.-K., Schwartz, C., & Wheeldon, I. (2017). Genome and metabolic engineering in non-conventional yeasts : Current advances and applications. *Synthetic and Systems Biotechnology*, 2(3), 198-207. <https://doi.org/10.1016/j.synbio.2017.08.002>

Madhavan, A., Jose, A. A., Binod, P., Sindhu, R., Sukumaran, R. K., Pandey, A., & Castro, G. E. (2017). Synthetic Biology and Metabolic Engineering Approaches and Its Impact on Non-Conventional Yeast and Biofuel Production. *Frontiers in Energy Research*, 5. <https://doi.org/10.3389/fenrg.2017.00008>

Madzak, C., Gaillardin, C., & Beckerich, J.-M. (2004). Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica* : A review. *Journal of Biotechnology*, 109(1-2), 63-81. <https://doi.org/10.1016/j.jbiotec.2003.10.027>

Magdouli, S., Brar, S. K., & Blais, J. F. (2016). Co-culture for lipid production : Advances and challenges. *Biomass and Bioenergy*, 92, 20-30. <https://doi.org/10.1016/j.biombioe.2016.06.003>

Magdouli, S., Guedri, T., Tarek, R., Brar, S. K., & Blais, J. F. (2017). Valorization of raw glycerol and crustacean waste into value added products by *Yarrowia lipolytica*. *Bioresource Technology*, 243, 57-68. <https://doi.org/10.1016/j.biortech.2017.06.074>

Magnan, C., Yu, J., Chang, I., Jahn, E., Kanomata, Y., Wu, J., Zeller, M., Oakes, M., Baldi, P., & Sandmeyer, S. (2016). Sequence Assembly of *Yarrowia lipolytica* Strain W29/CLIB89 Shows Transposable Element Diversity. *PLOS ONE*, 11(9), e0162363. <https://doi.org/10.1371/journal.pone.0162363>

Mantzouridou, F., Naziri, E., & Tsimidou, M. Z. (2008). Industrial glycerol as a supplementary carbon source in the production of beta-carotene by *Blakeslea trispora*. *Journal of Agricultural and Food Chemistry*, 56(8), 2668-2675. <https://doi.org/10.1021/jf703667d>

Mao, X., Cai, T., Olyarchuk, J. G., & Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*, 21(19), 3787-3793. <https://doi.org/10.1093/bioinformatics/bti430>

Marschall, T., Marz, M., Abeel, T., Dijkstra, L., Dutilh, B. E., Ghaffaari, A., Kersey, P., Kloosterman, W. P., Mäkinen, V., Novak, A. M., Paten, B., Porubsky, D., Rivals, E., Alkan, C., Baaijens, J. A., De Bakker, P. I. W., Boeva, V., Bonnal, R. J. P., Chiaromonte, F., ... Schönhuth, A. (2018). Computational pan-genomics : Status,

promises and challenges. *Briefings in Bioinformatics*, 19(1), 118-135.
<https://doi.org/10.1093/bib/bbw089>

Martin, R., Hellwig, D., Schaub, Y., Bauer, J., Walther, A., & Wendland, J. (2007). Functional analysis of *Candida albicans* genes whose *Saccharomyces cerevisiae* homologues are involved in endocytosis. *Yeast* (Chichester, England), 24(6), 511-522.
<https://doi.org/10.1002/yea.1489>

Matatkova, O., Gharwalova, L., Zimola, M., Rezanka, T., Masak, J., & Kolouchova, I. (2017). Using Odd-Alkanes as a Carbon Source to Increase the Content of Nutritionally Important Fatty Acids in *Candida krusei*, *Trichosporon cutaneum*, and *Yarrowia lipolytica*. *International Journal of Analytical Chemistry*, 2017.
<https://doi.org/10.1155/2017/8195329>

McCarthy, C. G. P., & Fitzpatrick, D. A. (2019a). Pan-genome analyses of model fungal species. *Microbial Genomics*, 5(2). <https://doi.org/10.1099/mgen.0.000243>

McCarthy, C. G. P., & Fitzpatrick, D. A. (2019b). Pangloss : A Tool for Pan-Genome Analysis of Microbial Eukaryotes. *Genes*, 10(7), 521.
<https://doi.org/10.3390/genes10070521>

McMillan, J. D., & Beckham, G. T. (2017). Thinking big : Towards ideal strains and processes for large-scale aerobic biofuels production. *Microbial Biotechnology*, 10(1), 40-42. <https://doi.org/10.1111/1751-7915.12471>

Menon, V., & Rao, M. (2012). Trends in bioconversion of lignocellulose : Biofuels, platform chemicals & biorefinery concept. *Progress in Energy and Combustion Science*, 38(4), 522-550. <https://doi.org/10.1016/j.pecs.2012.02.002>

Meunchan, M., Michely, S., Devillers, H., Nicaud, J.-M., Marty, A., & Neuvéglise, C. (2015). Comprehensive Analysis of a Yeast Lipase Family in the *Yarrowia* Clade. *PLOS ONE*, 10(11), e0143096. <https://doi.org/10.1371/journal.pone.0143096>

Min, B., Grigoriev, I. V., & Choi, I.-G. (2017). FunGAP : Fungal Genome Annotation Pipeline using evidence-based gene model evaluation. *Bioinformatics*, 33(18), 2936-2937.
<https://doi.org/10.1093/bioinformatics/btx353>

Miranda, C., Bettencourt, S., Pozdniakova, T., Pereira, J., Sampaio, P., Franco-Duarte, R., & Pais, C. (2020). Modified high-throughput Nile red fluorescence assay for the rapid screening of oleaginous yeasts using acetic acid as carbon source. *BMC Microbiology*, 20(1), 60. <https://doi.org/10.1186/s12866-020-01742-6>

Mishra, P., Lee, N.-R., Lakshmanan, M., Kim, M., Kim, B.-G., & Lee, D.-Y. (2018). Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica*. *BMC Systems Biology*, 12(Suppl 2). <https://doi.org/10.1186/s12918-018-0542-5>

Miyagi, T., & Yamaguchi, K. (2012). Mammalian sialidases: Physiological and pathological roles in cellular functions. *Glycobiology*, 22(7), 880-896. <https://doi.org/10.1093/glycob/cws057>

Molano, E. P. L., Cabrera, O. G., Jose, J., do Nascimento, L. C., Carazzolle, M. F., Teixeira, P. J. P. L., Alvarez, J. C., Tiburcio, R. A., Tokimatu Filho, P. M., de Lima, G. M. A., Guido, R. V. C., Corrêa, T. L. R., Leme, A. F. P., Mieczkowski, P., & Pereira, G. A. G. (2018). *Ceratocystis cacaofunesta* genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). *BMC Genomics*, 19(1), 58. <https://doi.org/10.1186/s12864-018-4440-4>

Molecular devices. (2021, août 23). Molecular devices spectramax m2 user manual. <https://www.manualslib.com/manual/1363060/Molecular-Devices-Spectramax-M2.html>

Montalbo-Lomboy, M., Kantekin Erdogan, M., & Wang, T. (2014). Lipid Estimation of Surfactant-Extracted Microalgae Oil Using Nile Red. *Journal of the American Oil Chemists' Society*, 91. <https://doi.org/10.1007/s11746-013-2395-9>

Monteiro, M. R., Kugelmeier, C. L., Pinheiro, R. S., Batalha, M. O., & da Silva César, A. (2018). Glycerol from biodiesel production: Technological paths for sustainability. *Renewable and Sustainable Energy Reviews*, 88, 109-122. <https://doi.org/10.1016/j.rser.2018.02.019>

Moon, C., Ahn, J.-H., Kim, S. W., Sang, B.-I., & Um, Y. (2010). Effect of Biodiesel-derived Raw Glycerol on 1,3-Propanediol Production by Different Microorganisms. *Applied Biochemistry and Biotechnology*, 161(1), 502-510. <https://doi.org/10.1007/s12010-009-8859-6>

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35(Web Server issue), W182-W185. <https://doi.org/10.1093/nar/gkm321>

Morschett, H., Wiechert, W., & Oldiges, M. (2016). Automation of a Nile red staining assay enables high throughput quantification of microalgal lipid production. *Microbial Cell Factories*, 15. <https://doi.org/10.1186/s12934-016-0433-7>

Mota, C., Pinto, B. P., & Lima, A. L. de. (2017). *Glycerol: A Versatile Renewable Feedstock for the Chemical Industry*. Springer International Publishing. <https://www.springer.com/br/book/9783319593746>

Możejko-Ciesielska, J., & Pokoj, T. (2018). Exploring nutrient limitation for polyhydroxyalkanoates synthesis by newly isolated strains of *Aeromonas* sp. Using biodiesel-derived glycerol as a substrate. *PeerJ*, 6, e5838. <https://doi.org/10.7717/peerj.5838>

Mukherjee, V., Radecka, D., Aerts, G., Verstrepen, K. J., Lievens, B., & Thevelein, J. M. (2017). Phenotypic landscape of non-conventional yeast species for different stress tolerance traits desirable in bioethanol fermentation. *Biotechnology for Biofuels*, 10. <https://doi.org/10.1186/s13068-017-0899-5>

Mullikin, J. C. (2014). The evolution of comparative genomics. *Molecular Genetics & Genomic Medicine*, 2(5), 363-368. <https://doi.org/10.1002/mgg3.112>

Muniraj, I. K., Uthandi, S. K., Hu, Z., Xiao, L., & Zhan, X. (2015). Microbial lipid production from renewable and waste materials for second-generation biodiesel feedstock. *Environmental Technology Reviews*, 4(1), 1-16. <https://doi.org/10.1080/21622515.2015.1018340>

Muranushi, N., Takagi, N., Muranishi, S., & Sezaki, H. (1981). Effect of fatty acids and monoglycerides on permeability of lipid bilayer. *Chemistry and Physics of Lipids*, 28(3), 269-279.

Nachtweide, S., & Stanke, M. (2019). Multi-Genome Annotation with AUGUSTUS. In M. Kollmar (Éd.), *Gene Prediction: Methods and Protocols* (p. 139-160). Springer. https://doi.org/10.1007/978-1-4939-9173-0_8

Nakase, T., & Komagata, K. (1971). Significance of DNA base composition in the classification of yeast genera *Cryptococcus* and *Rhodotorula*. *Journal of General and Applied Microbiology - J GEN APPL MICROBIOL TOKYO*, 17, 121-130. <https://doi.org/10.2323/jgam.17.121>

Navarrete, C., Jacobsen, I. H., Martínez, J. L., & Procentese, A. (2020). Cell Factories for Industrial Production Processes: Current Issues and Emerging Solutions. *Processes*, 8(7), 768. <https://doi.org/10.3390/pr8070768>

Nicaud, J.-M., Coq, A.-M. C. L., Rossignol, T., & Morin, N. (2014). Protocols for monitoring growth and lipid accumulation in oleaginous yeasts. Editions Springer. https://doi.org/10.1007/8623_2014_40

Nicol, R. W., Marchand, K., & Lubitz, W. D. (2012). Bioconversion of crude glycerol by fungi. *Applied Microbiology and Biotechnology*, 93(5), 1865-1875. <https://doi.org/10.1007/s00253-012-3921-7>

Ning, P., Yang, G., Hu, L., Sun, J., Shi, L., Zhou, Y., Wang, Z., & Yang, J. (2021). Recent advances in the valorization of plant biomass. *Biotechnology for Biofuels*, 14(1), 102. <https://doi.org/10.1186/s13068-021-01949-3>

Niu, B., Zhang, H., Zhou, G., Zhang, S., Yang, Y., Deng, X., & Chen, Q. (2021). Safety risk assessment and early warning of chemical contamination in vegetable oil. *Food Control*, 125, 107970. <https://doi.org/10.1016/j.foodcont.2021.107970>

OCDE & FAO. (2020). OECD-FAO Agricultural Outlook 2020-2029. <https://www.oecd-ilibrary.org/content/publication/1112c23b-en>

Opulente, D. A., Morales, C. M., Carey, L. B., & Rest, J. S. (2013). Coevolution Trumps Pleiotropy: Carbon Assimilation Traits Are Independent of Metabolic Network Structure in Budding Yeast. *PLOS ONE*, 8(1), e54403. <https://doi.org/10.1371/journal.pone.0054403>

Opulente, D. A., Rollinson, E. J., Bernick-Roehr, C., Hulfachor, A. B., Rokas, A., Kurtzman, C. P., & Hittinger, C. T. (2018). Factors driving metabolic diversity in the budding yeast subphylum. *BMC Biology*, 16(1), 26. <https://doi.org/10.1186/s12915-018-0498-3>

Pagliari, M., & Rossi, M. (2008). *The Future of Glycerol: New Uses of a Versatile Raw Material*. Royal Society of Chemistry.

Pagliari, M. (2017). *Glycerol: The Renewable Platform Chemical*. Elsevier.

Panis, G., & Carreon, J. R. (2016). Commercial astaxanthin production derived by green alga *Haematococcus pluvialis*: A microalgae process model and a techno-economic assessment all through production line. *Algal Research*, 18, 175-190. <https://doi.org/10.1016/j.algal.2016.06.007>

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., & Aggelis, G. (2006). Influence of Glucose and Saturated Free-Fatty Acid Mixtures on

Citric Acid and Lipid Production by *Yarrowia lipolytica*. *Current Microbiology*, 52(2), 134-142. <https://doi.org/10.1007/s00284-005-0223-7>

Park, Y., Ledesma-Amaro, R., & Nicaud, J.-M. (2020). De novo Biosynthesis of Odd-Chain Fatty Acids in *Yarrowia lipolytica* Enabled by Modular Pathway Engineering. *Frontiers in Bioengineering and Biotechnology*, 7, 484. <https://doi.org/10.3389/fbioe.2019.00484>

Patel, A., Antonopoulou, I., Enman, J., Rova, U., Christakopoulos, P., & Matsakas, L. (2019). Lipids detection and quantification in oleaginous microorganisms : An overview of the current state of the art. *BMC Chemical Engineering*, 1(1), 13. <https://doi.org/10.1186/s42480-019-0013-9>

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). <https://doi.org/10.1128/genomeA.00046-14>

Pfleger, B. F., Gossing, M., & Nielsen, J. (2015). Metabolic engineering strategies for microbial synthesis of oleochemicals. *Metabolic Engineering*, 29(Supplement C), 1-11. <https://doi.org/10.1016/j.ymben.2015.01.009>

Piškur, J., & Langkjær, R. B. (2004). Yeast genome sequencing : The power of comparative genomics. *Molecular Microbiology*, 53(2), 381-389. <https://doi.org/10.1111/j.1365-2958.2004.04182.x>

Poontawee, R., & Limtong, S. (2020). Feeding Strategies of Two-Stage Fed-Batch Cultivation Processes for Microbial Lipid Production from Sugarcane Top Hydrolysate and Crude Glycerol by the Oleaginous Red Yeast *Rhodospiridiobolus fluvialis*. *Microorganisms*, 8(2), Article 2. <https://doi.org/10.3390/microorganisms8020151>

Posada, J. A., Higuera, J. C., & Cardona, C. A. (2011). Optimization on the Use of Crude Glycerol from the Biodiesel Production to Obtain Poly-3-Hydroxybutyrate. https://ep.liu.se/konferensartikel.aspx?series=e&issue=57&volume=1&Article_No=44

- Proux-Wéra, E., Armisen, D., Byrne, K. P., & Wolfe, K. H. (2012). A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics*, 13(1), 237. <https://doi.org/10.1186/1471-2105-13-237>
- Punta, M., & Ofran, Y. (2008). The rough guide to in silico function prediction, or how to use sequence and structure information to predict protein function. *PLoS Computational Biology*, 4(10), e1000160. <https://doi.org/10.1371/journal.pcbi.1000160>
- Putri, D. N., Sahlan, M., Montastruc, L., Meyer, M., Negny, S., & Hermansyah, H. (2020). Progress of fermentation methods for bio-succinic acid production using agro-industrial waste by *Actinobacillus succinogenes*. *Energy Reports*, 6, 234-239. <https://doi.org/10.1016/j.egy.2019.08.050>
- Pyle, D. J., Garcia, R. A., & Wen, Z. (2008). Producing Docosahexaenoic Acid (DHA)-Rich Algae from Biodiesel-Derived Crude Glycerol: Effects of Impurities on DHA Production and Algal Biomass Composition. *Journal of Agricultural and Food Chemistry*, 56(11), 3933-3939. <https://doi.org/10.1021/jf800602s>
- Qadariyah, L., Mahfud, Sumarno, Machmudah, S., Wahyudiono, Sasaki, M., & Goto, M. (2011). Degradation of glycerol using hydrothermal process. *Bioresource Technology*, 102(19), 9267-9271. <https://doi.org/10.1016/j.biortech.2011.06.066>
- Qi, F., Kitahara, Y., Wang, Z., Zhao, X., Du, W., & Liu, D. (2014). Novel mutant strains of *Rhodospiridium toruloides* by plasma mutagenesis approach and their tolerance for inhibitors in lignocellulosic hydrolyzate. *Journal of Chemical Technology & Biotechnology*, 89(5), 735-742. <https://doi.org/10.1002/jctb.4180>
- Qi, F., Zhao, X., Kitahara, Y., Li, T., Ou, X., Du, W., Liu, D., & Huang, J. (2017). Integrative transcriptomic and proteomic analysis of the mutant lignocellulosic hydrolyzate-tolerant *Rhodospiridium toruloides*. *Engineering in Life Sciences*, 17(3), 249-261. <https://doi.org/10.1002/elsc.201500143>
- Qin, L., Liu, L., Zeng, A., & Wei, D. (2017). From low-cost substrates to Single Cell Oils synthesized by oleaginous yeasts. *Bioresource Technology*, 245. <https://doi.org/10.1016/j.biortech.2017.05.163>
- Radecka, D., Mukherjee, V., Mateo, R. Q., Stojiljkovic, M., Foulquié-Moreno, M. R., & Thevelein, J. M. (2015). Looking beyond *Saccharomyces*: The potential of non-conventional yeast species for desirable traits in bioethanol fermentation. *FEMS Yeast Research*, 15(6). <https://doi.org/10.1093/femsyr/fov053>

Raimondi, S., Rossi, M., Leonardi, A., Bianchi, M. M., Rinaldi, T., & Amaretti, A. (2014). Getting lipids from glycerol: New perspectives on biotechnological exploitation of *Candida freyschussii*. *Microbial Cell Factories*, 13, 83. <https://doi.org/10.1186/1475-2859-13-83>

Ratledge, C. (2002). Regulation of lipid accumulation in oleaginous micro-organisms. *Biochemical Society Transactions*, 30(Pt 6), 1047-1050. <https://doi.org/10.1042/>

Ratledge, C. (2013). Microbial oils: An introductory overview of current status and future prospects. *OCL*, 20(6), D602. <https://doi.org/10.1051/ocl/2013029>

Razavi, S., & Marc, I. (2006). Effect of Temperature and pH on the Growth Kinetics and Carotenoid Production by *Sporobolomyces ruberrimus* H110 Using Technical Glycerol as Carbon Source. *Iranian Journal of Chemistry and Chemical Engineering*, 23.

Reimand, J., Vaquerizas, J. M., Todd, A. E., Vilo, J., & Luscombe, N. M. (2010). Comprehensive reanalysis of transcription factor knockout expression data in *Saccharomyces cerevisiae* reveals many new targets. *Nucleic Acids Research*, 38(14), 4768-4777. <https://doi.org/10.1093/nar/gkq232>

Riley, R., Haridas, S., Wolfe, K. H., Lopes, M. R., Hittinger, C. T., Göker, M., Salamov, A. A., Wisecaver, J. H., Long, T. M., Calvey, C. H., Aerts, A. L., Barry, K. W., Choi, C., Clum, A., Coughlan, A. Y., Deshpande, S., Douglass, A. P., Hanson, S. J., Klenk, H.-P., ... Jeffries, T. W. (2016). Comparative genomics of biotechnologically important yeasts. *Proceedings of the National Academy of Sciences of the United States of America*, 113(35), 9882-9887. <https://doi.org/10.1073/pnas.1603941113>

Rostron, K. A., & Lawrence, C. L. (2017). Nile Red Staining of Neutral Lipids in Yeast. In *Histochemistry of Single Molecules* (p. 219-229). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-6788-9_16

Ruiz-Perez, C. A., Conrad, R. E., & Konstantinidis, K. T. (2021). MicrobeAnnotator: A user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics*, 22(1), 11. <https://doi.org/10.1186/s12859-020-03940-5>

Rumin, J., Bonnefond, H., Saint-Jean, B., Rouxel, C., Sciandra, A., Bernard, O., Cadoret, J.-P., & Bougaran, G. (2015). The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae. *Biotechnology for Biofuels*, 8(1), 42. <https://doi.org/10.1186/s13068-015-0220-4>

Runguphan, W., & Keasling, J. D. (2014). Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metabolic Engineering*, 21, 103-113. <https://doi.org/10.1016/j.ymben.2013.07.003>

Ruy, A. D. da S., Ferreira, A. L. F., Bresciani, A. É., Alves, R. M. de B., & Pontes, L. A. M. (2020). Market Prospecting and Assessment of the Economic Potential of Glycerol from Biodiesel. In *Biotechnological Applications of Biomass*. IntechOpen. <https://doi.org/10.5772/intechopen.93965>

Sackett, D. L., & Wolff, J. (1987). Nile red as a polarity-sensitive fluorescent probe of hydrophobic protein surfaces. *Analytical Biochemistry*, 167(2), 228-234. [https://doi.org/10.1016/0003-2697\(87\)90157-6](https://doi.org/10.1016/0003-2697(87)90157-6)

Saini, R., Osorio-Gonzalez, C. S., Hegde, K., Brar, S. K., Magdouli, S., Vezina, P., & Avalos-Ramirez, A. (2020). Lignocellulosic Biomass-Based Biorefinery : An Insight into Commercialization and Economic Standout. *Current Sustainable/Renewable Energy Reports*, 7(4), 122-136. <https://doi.org/10.1007/s40518-020-00157-1>

Salzberg, S. L. (2019). Next-generation genome annotation : We still struggle to get it right. *Genome Biology*, 20(1), 92. <https://doi.org/10.1186/s13059-019-1715-2>

Samul, D., Leja, K., & Grajek, W. (2014). Impurities of crude glycerol and their effect on metabolite production. *Annals of Microbiology*, 64(3), 891-898. <https://doi.org/10.1007/s13213-013-0767-x>

San, J. E., Baichoo, S., Kanzi, A., Moosa, Y., Lessells, R., Fonseca, V., Mogaka, J., Power, R., & de Oliveira, T. (2020). Current Affairs of Microbial Genome-Wide Association Studies : Approaches, Bottlenecks and Analytical Pitfalls. *Frontiers in Microbiology*, 10, 3119. <https://doi.org/10.3389/fmicb.2019.03119>

Santos, C. A., Ferreira, M. E., Lopes da Silva, T., Gouveia, L., Novais, J. M., & Reis, A. (2011). A symbiotic gas exchange between bioreactors enhances microalgal biomass and lipid productivities : Taking advantage of complementary nutritional modes. *Journal of Industrial Microbiology & Biotechnology*, 38(8), 909-917. <https://doi.org/10.1007/s10295-010-0860-0>

Santos, C. A., Caldeira, M. L., Lopes da Silva, T., Novais, J. M., & Reis, A. (2013). Enhanced lipidic algae biomass production using gas transfer from a fermentative *Rhodospiridium toruloides* culture to an autotrophic *Chlorella protothecoides* culture. *Bioresource Technology*, 138, 48-54. <https://doi.org/10.1016/j.biortech.2013.03.135>

Sasson, O., Kaplan, N., & Linial, M. (2006). Functional annotation prediction: All for one and one for all. *Protein Science: A Publication of the Protein Society*, 15(6), 1557-1562. <https://doi.org/10.1110/ps.062185706>

Savojardo, C., Martelli, P. L., Fariselli, P., Profiti, G., & Casadio, R. (2018). BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Research*, 46(W1), W459-W466. <https://doi.org/10.1093/nar/gky320>

Scheyder, E. (2009, août 11). BP to fund Martek research on biodiesel. Reuters. <https://www.reuters.com/article/uk-bp-martek-sb-idUKTRE57A3AX20090811>

Schneider, J., Blom, J., Jaenicke, S., Linke, B., Brinkrolf, K., Neuweger, H., Tauch, A., & Goesmann, A. (2011). RPYD — Rapid Annotation Platform for Yeast Data. *Journal of Biotechnology*, 155(1), 118-126. <https://doi.org/10.1016/j.jbiotec.2010.10.076>

Schulze, I. (2014). Microbial Lipid Production with Oleaginous Yeasts. <https://publikationen.bibliothek.kit.edu/1000043118>

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

Seif, Y., Kavvas, E., Lachance, J.-C., Yurkovich, J. T., Nuccio, S.-P., Fang, X., Catoi, E., Raffatellu, M., Palsson, B. O., & Monk, J. M. (2018). Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. *Nature Communications*, 9(1), 3771. <https://doi.org/10.1038/s41467-018-06112-5>

Shahzadi, T., Mehmood, S., Irshad, M., Anwar, Z., Afroz, A., Zeeshan, nadia, Rashid, U., & Sughra, K. (2014). Advances in lignocellulosic biotechnology: A brief review on lignocellulosic biomass and cellulases. *Advances in Bioscience and Biotechnology*, 05, 246-251. <https://doi.org/10.4236/abb.2014.53031>

Shen, H., Gong, Z., Yang, X., Jin, G., Bai, F., & Zhao, Z. K. (2013). Kinetics of continuous cultivation of the oleaginous yeast *Rhodospiridium toruloides*. *Journal of Biotechnology*, 168(1), 85-89. <https://doi.org/10.1016/j.jbiotec.2013.08.010>

Shen, Q., Chen, Y., Jin, D., Lin, H., Wang, Q., & Zhao, Y.-H. (2016). Comparative genome analysis of the oleaginous yeast *Trichosporon fermentans* reveals its potential applications in lipid accumulation. *Microbiological Research*, 192, 203-210. <https://doi.org/10.1016/j.micres.2016.07.005>

Shi, S., & Zhao, H. (2017). Metabolic Engineering of Oleaginous Yeasts for Production of Fuels and Chemicals. *Frontiers in Microbiology*, 8. <https://doi.org/10.3389/fmicb.2017.02185>

Signori, L., Ami, D., Posterl, R., Giuzzi, A., Mereghetti, P., Porro, D., & Branduardi, P. (2016). Assessing an effective feeding strategy to optimize crude glycerol utilization as sustainable carbon source for lipid accumulation in oleaginous yeasts. *Microbial Cell Factories*, 15(1), 75. <https://doi.org/10.1186/s12934-016-0467-x>

Silva, A., & Fonseca, G. (2018). Screening of an adapted culture medium composed by different carbon sources for heterotrophic cultivation of *Chlorella vulgaris* using a microplate assay. *Acta Scientiarum Biological Sciences*, 40, e39401. <https://doi.org/10.4025/actascibiols.v40i1.39401>

Simm, S., Orrico, A. C. A., Orrico Junior, M. A. P., Sunada, N. da S., Schwingel, A. W., Mendonça Costa, M. S. S. de, Simm, S., Orrico, A. C. A., Orrico Junior, M. A. P., Sunada, N. da S., Schwingel, A. W., & Mendonça Costa, M. S. S. de. (2017). Crude glycerin in anaerobic co-digestion of dairy cattle manure increases methane production. *Scientia Agricola*, 74(3), 175-179. <https://doi.org/10.1590/1678-992x-2016-0057>

Simon, M., Adam, G., Rapatz, W., Spevak, W., & Ruis, H. (1991). The *Saccharomyces cerevisiae* ADR1 gene is a positive regulator of transcription of genes encoding peroxisomal proteins. *Molecular and Cellular Biology*, 11(2), 699-704.

Singh, G., Jawed, A., Paul, D., Bandyopadhyay, K. K., Kumari, A., & Haque, S. (2016). Concomitant Production of Lipids and Carotenoids in *Rhodospiridium toruloides* under Osmotic Stress Using Response Surface Methodology. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.01686>

Singh, R., Kaushik, S., Wang, Y., Xiang, Y., Novak, I., Komatsu, M., Tanaka, K., Cuervo, A. M., & Czaja, M. J. (2009). Autophagy regulates lipid metabolism. *Nature*, 458(7242), 1131-1135. <https://doi.org/10.1038/nature07976>

Sitepu, I. R., Ignatia, L., Franz, A. K., Wong, D. M., Faulina, S. A., Tsui, M., Kanti, A., & Boundy-Mills, K. (2012). An improved high-throughput Nile red fluorescence assay for estimating intracellular lipids in a variety of yeast species. *Journal of Microbiological Methods*, 91(2), 321-328. <https://doi.org/10.1016/j.mimet.2012.09.001>

Sitepu, I. R., Garay, L. A., Sestric, R., Levin, D., Block, D. E., German, J. B., & Boundy-Mills, K. L. (2014a). Oleaginous yeasts for biodiesel: Current and future trends in

biology and production. *Biotechnology Advances*, 32(7), 1336-1360. <https://doi.org/10.1016/j.biotechadv.2014.08.003>

Sitepu, I. R., Jin, M., Fernandez, J. E., da Costa Sousa, L., Balan, V., & Boundy-Mills, K. L. (2014b). Identification of oleaginous yeast strains able to accumulate high intracellular lipids when cultivated in alkaline pretreated corn stover. *Applied microbiology and biotechnology*, 98(17), 7645-7657. <https://doi.org/10.1007/s00253-014-5944-8>

Sitepu, I., Selby, T., Lin, T., Zhu, S., & Boundy-Mills, K. (2014c). Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. *Journal of Industrial Microbiology & Biotechnology*, 41(7), 1061-1070. <https://doi.org/10.1007/s10295-014-1447-y>

Sitepu, I. R., Garay, A. L., Cajka, T., Fiehn, O., & Boundy-Mills, K. L. (2019). Laboratory Screening Protocol to Identify Novel Oleaginous Yeasts. In V. Balan (Éd.), *Microbial Lipid Production: Methods and Protocols* (p. 33-50). Springer. https://doi.org/10.1007/978-1-4939-9484-7_2

Sivasankaran, C., Bharathiraja, Dr. B., Ramanujam, P. K., Varjani, S., Mani, D., & Elavazhagan, S. (2018). Utilization of Crude Glycerol from Biodiesel Industry for the Production of Value-Added Bioproducts. https://doi.org/10.1007/978-981-10-7431-8_4

Sivasankaran, C., Ramanujam, P. K., Balasubramanian, B., & Mani, J. (2019). Recent progress on transforming crude glycerol into high value chemicals: A critical review. *Biofuels*, 10(3), 309-314. <https://doi.org/10.1080/17597269.2016.1174018>

Solaiman, D. K. Y., Ashby, R. D., Foglia, T. A., & Marmer, W. N. (2006). Conversion of agricultural feedstock and coproducts into poly(hydroxyalkanoates). *Applied Microbiology and Biotechnology*, 71(6), 783-789. <https://doi.org/10.1007/s00253-006-0451-1>

Song, G., Dickins, B. J. A., Demeter, J., Engel, S., Dunn, B., & Cherry, J. M. (2015). AGAPE (Automated Genome Analysis PipelinE) for Pan-Genome Analysis of *Saccharomyces cerevisiae*. *PLOS ONE*, 10(3), e0120671. <https://doi.org/10.1371/journal.pone.0120671>

Sprouffs, K., & Wagner, A. (2016). Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, 17(1), 172. <https://doi.org/10.1186/s12859-016-1016-7>

Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33(Web Server issue), W465-W467. <https://doi.org/10.1093/nar/gki458>

Steensels, J., Snoek, T., Meersman, E., Nicolino, M. P., Voordeckers, K., & Verstrepen, K. J. (2014). Improving industrial yeast strains: Exploiting natural and artificial diversity. *FEMS Microbiology Reviews*, 38(5), 947-995. <https://doi.org/10.1111/1574-6976.12073>

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026-1028. <https://doi.org/10.1038/nbt.3988>

Sudiyani, Y., Styarini, D., Triwahyuni, E., Sudiyarmanto, Sembiring, K. C., Aristiawan, Y., Abimanyu, H., & Han, M. H. (2013). Utilization of Biomass Waste Empty Fruit Bunch Fiber of Palm Oil for Bioethanol Production Using Pilot-Scale Unit. *Energy Procedia*, 32, 31-38. <https://doi.org/10.1016/j.egypro.2013.05.005>

Sun, Z., Harris, H. M. B., McCann, A., Guo, C., Argimón, S., Zhang, W., Yang, X., Jeffery, I. B., Cooney, J. C., Kagawa, T. F., Liu, W., Song, Y., Salvetti, E., Wrobel, A., Rasinkangas, P., Parkhill, J., Rea, M. C., O'Sullivan, O., Ritari, J., ... O'Toole, P. W. (2015). Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nature Communications*, 6, 8322. <https://doi.org/10.1038/ncomms9322>

Sutherland, F. C., Lages, F., Lucas, C., Luyten, K., Albertyn, J., Hohmann, S., Prior, B. A., & Kilian, S. G. (1997). Characteristics of Fps1-dependent and -independent glycerol transport in *Saccharomyces cerevisiae*. *Journal of Bacteriology*, 179(24), 7790-7795.

Swinnen, S., Ho, P.-W., Klein, M., & Nevoigt, E. (2016). Genetic determinants for enhanced glycerol growth of *Saccharomyces cerevisiae*. *Metabolic Engineering*, 36, 68-79. <https://doi.org/10.1016/j.ymben.2016.03.003>

Szymanowska-Powałowska, D., & Białas, W. (2014). Scale-up of anaerobic 1,3-propanediol production by *Clostridium butyricum* DSP1 from crude glycerol. *BMC Microbiology*, 14(1), 45. <https://doi.org/10.1186/1471-2180-14-45>

Tachibana, C., Yoo, J. Y., Tagne, J.-B., Kacherovsky, N., Lee, T. I., & Young, E. T. (2005). Combined global localization analysis and transcriptome data identify genes that

are directly coregulated by Adr1 and Cat8. *Molecular and Cellular Biology*, 25(6), 2138-2146. <https://doi.org/10.1128/MCB.25.6.2138-2146.2005>

Tan, H. W., Abdul Aziz, A. R., & Aroua, M. K. (2013). Glycerol production and its applications as a raw material: A review. *Renewable and Sustainable Energy Reviews*, 27, 118-127. <https://doi.org/10.1016/j.rser.2013.06.035>

Tao, B. Y. (2007). Chapter 24—Industrial Applications for Plant Oils and Lipids. In S.-T. Yang (Éd.), *Bioprocessing for Value-Added Products from Renewable Resources* (p. 611-627). Elsevier. <https://doi.org/10.1016/B978-044452114-9/50025-6>

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., & Natale, D. A. (2003). The COG database: An updated version includes eukaryotes. *BMC Bioinformatics*, 4(1), 41. <https://doi.org/10.1186/1471-2105-4-41>

Telford, J. C., Yeung, J. H. F., Xu, G., Kiefel, M. J., Watts, A. G., Hader, S., Chan, J., Bennet, A. J., Moore, M. M., & Taylor, G. L. (2011). The *Aspergillus fumigatus* sialidase is a 3-deoxy-D-glycero-D-galacto-2-nonulosonic acid hydrolase (KDNase): Structural and mechanistic insights. *The Journal of Biological Chemistry*, 286(12), 10783-10792. <https://doi.org/10.1074/jbc.M110.207043>

Teste, M.-A., François, J. M., & Parrou, J.-L. (2010). Characterization of a new multigene family encoding isomaltases in the yeast *Saccharomyces cerevisiae*, the IMA family. *The Journal of Biological Chemistry*, 285(35), 26815-26824. <https://doi.org/10.1074/jbc.M110.145946>

Thevenieau, F., & Nicaud, J.-M. (2013). Microorganisms as sources of oils. *OCL*, 20(6), D603. <https://doi.org/10.1051/ocl/2013034>

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1), 93-121. <https://doi.org/10.1038/nprot.2009.203>

Thompson, J. C., & He, B. B. (2006). Characterization of crude glycerol from biodiesel production from multiple feedstocks. *Applied Engineering in Agriculture*, 22(2), 261-265. Scopus.

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

Unrean, P., & Champreda, V. (2017). High-Throughput Screening and Dual Feeding Fed-Batch Strategy for Enhanced Single-Cell Oil Accumulation in *Yarrowia lipolytica*. *BioEnergy Research*, 10(4), 1057-1065. <https://doi.org/10.1007/s12155-017-9865-0>

Valencia, A. (2005). Automatic annotation of protein function. *Current Opinion in Structural Biology*, 15(3), 267-274. <https://doi.org/10.1016/j.sbi.2005.05.010>

Valerio, O., Horvath, T., Pond, C., Manjusri Misra, & Mohanty, A. (2015). Improved utilization of crude glycerol from biodiesel industries : Synthesis and characterization of sustainable biobased polyesters. *Industrial Crops and Products*, 78, 141-147. <https://doi.org/10.1016/j.indcrop.2015.10.019>

Venkataramanan, K. P., Boatman, J. J., Kurniawan, Y., Taconi, K. A., Bothun, G. D., & Scholz, C. (2012). Impact of impurities in biodiesel-derived crude glycerol on the fermentation by *Clostridium pasteurianum* ATCC 6013. *Applied Microbiology and Biotechnology*, 93(3), 1325-1335. <https://doi.org/10.1007/s00253-011-3766-5>

Venters, B. J., Wachi, S., Mavrich, T. N., Andersen, B. E., Jena, P., Sinnamon, A. J., Jain, P., Roller, N. S., Jiang, C., Hemeryck-Walsh, C., & Pugh, B. F. (2011). A comprehensive genomic binding map of gene and chromatin regulatory proteins in *Saccharomyces*. *Molecular Cell*, 41(4), 480-492. <https://doi.org/10.1016/j.molcel.2011.01.015>

Vieira, E. R. (2013). *Elementary Food Science*. Springer Science & Business Media.

Viigand, K., Põšnograjeva, K., Visnapuu, T., & Alamäe, T. (2018). Genome Mining of Non-Conventional Yeasts : Search and Analysis of MAL Clusters and Proteins. *Genes*, 9(7). <https://doi.org/10.3390/genes9070354>

Vuralhan, Z., Morais, M. A., Tai, S.-L., Piper, M. D. W., & Pronk, J. T. (2003). Identification and Characterization of Phenylpyruvate Decarboxylase Genes in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, 69(8), 4534-4541. <https://doi.org/10.1128/AEM.69.8.4534-4541.2003>

Wadekar, S. D., Kale, S. B., Lali, A. M., Bhowmick, D. N., & Pratap, A. P. (2012). Utilization of sweetwater as a cost-effective carbon source for sophorolipids production by *Starmerella bombicola* (ATCC 22214). *Preparative Biochemistry & Biotechnology*, 42(2), 125-142. <https://doi.org/10.1080/10826068.2011.577883>

Wang, Q., Guo, F.-J., Rong, Y.-J., & Chi, Z.-M. (2012). Lipid production from hydrolysate of cassava starch by *Rhodospiridium toruloides* 21167 for biodiesel making. *Renewable Energy*, 46(Supplement C), 164-168. <https://doi.org/10.1016/j.renene.2012.03.002>

Wang, X., Xia, K., Yang, X., & Tang, C. (2019). Growth strategy of microbes on mixed carbon sources. *Nature Communications*, 10(1), 1279. <https://doi.org/10.1038/s41467-019-09261-3>

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

Warringer, J., Zörgö, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., Forsmark, A., Durbin, R., Omholt, S. W., Louis, E. J., Liti, G., Moses, A., & Blomberg, A. (2011). Trait Variation in Yeast Is Defined by Population History. *PLoS Genetics*, 7(6), Article 6. <https://doi.org/10.1371/journal.pgen.1002111>

Wen, Z., Zhang, S., Odoh, C. K., Jin, M., & Zhao, Z. K. (2020). *Rhodospiridium toruloides*—A potential red yeast chassis for lipids and beyond. *FEMS Yeast Research*, 20(5), Article 5. <https://doi.org/10.1093/femsyr/foaa038>

Werpy, T., & Petersen, G. (2004). *Top Value Added Chemicals from Biomass : Volume I -- Results of Screening for Potential Candidates from Sugars and Synthesis Gas (DOE/GO-102004-1992)*. National Renewable Energy Lab., Golden, CO (US). <https://doi.org/10.2172/15008859>

Westbrook, A. W., Miscevic, D., Kilpatrick, S., Bruder, M. R., Moo-Young, M., & Chou, C. P. (2018). Strain engineering for microbial production of value-added chemicals and fuels from glycerol. *Biotechnology Advances*. <https://doi.org/10.1016/j.biotechadv.2018.10.006>

Wicke, B., Sikkema, R., Dornburg, V., & Faaij, A. P. C. (2011). Exploring land use changes and the role of palm oil production in Indonesia and Malaysia. *Land Use Policy*, 28, 193-206. <https://doi.org/10.1016/j.landusepol.2010.06.001>

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

Williams, T. J., Allen, M., Tschitschko, B., & Cavicchioli, R. (2017). Glycerol metabolism of haloarchaea. *Environmental Microbiology*, 19(3), 864-877. <https://doi.org/10.1111/1462-2920.13580>

Willke, T., & Vorlop, K. D. (2001). Biotechnological production of itaconic acid. *Applied Microbiology and Biotechnology*, 56(3-4), 289-295. <https://doi.org/10.1007/s002530100685>

Wolfe, K. H. (2006). Comparative genomics and genome evolution in yeasts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1467), 403-412. <https://doi.org/10.1098/rstb.2005.1799>

Wrede, D., Taha, M., Miranda, A. F., Kadali, K., Stevenson, T., Ball, A. S., & Mouradov, A. (2014). Co-Cultivation of Fungal and Microalgal Cells as an Efficient System for Harvesting Microalgal Cells, Lipid Production and Wastewater Treatment. *PLOS ONE*, 9(11), e113497. <https://doi.org/10.1371/journal.pone.0113497>

Wu, S., Hu, C., Jin, G., Zhao, X., & Zhao, Z. K. (2010). Phosphate-limitation mediated lipid production by *Rhodospiridium toruloides*. *Bioresource Technology*, 101(15), 6124-6129. <https://doi.org/10.1016/j.biortech.2010.02.111>

Wu, S., Zhao, X., Shen, H., Wang, Q., & Zhao, Z. K. (2011). Microbial lipid production by *Rhodospiridium toruloides* under sulfate-limited conditions. *Bioresource Technology*, 102(2), 1803-1807. <https://doi.org/10.1016/j.biortech.2010.09.033>

Wynn, J. P., & Ratledge, C. (2005). Oils from Microorganisms. In *Bailey's Industrial Oil and Fat Products*. American Cancer Society. <https://doi.org/10.1002/047167849X.bio006>

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C.-Y., & Wei, L. (2011). KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, 39(Web Server issue), W316-W322. <https://doi.org/10.1093/nar/gkr483>

Xin, B., Wang, Y., Tao, F., Li, L., Ma, C., & Xu, P. (2016). Co-utilization of glycerol and lignocellulosic hydrolysates enhances anaerobic 1,3-propanediol production by *Clostridium diolis*. *Scientific Reports*, 6, 19044. <https://doi.org/10.1038/srep19044>

Xu, J., Zhao, X., Wang, W., Du, W., & Liu, D. (2012). Microbial conversion of biodiesel byproduct glycerol to triacylglycerols by oleaginous yeast *Rhodospiridium toruloides*

and the individual effect of some impurities on lipid production. *Biochemical Engineering Journal*, 65, 30-36. <https://doi.org/10.1016/j.bej.2012.04.003>

Yamada, R., Kashihara, T., & Ogino, H. (2017). Improvement of lipid production by the oleaginous yeast *Rhodospiridium toruloides* through UV mutagenesis. *World Journal of Microbiology and Biotechnology*, 33(5), 99. <https://doi.org/10.1007/s11274-017-2269-7>

Yan, Q., & Pflieger, B. F. (2020). Revisiting metabolic engineering strategies for microbial synthesis of oleochemicals. *Metabolic Engineering*, 58, 35-46. <https://doi.org/10.1016/j.ymben.2019.04.009>

Yandell, M., & Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics*, 13(5), 329-342. <https://doi.org/10.1038/nrg3174>

Yang, F., Hanna, M. A., & Sun, R. (2012). Value-added uses for crude glycerol—A byproduct of biodiesel production. *Biotechnology for Biofuels*, 5. <https://doi.org/10.1186/1754-6834-5-13>

Yazdani, S. S., & Gonzalez, R. (2007). Anaerobic fermentation of glycerol: A path to economic viability for the biofuels industry. *Current Opinion in Biotechnology*, 18(3), 213-219. <https://doi.org/10.1016/j.copbio.2007.05.002>

Yen, H.-W., Liu, Y. X., & Chang, J.-S. (2015). The effects of feeding criteria on the growth of oleaginous yeast—*Rhodotorula glutinis* in a pilot-scale airlift bioreactor. *Journal of the Taiwan Institute of Chemical Engineers*, 49, 67-71. <https://doi.org/10.1016/j.jtice.2014.11.019>

Yeung, J. (2015). Sialic acid metabolism in the opportunistic fungal pathogen, *Aspergillus fumigatus* [Thesis, Science: Biological Sciences Department]. <https://summit.sfu.ca/item/16993>

Yong, K. C., Ooi, T. L., Dzulkefly, K., Yunus, W., & Hassan, H. (2001). Characterization of glycerol residue from a palm kernel oil methyl ester plant. *J. Oil Palm Res.*, 13, 1-6.

Yoon, S. H., & Rhee, J. S. (1983). Lipid from yeast fermentation: Effects of cultural conditions on lipid production and its characteristics of *rhodotorula glutinis*. *Journal of the American Oil Chemists' Society*, 60(7), 1281-1286. <https://doi.org/10.1007/BF02702101>

Yoon, S. A., & Kim, H. K. (2013). Development of a bioconversion system using *Saccharomyces cerevisiae* Reductase YOR120W and *Bacillus subtilis* glucose dehydrogenase for chiral alcohol synthesis. *Journal of Microbiology and Biotechnology*, 23(10), 1395-1402. <https://doi.org/10.4014/jmb.1305.05030>

Yoshino, M., & Murakami, K. (1982). AMP deaminase as a control system of glycolysis in yeast. Mechanism of the inhibition of glycolysis by fatty acid and citrate. *The Journal of Biological Chemistry*, 257(18), 10644-10649.

Young, E. T., Dombek, K. M., Tachibana, C., & Ideker, T. (2003). Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *The Journal of Biological Chemistry*, 278(28), 26146-26158. <https://doi.org/10.1074/jbc.M301981200>

Yu, A.-Q., Pratomo Juwono, N. K., Leong, S. S. J., & Chang, M. W. (2014). Production of Fatty Acid-Derived Valuable Chemicals in Synthetic Microbes. *Frontiers in Bioengineering and Biotechnology*, 2. <https://doi.org/10.3389/fbioe.2014.00078>

Yurimoto, H., & Sakai, Y. (2019). Methylotrophic Yeasts : Current Understanding of Their C1-Metabolism and its Regulation by Sensing Methanol for Survival on Plant Leaves. *Current Issues in Molecular Biology*, 33, 197-210. <https://doi.org/10.21775/cimb.033.197>

Zainuddin, M. F., Fai, C. K., Ariff, A. B., Rios-Solis, L., & Halim, M. (2021). Current Pretreatment/Cell Disruption and Extraction Methods Used to Improve Intracellular Lipid Recovery from Oleaginous Yeasts. *Microorganisms*, 9(2), 251. <https://doi.org/10.3390/microorganisms9020251>

Zaky, A. S., Greetham, D., Tucker, G. A., & Du, C. (2018). The establishment of a marine focused biorefinery for bioethanol production using seawater and a novel marine yeast strain. *Scientific Reports*, 8(1), 12127. <https://doi.org/10.1038/s41598-018-30660-x>

Zambanini, T., Hosseinpour Tehrani, H., Geiser, E., Merker, D., Schleese, S., Krabbe, J., Buescher, J. M., Meurer, G., Wierckx, N., & Blank, L. M. (2017). Efficient itaconic acid production from glycerol with *Ustilago vetiveriae* TZ1. *Biotechnology for Biofuels*, 10(1), 131. <https://doi.org/10.1186/s13068-017-0809-x>

Zhang, M. Q. (2002). Computational prediction of eukaryotic protein-coding genes. *Nature Reviews Genetics*, 3(9), 698-709. <https://doi.org/10.1038/nrg890>

Zhang, X. L., Yan, S., Tyagi, R. D., & Surampalli, R. Y. (2013). Biodiesel production from heterotrophic microalgae through transesterification and nanotechnology application in the production. *Renewable and Sustainable Energy Reviews*, 26, 216-223. <https://doi.org/10.1016/j.rser.2013.05.061>

Zhang, H., Zhang, L., Chen, H., Chen, Y. Q., Chen, W., Song, Y., & Ratledge, C. (2014). Enhanced lipid accumulation in the yeast *Yarrowia lipolytica* by over-expression of ATP:citrate lyase from *Mus musculus*. *Journal of Biotechnology*, 192 Pt A, 78-84. <https://doi.org/10.1016/j.jbiotec.2014.10.004>

Zhang, S., Ito, M., Skerker, J. M., Arkin, A. P., & Rao, C. V. (2016a). Metabolic engineering of the oleaginous yeast *Rhodospiridium 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. (2016b). Engineering *Rhodospiridium toruloides* for increased lipid production. *Biotechnology and Bioengineering*, 113(5), 1056-1066. <https://doi.org/10.1002/bit.25864>

Zhang, Y.-H. P., Sun, J., & Ma, Y. (2017). Biomufacturing : History and perspective. *Journal of Industrial Microbiology and Biotechnology*, 44(4-5), 773-784. <https://doi.org/10.1007/s10295-016-1863-2>

Zhao, X., Wu, S., Hu, C., Wang, Q., Hua, Y., & Zhao, Z. K. (2010). Lipid production from Jerusalem artichoke by *Rhodospiridium toruloides* Y4. *Journal of Industrial Microbiology & Biotechnology*, 37(6), 581-585. <https://doi.org/10.1007/s10295-010-0704-y>

Zhao, X., Hu, C., Wu, S., Shen, H., & Zhao, Z. K. (2011). Lipid production by *Rhodospiridium toruloides* Y4 using different substrate feeding strategies. *Journal of Industrial Microbiology & Biotechnology*, 38(5), 627-632. <https://doi.org/10.1007/s10295-010-0808-4>

Zhao, X., Peng, F., Du, W., Liu, C., & Liu, D. (2012). Effects of some inhibitors on the growth and lipid accumulation of oleaginous yeast *Rhodospiridium toruloides* and preparation of biodiesel by enzymatic transesterification of the lipid. *Bioprocess and Biosystems Engineering*, 35(6), 993-1004. <https://doi.org/10.1007/s00449-012-0684-6>

Zhao, C., Luo, M.-T., Huang, C., Chen, X.-F., Xiong, L., Li, H.-L., & Chen, X.-D. (2019). Determining intracellular lipid content of different oleaginous yeasts by one

simple and accurate Nile Red fluorescent method. *Preparative Biochemistry & Biotechnology*, 49(6), 597-605. <https://doi.org/10.1080/10826068.2019.1587624>

Zheng, W., Zhao, H., Mancera, E., Steinmetz, L. M., & Snyder, M. (2010). Genetic analysis of variation in transcription factor binding in yeast. *Nature*, 464(7292), 1187-1191. <https://doi.org/10.1038/nature08934>

Zhu, Z., Zhang, S., Liu, H., Shen, H., Lin, X., Yang, F., Zhou, Y. J., Jin, G., Ye, M., Zou, H., Zou, H., & Zhao, Z. K. (2012). A multi-omic map of the lipid-producing yeast *Rhodospiridium toruloides*. *Nature Communications*, 3, 1112. <https://doi.org/10.1038/ncomms2112>

Zhu, C., Chiu, S., Nakas, J., & Nomura, C. (2013). Bioplastics from waste glycerol derived from biodiesel industry. *Journal of Applied Polymer Science*, 130. <https://doi.org/10.1002/app.39157>

Zhu, Z., Ding, Y., Gong, Z., Yang, L., Zhang, S., Zhang, C., Lin, X., Shen, H., Zou, H., Xie, Z., Yang, F., Zhao, X., Liu, P., & Zhao, Z. K. (2015). Dynamics of the Lipid Droplet Proteome of the Oleaginous Yeast *Rhodospiridium toruloides*. *Eukaryotic Cell*, 14(3), 252-264. <https://doi.org/10.1128/EC.00141-14>