EDUARDO LUÍS MENEZES DE ALMEIDA

ASSESSMENT OF METABOLIC REGULATION AND STRESS RESPONSE IN OLEAGINOUS YEASTS BY SYSTEMS BIOLOGY APPROACHES

Thesis submitted to the Agricultural Microbiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Advisor: Wendel Batista da Silveira

Co-advisor: Marcelo Mendes Brandão

VIÇOSA - MINAS GERAIS

2024

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
A447a
2024

Almeida, Eduardo Luís Menezes de, 1997-
Assessment of metabolic regulation and stress response in oleaginous yeasts by systems biology approaches / Eduardo Luís Menezes de Almeida. – Viçosa, MG, 2024.
1 tese eletrônica (189 f.): il. (algumas color.).

Texto em inglês.
Orientador: Wendel Batista da Silveira.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Microbiologia, 2024.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2024.019
Modo de acesso: World Wide Web.

1. Leveduras. 2. Metabolismo - Regulação. 3. Perfilação da expressão gênica. I. Silveira, Wendel Batista da, 1979-. II. Universidade Federal de Viçosa. Departamento de Microbiologia. Programa de Pós-Graduação em Microbiologia Agrícola. III. Título.

CDD 22. ed. 579.562

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

EDUARDO LUÍS MENEZES DE ALMEIDA

ASSESSMENT OF METABOLIC REGULATION AND STRESS RESPONSE IN OLEAGINOUS YEASTS BY SYSTEMS BIOLOGY APPROACHES

Thesis submitted to the Agricultural Microbiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: 30 de janeiro de 2024

gov.br Documento assinado digitalmente
EDUARDO LUIS MENEZES DE ALMEIDA
Data: 31/01/2024 11:00:05-0300
Verifique em https://validar.iti.gov.br

Eduardo Luís Menezes de Almeida

Author

gov.br Documento assinado digitalmente
WENDEL BATISTA DA SILVEIRA
Data: 31/01/2024 12:44:42-0300
Verifique em https://validar.iti.gov.br

Wendel Batista da Silveira

Advisor

A todos que fizeram possível esse caminho, especialmente à minha família.

ACKNOWLEDGEMENTS

À Universidade Federal de Viçosa, pela excelência em ensino, pesquisa e extensão.

Ao Programa de Pós-Graduação em Microbiologia Agrícola e ao Departamento de Microbiologia, pela oportunidade de realização do curso de doutorado.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for granting the scholarship - Finance Code 140538/2021-6.

To the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for the financial suport - Finance Code APQ-00326-21.

To Professor Wendel Batista da Silveira for his supervision.

To Professor Marcelo Mendes Brandão for his co-supervision.

To my lab colleagues, for their discussions, comments, and contributions.

To Professor Monique Renon Eller for the guidance.

To my friends, for their patience and guidance.

To my family and my partner, Thaynara Lorenzoni Entringer, for all their support.

“Through others, we become ourselves.”

Lev Vygotsky

ABSTRACT

ALMEIDA, Eduardo Luís Menezes de, D.Sc., Universidade Federal de Viçosa, January, 2024. **Assessment of metabolic regulation and stress response in oleaginous yeasts by systems biology approaches**. Advisor: Wendel Batista da Silveira. Co-advisor: Marcelo Mendes Brandão.

The current environmental crises promoted by the extensive use of oil have highlighted the necessity of alternative and sustainable production chains. Lipids and fatty acids from oleaginous yeasts produced from lignocellulosic hydrolysates are promising sources for oleochemicals. Hence, selecting and developing robust yeast strains that can grow and produce lipids in lignocellulosic hydrolysates is pivotal to broaden their applicability and viability. *Lipomyces starkeyi* is an oleaginous yeast capable of growing and producing lipids using a diverse range of carbon sources. Its growth and lipid production have been demonstrated in lignocellulosic biomasses. *Papiliotrema laurentii* can also assimilate sugars derived from agricultural wastes, such as glucose and xylose from lignocellulosic biomasses, and convert them into high lipid amounts. However, wild *P. laurentii* strains are highly sensitive to acetic acid, one of the primary inhibitors in lignocellulosic hydrolysates. The understanding of multifactorial stress responses and yeast physiology can be facilitated by the application of holistic approaches, including biological networks and high-throughput data. Therefore, we hypothesized that the use of systems biology approaches, including genome-scale metabolic models (GEMs) and transcriptomics, as well as their integration, can support us to understand the metabolism and stress responses of oleaginous yeasts and identify suitable targets for metabolic engineering. Herein, we propose the first GEM of *L. starkeyi, lista*-GEM. We reconstructed *lista*-GEM using two high-quality oleaginous yeast models as templates and curated it to reflect the metabolism of *L. starkeyi*. The simulated phenotypes and predicted flux distributions were in good accordance with experimental data. Then, we predicted targets to improve lipid production in glucose, xylose, and glycerol. Enzymes related to lipid synthesis in the endoplasmic reticulum, such as stearoyl-CoA desaturase, fatty-acyl-CoA synthase, diacylglycerol acyltransferase, and glycerol-3-phosphate acyltransferase, were the main targets to improve lipid production. Glycolytic genes were also predicted as targets for overexpression. Pyruvate decarboxylase, acetaldehyde dehydrogenase, acetyl-CoA synthetase, adenylate kinase, inorganic diphosphatase, and triose-phosphate isomerase were predicted only when glycerol was the carbon source. Hence, *lista*-GEM provides multiple metabolic engineering

targets to improve lipid production by *L. starkeyi* using carbon sources from agricultural and industrial wastes. Furthermore, we combined transcriptome and genome-scale metabolic modeling to deepen our understanding regarding the targets of acetic acid stress, as well as the adaptive responses in *P. laurentii*. Acetic acid stress promoted global expression changes and most repressed genes were related to transcriptional and translational processes. Under stress, the sensitive strain induced DNA mismatch repair mechanisms and meiosis, while the tolerant strain negatively regulated autophagy and the cell cycle. The tolerant strain induced processes responsible for increasing the intracellular pH (e.g., arginase, ornithine metabolism, urea cycle), detoxification of toxic compounds (e.g., glutathione metabolism), and proton efflux. The tolerant strain also presented a remarkable NAD(P)H pool in the metabolic modeling analysis, which might support the reducing power required by tolerance mechanisms. Otherwise, the sensitive strain induced genes related to cell wall biogenesis and cobalamin synthesis. Overall, the genes and pathways described herein as tolerant-related can be useful in future metabolic engineering strategies to improve the tolerance of *P. laurentii* to weak acids, boosting its application in lignocellulosic-based biorefineries.

Keywords: Metabolic modeling; Transcriptomics; Non-*Saccharomyces*.

RESUMO

ALMEIDA, Eduardo Luís Menezes de, D.Sc., Universidade Federal de Viçosa, January, 2024. **Avaliação da regulação metabólica e da resposta ao stress de leveduras oleaginosas através da biologia de sistemas**. Orientador: Wendel Batista da Silveira. Coorientador: Marcelo Mendes Brandão.

As atuais crises ambientais promovidas pelo uso extensivo de petróleo têm realçado a necessidade de cadeias produtivas sustentáveis. Lipídios e ácidos graxos produzidos a partir de hidrolisados lignocelulósicos por leveduras oleaginosas são fontes promissoras de oleoquímicos. Portanto, a seleção e construção de linhagens robustas de leveduras capazes de crescer e produzir lipídios em hidrolisados lignocelulósicos é essencial para ampliar sua aplicação e viabilidade. *Lipomyces starkeyi* é uma levedura oleaginosa que cresce e produz lipídios a partir de diversas fontes de carbono. O seu crescimento e produção de lipídios já foram demonstrados em biomassas lignocelulósicas. *Papiliotrema laurentii* também pode assimilar açúcares derivados de subprodutos agrícolas, como a glucose e a xilose de biomassas lignocelulósicas, e convertê-los em elevadas concentrações de lipídios. No entanto, linhagens selvagens de *P. laurentii* são sensíveis ao ácido acético, um dos principais inibidores presentes em hidrolisados lignocelulósicos. A compreensão das respostas adaptativas e da fisiologia dessas leveduras pode ser facilitada por abordagens holísticas, incluindo redes biológicas e ômicas. Portanto, a nossa hipótese é que o emprego de abordagens de biologia de sistemas, incluindo modelos metabólicos em escala genômica (MMEG) e transcriptômica, bem como a sua integração, pode auxiliar na compreensão do metabolismo e das respostas ao estresse de leveduras oleaginosas e determinação de alvos para a engenharia metabólica. Neste trabalho, propomos o primeiro MMEG de *L. starkeyi*, *lista*-GEM. Reconstruímos o *lista*-GEM usando dois modelos de leveduras oleaginosas de alta qualidade como templates e conduzimos a curadoria manual para refletir o metabolismo de *L. starkeyi.* Os fenótipos simulados e as distribuições de fluxo previstas apresentaram boa conformidade com os dados experimentais. Em seguida, prevemos alvos para melhorar a produção de lípidos em glicose, xilose e glicerol. As enzimas relacionadas com a síntese de lípidos no retículo endoplasmático, como estearoil-CoA dessaturase, acil-CoA sintase, diacilglicerol aciltransferase e glicerol-3-fosfato aciltransferase, foram os principais alvos para melhorar a produção de lípidos. Genes glicolíticos também foram previstos como alvos de superexpressão. Piruvato descarboxilase,

acetaldeído desidrogenase, acetil-CoA sintetase, adenilato quinase, difosfatase inorgânica e triose-fosfato isomerase foram previstas apenas quando o glicerol era a fonte de carbono. Dessa forma, o *lista*-GEM fornece múltiplos alvos de engenharia metabólica para melhorar a produção de lípidos por *L. starkeyi* a partir de resíduos agrícolas e industriais. Em seguida, combinamos o transcriptoma e a modelação metabólica em escala genômica para aprofundar o nosso conhecimento sobre os alvos do estresse por ácido acético, bem como sobre as respostas adaptativas de *P. laurentii*. O ácido acético promoveu alterações globais de expressão e a maioria dos genes reprimidos estava relacionada com processos transcricionais e translacionais. Sob estresse, a linhagem sensível induziu mecanismos de reparo do DNA e meiose, enquanto a linhagem tolerante regulou negativamente a autofagia e o ciclo celular. A linhagem tolerante induziu também processos responsáveis pelo aumento do pH intracelular (ex. arginase, metabolismo da ornitina, ciclo da ureia), detoxificação (ex. metabolismo da glutationa) e efluxo de prótons. A linhagem tolerante também apresentou uma reserva notável de NAD(P)H na análise de modelagem metabólica, o que suportaria o poder redutor exigido pelos mecanismos de tolerância. Por outro lado, a estirpe sensível induziu genes relacionados com a biogênese da parede celular e a síntese de cobalamina. Em geral, os genes e vias aqui descritos como relacionados com a tolerância podem ser úteis em futuras estratégias de engenharia metabólica para melhorar a tolerância de *P. laurentii* a ácidos fracos, impulsionando a sua aplicação em biorrefinarias baseadas em lignocelulose.

Palavras-chave: Modelagem metabólica; Transcriptômica; Não-*Saccharomyces*.

SUMMARY

INTRODUCTION

The current environmental crises promoted by the extensive use of oil derivatives (*e.g.,* fuels and chemicals) and emission of greenhouse gasses (GHG) have highlighted the necessity of alternative and more sustainable production chains using agricultural and industrial byproducts (CHERUBINI, 2010; YAMAKAWA et al., 2020). Converting these feedstocks into valuable products using yeasts is a promising strategy. Yeasts have been long used in bioprocessing, especially *Saccharomyces cerevisiae,* due to their tolerance to ethanol, capacity to ferment high concentrations of hexose, and Generally Regarded as Safe (GRAS) status (NIELSEN, 2019). However, wild *S. cerevisiae* strains cannot ferment some sugars abundant in agricultural byproducts, such as xylose from lignocellulosic biomasses, or industrial wastes, such as lactose from whey derivatives. They also cannot produce high contents of lipids, impairing their use as oil sources for biodiesel and oleochemicals (GEIJER; LEDESMA-AMARO; TOMAS-PEJO, 2022; NAVARRETE; L. MARTÍNEZ, 2020). These drawbacks require expansive and time-demanding development of engineered strains. Hence, using other organisms that are naturally more suitable for the processes described above is imperative. Non-conventional yeasts, that is, yeasts that are not *Saccharomyces,* have been extensively evaluated for bioprocessing, including *Yarrowia lipolytica, Rhodotorula toruloides, Kluyveromyces marxianus, Kluyveromyces lactis, Komagataella phaffii* (previously *Pichia pastoris*), and *Ogataea polymorpha*.

Lipomyces starkeyi is an oleaginous yeast capable of growing and producing lipids using a diverse range of carbon sources, such as glucose, galactose, arabinose, xylose, glycerol, mannose, cellobiose, and sucrose (SMITH; KURTZMAN, 2011). Its growth and lipid production have been demonstrated in lignocellulosic biomasses, including corn stover (POMRANING et al., 2019), wheat straw (YU et al., 2011), lignin derivatives (PUTRA et al., 2023); glycerol (LIU et al., 2017; MARUYAMA et al., 2018); and sewage sludge (ANGERBAUER et al., 2008). Importantly, *L. starkeyi* tolerates inhibitors found in lignocellulosic hydrolysates, such as hydroxymethylfurfural (HMF), furfural, and phenolic compounds (PUTRA et al., 2023; RAHMAN et al., 2017). Besides, *L. starkeyi* can use levoglucosan, a major product from lignocellulose pyrolysis, as a carbon source (NING et al., 2008).

Moreover, given the potential of non-conventional yeasts for bioprocessing, our research team has isolated oleaginous yeasts capable of assimilating sugars derived from agricultural wastes, such as xylose from lignocellulosic biomasses, and converting them into high lipid contents, including *Papiliotrema laurentii* (VIEIRA et al., 2020). Thus, we have previously selected acetic acid tolerant strains (ATS) by adaptive laboratory evolution (ALE), and one strain stood out (ATS) due to its tolerance and presence of the oleaginous phenotype in all conditions tested (ALMEIDA et al., 2023). We have also recently reconstructed the genome-scale metabolic model (GEM) of *P. laurentii* (*papla*-GEM), an excellent tool for understanding its physiology and determining metabolic engineering targets (VENTORIM et al., 2022).

Furthermore, understanding the stress response and physiology of these yeasts is crucial for developing robust strains and making feasible the lipid production by oleaginous yeasts in lignocellulosic biorefineries (THORWALL et al., 2020). Systems biology approaches are suitable for holistically studying metabolism and the underlying mechanisms and complex interactions in stressed conditions. The reconstruction of biological networks, mathematical models, and high-throughput data (omics data) are the cornerstone of system biology (NIELSEN, 2017). GEMs combine an organism's genome and metabolic landscape information into a comprehensive metabolic map. These models are a great framework not only for performing simulations to predict extra and intracellular fluxes but also for the integration of omics data, such as transcriptomics (PALSSON, 2009). Therefore, we hypothesized that the use of systems biology approaches, including GEMs and transcriptomics, as well as their integration, can help us understand the metabolism and stress responses of oleaginous yeasts and determine suitable targets for metabolic engineering.

This thesis is organized in four chapters. In the first chapter, we present a literature review on the reconstruction and validations of GEMs of non-*Saccharomyces* yeasts, underscoring the physiological insights, as well as the identification of both metabolic engineering and drug targets. We also discuss the challenges and knowledge gaps and propose strategies to boost their use and novel reconstructions. In the second chapter, we propose *lista*-GEM, the first genome-scale metabolic model of *L. starkeyi*. We show that *lista*-GEM successfully captures the growth and lipid-producing phenotype of *L. starkeyi* and, therefore, is a useful platform for *in silico* metabolic engineering of this yeast. The third chapter comprises a literature review of the current knowledge about the physiology, genetics, ecology, and biotechnological applications of *P. laurentii*. In the fourth chapter, we combine transcriptome

and genome-scale metabolic modeling to deepen our understanding regarding the targets of acetic acid stress, as well as the adaptive responses in *P. laurentii*. We describe genes and pathways that can be used in future genetic and metabolic engineering rounds to improve the tolerance of this yeast to weak acids and boost its application in lignocellulosic-based biorefineries.

References

ALMEIDA, E. L. M. et al. New *Papiliotrema laurentii* UFV-1 strains with improved acetic acid tolerance selected by adaptive laboratory evolution. **Fungal Genetics and Biology**, v. 164, p. 103765, jan. 2023.

ANGERBAUER, C. et al. Conversion of sewage sludge into lipids by *Lipomyces starkeyi* for biodiesel production. **Bioresource Technology**, v. 99, n. 8, p. 3051–3056, maio 2008.

CHERUBINI, F. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. **Energy Conversion and Management**, v. 51, n. 7, p. 1412–1421, jul. 2010.

GEIJER, C.; LEDESMA-AMARO, R.; TOMAS-PEJO, E. Unraveling the potential of non-conventional yeasts in biotechnology. **FEMS Yeast Research**, v. 22, n. 1, p. 1–6, 2022.

LIU, L. et al. Efficient microbial oil production on crude glycerol by *Lipomyces starkeyi* AS 2.1560 and its kinetics. **Process Biochemistry**, v. 58, p. 230–238, jul. 2017.

MARUYAMA, Y. et al. Characterization of oil-producing yeast *Lipomyces starkeyi* on glycerol carbon source based on metabolomics and 13C-labeling. **Applied Microbiology and Biotechnology**, v. 102, n. 20, p. 8909–8920, out. 2018.

NAVARRETE, C.; L. MARTÍNEZ, J. Non-conventional yeasts as superior production platforms for sustainable fermentation based bio-manufacturing processes. **AIMS Bioengineering**, v. 7, n. 4, p. 289–305, 2020.

NIELSEN, J. Systems Biology of Metabolism. **Annual Review of Biochemistry**, v. 86, n. 1, p. 245–275, 20 jun. 2017.

NIELSEN, J. Yeast Systems Biology: Model Organism and Cell Factory. **Biotechnology Journal**, v. 14, n. 9, 2019.

PALSSON, B. Metabolic systems biology. **FEBS Letters**, v. 583, n. 24, p. 3900–3904, 2009.

POMRANING, K. R. et al. Transcriptomic analysis of the oleaginous yeast *Lipomyces starkeyi* during lipid accumulation on enzymatically treated corn stover hydrolysate. **Biotechnology for Biofuels**, v. 12, n. 1, p. 162, dez. 2019.

PUTRA, F. J. N. et al. The bioconversion of lignin derivative aldehydes into high-value aromatic alcohols and lipids via *Lipomyces starkeyi*. **Biochemical Engineering Journal**, p. 109065, ago. 2023.

RAHMAN, S. et al. Microbial lipid production from lignocellulosic hydrolyzates: effect of carbohydrate mixtures and acid-hydrolysis byproducts on cell growth and lipid production by *Lipomyces starkeyi*: Microbial lipid production from lignocellulosic hydrolyzates. **Journal of Chemical Technology & Biotechnology**, v. 92, n. 8, p. 1980–1989, ago. 2017.

SMITH, M. TH.; KURTZMAN, C. P. Lipomyces Lodder & Kreger-van Rij (1952). Em: **The Yeasts**. [s.l.] Elsevier, 2011. p. 545–560.

THORWALL, S. et al. Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. **Nature Chemical Biology**, v. 16, n. 2, p. 113–121, 1 fev. 2020.

VENTORIM, R. Z. et al. Genome-scale metabolic model of oleaginous yeast *Papiliotrema laurentii*. **Biochemical Engineering Journal**, v. 180, n. December 2021, p. 108353, 1 jan. 2022.

VIEIRA, N. M. et al. Isolation of a new *Papiliotrema laurentii* strain that displays capacity to achieve high lipid content from xylose. **3 Biotech**, v. 10, n. 9, p. 1–14, 2020.

YAMAKAWA, C. K. et al. Exploiting new biorefinery models using non-conventional yeasts and their implications for sustainability. **Bioresource Technology**, v. 309, p. 123374, ago. 2020.

YU, X. et al. Oil production by oleaginous yeasts using the hydrolysate from pretreatment of wheat straw with dilute sulfuric acid. **Bioresource Technology**, v. 102, n. 10, p. 6134–6140, maio 2011.

ZAKHARTSEV, M. et al. Metabolic efficiency in yeast *Saccharomyces cerevisiae* in relation to temperature dependent growth and biomass yield. **Journal of Thermal Biology**, v. 52, p. 117–129, 2015.

CHAPTER 1 - RECONSTRUCTION OF GENOME-SCALE METABOLIC MODELS OF NON-CONVENTIONAL YEASTS: CURRENT STATE, CHALLENGES, AND PERSPECTIVES

E.L.M. Almeida[a], E.J. Kerkhoven[b,c], W.B. Silveira[a]*.

[a]Laboratory of Microbial Physiology, Department of Microbiology, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

[b]Systems and Synthetic Biology, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden.

[c]SciLifeLab, Chalmers University of Technology, Gothenburg, Sweden.

*Correspondence to: Universidade Federal de Viçosa, Av. P. H. Rolfs, s/n, Viçosa, Minas Gerais 36570-900, Brazil. E-mail address: wendel.silveira@ufv.br (W.B. da Silveira); Phone: +55 31 3612-2427

Accepted for publication in Biotechnology and Bioprocess Engineering.

1.1. Abstract

Non-conventional yeasts are promising cell factories to produce lipids and oleochemicals, metabolites of industrial interest (e.g., organics acids, esters, and alcohols), and enzymes. They can also use different agro-industrial by-products as substrates within the context of a circular economy. Some of these yeasts can also comprise economic and health burdens as pathogens. Genome-scale metabolic models (GEMs), networks reconstructed based on the genomic and metabolic information of one or more organisms, are great tools to understand metabolic functions and landscapes, as well as propose engineering targets to improve metabolite production or propose novel drug targets. Previous reviews on yeast GEMs have mainly focused on the history and the evaluation of *S. cerevisiae* modeling paradigms or the accessibility and usability of yeast GEMs. However, they did not describe the reconstruction strategies, limitations, validations, challenges, and research gaps of non-conventional yeast GEMs. Herein, we focused on the reconstruction of available non-*Saccharomyces* GEMs, their validation, underscoring the physiological insights, as well as the identification of both metabolic engineering and drug targets. We also discuss the challenges and knowledge gaps and propose strategies to boost their use and novel reconstructions.

Keywords: metabolic engineering; yeast; metabolic modeling; non-*Saccharomyces*

1.2. Introduction

Yeasts have been used in various biotechnological processes, including food fermentation, the production of biofuels, pharmaceuticals, and bioingredients, as well as bioremediation and biocontrol. Most yeast research has been directed to *Saccharomyces cerevisiae*, a model organism widely applied in bioprocesses. However, it does not represent the physiological, morphological, and molecular diversity of the more than 2000 currently accepted yeast species (BOEKHOUT et al., 2022). Non-conventional yeasts, that is, other than *S. cerevisiae* (non-*Saccharomyces*), have received increasing attention due to their potential for producing metabolites, lipids, and proteins from both industrial by-products (e.g., lignocellulosic biomasses) and domestic wastes. Importantly, some non-conventional yeasts tolerate stress conditions such as high temperatures (e.g., *Kluyveromyces marxianus, Ogataea polymorpha*) and organic acids (e.g., *Rhodotorula toruloides, Issatchenkia orientalis*). However, some non-conventional yeasts cause food spoilage and diseases in humans, animals, and plants (GEIJER; LEDESMA-AMARO; TOMAS-PEJO, 2022; NAVARRETE; L. MARTÍNEZ, 2020). For example, *Cryptococcus neoformans* causes cryptococcosis (TEZCAN et al., 2023) and *Candida albicans* is the main etiological agent of fungemia in humans (VIANA et al., 2020); nevertheless, these infections are easily treated in healthy people and are more concerning in immunocompromised individuals. In plants, *Ustilago maydis* is associated with the corn smut disease, growing in the plant tissue and extracting substrates for its metabolism (LIEBAL et al., 2022).

Genome-scale metabolic models (GEMs) are networks reconstructed from a given organism's genome and metabolic information, which can be used to comprehend its physiology, gene essentiality, and metabolic landscape, uncover new engineering and drug targets, and understand evolutionary plasticity. The reconstruction of a GEM requires curated metabolic data and an annotated genome of the desired organism (FANG; LLOYD; PALSSON, 2020). The genome sequence can be mapped to a knowledge database [e.g., Kyoto Encyclopedia of Genes and Genomes (KEGG)] or a well-curated GEM of closely related species, generating a network containing all identified metabolic reactions. This first network is called a draft, which will be further refined to obtain a high-quality model. This process of model refinement (or curation) usually involves verifying reaction directionality, elemental and charge balance, assigning correct gene-protein-reaction (GPR) rules, incorporating novel reactions and metabolites, identifying and correcting gaps and dead-end cycles in the network, adding pseudo-reactions (e.g., exchange of substrates and metabolites, biomass equation to

represent growth), and assigning adenosine triphosphate (ATP) requirements. This process is iterative and continuous, alongside different validations, such as gene essentiality predictions and growth simulations. For example, the consensus network of *S. cerevisiae* is currently in version 8.6.2 (LU et al., 2019) (https://github.com/SysBioChalmers/yeast-GEM). For a more detailed description of GEM reconstruction protocols, see Thiele and Palsson (THIELE; PALSSON, 2010) and Heirendt et al. (HEIRENDT et al., 2019). A brief reconstruction workflow is presented in Figure 1.

The reconstructed metabolic network is mathematically represented by a stoichiometric matrix with the columns representing reactions, the lines representing metabolites, and the entries representing the coefficients of a specific metabolite in a reaction. Flux balance analysis (FBA) is the most common framework for making predictions using GEMs. In FBA, metabolic fluxes are estimated using inequalities that constrain the system, considering cells at a steady state. Then, the metabolic fluxes are optimized towards a desired phenotype (commonly growth/biomass production), known as the objective function, usually by linear programming (ORTH; THIELE; PALSSON, 2010). For a more detailed description of simulation strategies and frameworks, see Becker et al. (BECKER et al., 2007) and Lewis et al. (LEWIS; NAGARAJAN; PALSSON, 2012). A brief description of modeling algorithms used for phenotype prediction and optimization used in the studies discussed here is presented in Table 1.

Previous reviews on yeast GEMs have mainly focused on the history and the evaluation of *S. cerevisiae* modeling paradigms (LOPES; ROCHA, 2017; CHEN; LI; NIELSEN, 2022) or the accessibility and usability of yeast GEMs (DOMENZAIN et al., 2021). However, they did not describe the reconstruction strategies, limitations, validations, challenges, and research gaps of non-conventional yeast GEMs. Therefore, in the present review, we describe the reconstruction of the available non-*Saccharomyces* GEMs, their validation, physiological insights, and prediction of both metabolic engineering and drug targets. Finally, we discuss the knowledge gaps, challenges, and perspectives to boost their use and novel reconstructions.

Figure 1 – Overview of the reconstruction process of genome-scale metabolic models (GEMs). (I) The reconstruction of GEMs is based on a genome annotation of at least one species or strain of interest. Thus, the quality of the genome assembly and annotation might affect GEM quality (FANG; LLOYD; PALSSON, 2020). **(II)** GEM reconstruction can be conducted using automatic tools (e.g., RAVEN (AGREN et al., 2013), ModelSEED (SEAVER et al., 2021), *merlin* (CAPELA et al., 2022), KBase (ARKIN et al., 2018) based on networks and/or data available in online databases (e.g., KEGG (KANEHISA; GOTO, 2000), BRENDA (CHANG et al., 2021), MetaCyc (KARP et al., 2018), KBase (ARKIN et al., 2018), manually, or combining these strategies (THIELE; PALSSON, 2010). **(III)** The first metabolic network (draft network) is iteratively modified and curated during the subsequent reconstruction steps (THIELE; PALSSON, 2010). **(IV)** The validation and curation of the model are based on experimental data, including biomass composition, growth, substrate uptake and production rates, gene essentiality, and energy requirements. Experiments can be designed considering the necessities detected during the reconstruction process, or the data can be retrieved from previous studies. **(V-VI)** The curation and validation of the model are performed continuously and iteratively with *in silico* and experimental data obtained during and after the reconstruction of the model (THIELE; PALSSON, 2010). These steps can be assisted by varied tools, such as MEMOTE (LIEVEN et al., 2020), or functions already incorporated in the COBRA (HEIRENDT et al., 2019) and RAVEN (AGREN et al., 2013) toolboxes, for example. **(VII)** After the model has been validated and can be used to perform biological predictions, the GEM can be distributed on a platform, such as GitHub, that allows collaborative and continuous network improvement based on novel studies, necessities, and research objectives.

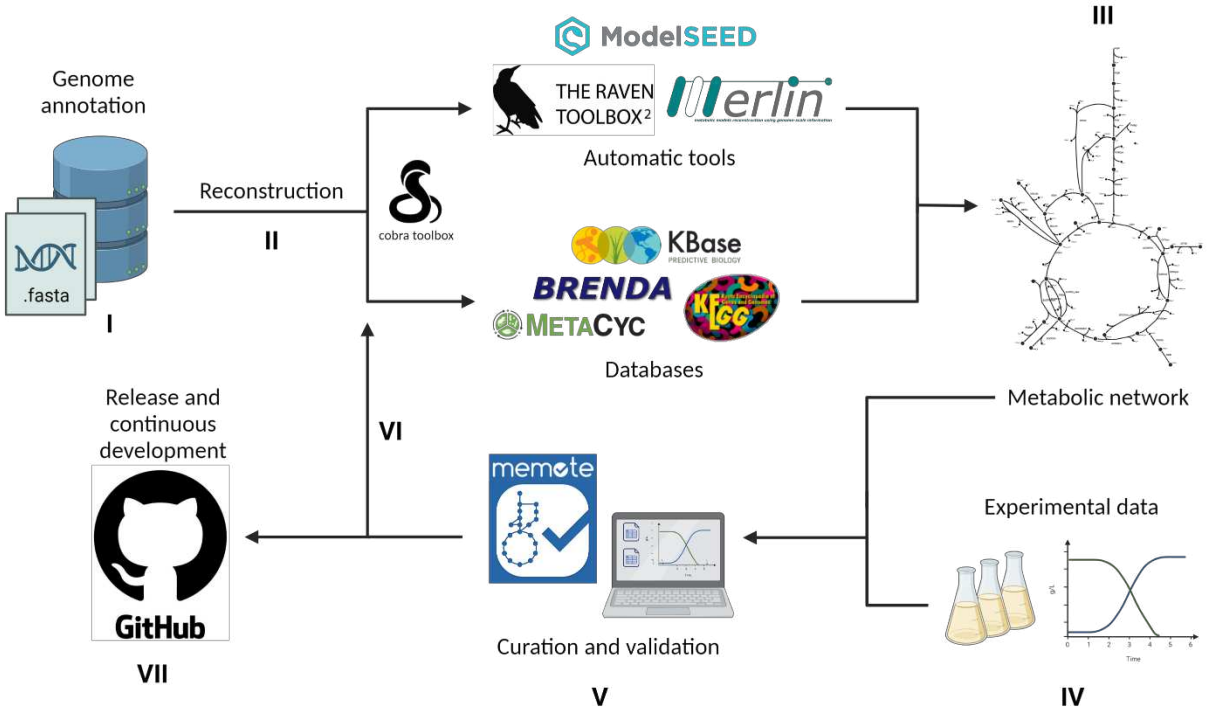


Table 1 – Summary of modeling algorithms used to simulate phenotypes using non-conventional yeast GEMs.

Algorithm	Description
Flux Balance Analysis (FBA)	Maximizes or minimizes an objective function (e.g., biomass production) via linear programming under the assumption of steady state (ORTH; THIELE; PALSSON, 2010).
Flux Variability Analysis (FVA)	Quantifies feasible flux ranges from FBA solutions (BURGARD; VAIDYARAMAN; MARANAS, 2001).
Dynamic Flux Balance Analysis (dFBA)	Incorporates the rate of flux changes when the kinetics of exchange reactions are well characterized (MAHADEVAN; EDWARDS; DOYLE, 2002)
Flux Scanning based on Enforced Objective Function (FSEOF)	Scans the fluxes when a given objective is enforced (product formation) under biomass optimization (CHOI et al., 2010).
OptGeneKnock	Combines a logic transformation model (LTM), a logical equivalent of the original GEM with GPR rules included in the stoichiometric matrix, with MILP to determine the best gene knockout to improve the production of a desired compound (ZHANG et al. 2015).
IdealKnock	Optimizes the times for determination of the best knockout strategy for overproduction by first scanning for mutants of interest, then iteratively combining these mutations (GU et al. 2016).
Analysis of Production and Growth Coupling (APGC)	Considers that growth and production are "compete" for metabolic resources. Thus, it first allocates part of the carbon uptake to optimize growth via FBA with stepwise increases while using the rest to optimize product formation (JIAN et al. 2016).
Genetic Design by Local Search (GDLS)	An heuristic algorithm that searches for multiple paths (genetic manipulations) from a defined or random starting point compared to different scenarios with *n* additional manipulations and the best strategy are selected via iterative MILP (LUN et al., 2009).
Transcriptomic-based Strain Optimization Tool (tSOT)	After the generation of condition specific models based on RNA-seq, it determines which deactivated reactions should be added to the to increase the production of a desired compound using MILP (LUN et al., 2009).
Cofactor Modification Analysis (CMA)	Determines the cofactor pools required for catalyzing reactions of interest based on mixed-integer linear programming (MILP) between an original metabolic state and one with altered cofactor specificity while both the product formation and biomass are maximized (LAKSHMANAN et al. 2013).
OptForce	Identifies the minimal set of genetic interventions to improve the product yield based on flux measurements of the wide-type strain. It determines the fluxes that should be increased, decreased, and turned off from FVA solutions (RANGANATHAN, S., P.F. SUTHERS, and C.D. MARANAS, 2010).
OptKnock	Suggests gene knockouts considering two metabolic objectives: maximization of biomass yield and minimization of metabolic adjustment (MOMA). It identifies and removes reactions that uncouple product and biomass formation combined with overproduction in a bilevel mixed-integer optimization (BURGARD; PHARKYA; MARANAS, 2003).

Minimization of Metabolic Adjustment (MOMA)	Identifies the minimal flux redistribution from a wide-type and to mutant strain by minimizing the Euclidean distance between their flux vectors (SEGRÈ; VITKUP; CHURCH, 2002).
Environmental version Minimization of Metabolic Adjustment (eMOMA)	Instead of identifying the flux between two strains, it determines the minimal flux redistribution from growth on a nutrient-rich and a nutrient-limited condition using the same mathematical formulation of MOMA (KIM et al. 2019).

1.3. Current knowledge

1.3.1. Komagataella phaffi

The non-conventional yeast with most GEM reconstructions is *Komagataella phaffi* (syn. *Pichia pastoris*) (Figure 2A). This methylotrophic yeast is widely used in recombinant protein production, single-cell protein (SCP), and biomass production from methanol. Currently, multiple expression vectors are designed for *Komagataella*, especially based on the alcohol oxidase 1 (*AOX1*) gene locus induced by methanol. *Komagataella* has also been proposed as a model organism for basic research. Besides the availability of genome editing tools and protocols and growth characteristics, *K. phaffii* evolved slower than *S. cerevisiae*, being closer to ancient yeasts, and shares more similarities with metazoans, highlighting its potential for eukaryotic molecular biology (BERNAUER et al., 2021).

SOHN et al. (2010) reconstructed the first *K. phaffi* model, PpaMBEL1254 (Table 2). The reconstruction was based on the genome annotation of *Pichia pastoris* DSMZ 70382, an automatic draft from KEGG (KANEHISA; GOTO, 2000) and the *Pichia* genome database (http://pichiagenome-ext.boku.ac.at:8080/apex/f?p=100:1:33508677146320), followed by manual curation and use of other databases, such as TransportDB (REN; KANG; PAULSEN, 2004). The biomass composition was determined using experimental and literature data, and the model was validated using batch and chemostat cultivations with glucose as the sole carbon source. The model could predict well the use of 25 carbon sources and the growth rate using glucose. The production of heterologous proteins was simulated by implementing additional reactions to describe the protein of interest, including nucleotides to represent the recombinant DNA and mRNA sequence and amino acids to represent the polypeptide, including the energetic requirements for polymerization. They predicted the maximum production rates of human serum albumin (HSA) and human superoxide dismutase (hSOD) and the effects of oxygen limitation. However, they did not propose target genes to improve protein production based on the simulations.

Table 2 – Summary of genome-scale metabolic models (GEMs) of *Komagataella phaffi*.

GEM	Genes	Metabolites	Reactions	Draft reconstruction strategy	Validation	Main insight(s)	Reference
PpaMBEL1254	540	1147	1254	Database-retrieved (KEGG, *Pichia* genome database).	Qualitative: growth on 25 carbon sources. Quantitative: batch and chemostat cultivations on glucose.	Effects of oxygen availability on heterologous protein production. In conditions with high levels of oxygen uptake (above 5 mmol/g DCW/h), the increase in protein production led to decrease in growth rate. In contrast, for oxygen uptake rates below 5 mmol/g DCW/h, the increase in protein production results in milder effects on growth.	(SOHN et al., 2010)
iPP668	668	1177	1361	Database-retrieved (BioSilico, BRENDA, ExPASy ENZYME, MetaCyc, KEGG).	Quantitative: chemostat cultivations on glucose and mixed glycerol/methanol.	Identification of metabolic targets to enhance 2,3-butanediol production, such as NADH and NADPH regeneration, acetoin supplementation, and overexpression of alcohol oxidase (AOD), formaldehyde dehydrogenase (FALDH), formate dehydrogenase (FDH), dihydroxyacetone kinase (DAK), glycerol-3-phosphate dehydrogenase (G3PD1ir), glycerol-3-phosphatase (G3PT), and glycerol dehydrogenase (GYCDy).	(CHUNG et al., 2010)

iLC915	915	1302	1423	Homology-driven (iIN800 of *S. cerevisiae*) and database-retrieved (KEGG) using the RAVEN Toolbox	Gene essentiality. Qualitative: viability on different carbon sources. Quantitative: batch and chemostat growth on glucose, methanol, glycerol, and mixed glycerol/methanol.	Effects of oxygen availability and mixed methanol and glycerol or sorbitol on heterologous protein production. Combinations of methanol-glycerol and methanol-sorbitol increases protein production compared to single carbon source use.	(CASPETA et al., 2012)
ihGlycopastoris	-	1443	1595	Manual curation and extension of iLC915 to incorporate native and humanized N-glycosylation using the RAVEN Toolbox	Quantitative: growth on varied carbon sources (e.g., glucose, glycerol, methanol) and oxygen uptake rates.	Effects of N-glycosylation on protein yield. Identification of targets to improve protein yield by FSEOF, including N-glycan metabolism, UDP-GlcNAc biosynthesis (GNA1, GFA1, UAP1, and PCM1), GDP-Man biosynthesis (GLK1, PSA1, PMI40 and PMM1), and pentose phosphate pathway (GND2 and GNK1).	(IRANI et al., 2016)
iMT1026	1026	1689	2035	Homology-driven. Consensus model of PpaMBEL1254, iPP668, and iLC915. Manual curation and extension.	Qualitative: growth on 47 carbon sources. Quantitative: chemostat cultivations on glucose, glycerol, and methanol. Gene essentiality in normoxic, limited oxygen, and hypoxic conditions.	Importance of continuous model refinement to improve simulation capabilities and precision on fatty acid biosynthesis, glycosylation, oxidative phosphorylation, and growth and maintenance energy.	(TOMÀS-GAMISANS; FERRER; ALBIOL, 2016)

iRY1243	1243	1740	2407	Homology-driven. Consensus model of PpaMBEL1254, iPP668, iLC915, and iMT1026. Manual curation and extension.	Qualitative: growth on 30 carbon and 21 nitrogen sources. Quantitative: chemostat cultivation on glucose and central metabolic fluxes compared to ^{13}C data. Gene coverage compared to RNA-seq. Gene essentiality.	Importance of continuous model refinement to improve simulation capabilities and precision. Effects of gene insertion, deletion, and overexpression on growth rate and S-adenosyl-l-methionine production for strain engineering.	(YE et al., 2017)
Kp.1.0	720	1221	1424	Homology-driven: consensus model of PpaMBEL1254, iPP668, and iLC915, and manual curation based on KEGG and MetaCyc.	Gene essentiality. Quantitative: growth of wild-type and recombinant strains on glucose or mixed methanol/glycerol or at different oxygen availability.	Effects of using multiple (12) biomass equations and consistent GPR rules for correct target prediction using FSEOF to improve human copper/zinc superoxide dismutase (hSOD). Histidine and lysine metabolic processes, as well as glutamine family amino acid biosynthesis were identified as possible targets.	(CANKORUR-CETINKAYA; DIKICIOGLU; OLIVER, 2017)
iMT1026v3	1026	1689	2035	Manual curation and extension of iMT1026 focused on methanol and glycerol growth.	Quantitative: chemostat cultivations on glycerol and methanol at different dilution rates.	Importance of continuous model refinement, especially biomass composition and exchange rates, and specialization for different applications. Calculations of average biomass compositions for growth on glycerol and methanol, and fine adjustments of growth and maintenance energy.	(TOMÀS-GAMISANS; FERRER; ALBIOL, 2018)

CHUNG et al. (2010) reconstructed the iPP668 GEM of *Pichia pastoris* GS115 to elucidate its metabolism and identify targets to enhance its physiological properties (Table 2). This reconstruction was based on a draft assembled by compiling annotated genes in BioSilico (HOU et al., 2004), BRENDA (CHANG et al., 2021), ExPASy ENZYME (https://enzyme.expasy.org/), MetaCyc (KARP et al., 2018), and KEGG, and was further refined by manual curation and gap-filling. The inclusion of essential reactions was performed interactively via FBA to ensure growth *in silico*. The carbohydrate and amino acid compositions were retrieved from experimental data, while lipid and sterol compositions were based on membrane structure. The growth-associated maintenance (GAM) was estimated considering protein, DNA, and RNA synthesis. Similar to Sohn et al. (SOHN et al., 2010), the model was validated using chemostat cultures in glucose minimal media. Predictions were also compared to literature data on mixed glycerol/methanol growth. However, discrepancies were detected when the methanol uptake was high. These results demonstrated the necessity to reconstruct a model capable of better-predicting methanol utilization.

Furthermore, CASPETA et al. (2012) reconstructed a GEM of *Pichia pastoris* GS115, iLC915, and evaluated for the first time both protein production and methanol metabolism (Table 2). The draft model was automatically generated using the RAVEN Toolbox (AGREN et al., 2013), leveraging on both the iIN800 GEM of *S. cerevisiae* as the reference network and the KEGG database for additional species-specific reactions. In this case, the homologous genes among yeasts were identified via bi-directional protein Basic Local Alignment Search Tool (BLASTp). This automatical reconstruction was supplemented and validated by biochemical and physiological evidence. This was crucial, for example, for peroxisomal metabolism, where AOX is localized. The production of heterologous proteins was modeled using a human monoclonal antibody, with four reactions to polymerize nucleotides, ribonucleotides, amino acids, and their assembly. Human monoclonal antibody 3H6 Fab fragment (FAB) production was evaluated using glucose and ammonia to estimate its metabolic burden and oxygen requirements. The simulation had a fixed protein amount, and the reaction exporting FAB was used as the objective function. Overall, the model accuracy for glucose and glycerol was similar to PpaMBEL1254 and superior to iPP668. Notably, iLC915 predictions were more similar to experimental data on methanol and mixed methanol and glycerol.

Moreover, IRANI et al. (2016) extended the iLC915 model to describe native *K. phaffi* and humanized N-glycosylation for recombinant proteins, named *ihGlycopastoris*, using RAVEN (Table 2). They assumed an average N-glycosylation on every 4691st amino acid,

estimated using proteomics data from *S. cerevisiae*. Recombinant protein production was determined by incorporating the polymerization reactions of several proteins with different sizes and glycosylation sites, either native or humanized. For simulations, the model was constrained using data on exchange fluxes and growth rates, and protein excretion was set as the objective function. After growth validation, they showed that the engineering of *K. phaffi* to produce humanized glycosylated proteins reduces theoretical protein yield due to the size of the N-glycan complex. Besides, they applied Flux Scanning based on Enforced Objective Function (FSEOF) (CHOI et al., 2010) to predict potential targets to improve protein yield. In this case, the model was constrained by the carbon uptake rate, while the protein yield was increased stepwise to the theoretical or experimental yield. Most identified reactions were related to amino acid biosynthesis, as anticipated, and N-glycan metabolism. Besides, two targets in the pentose phosphate pathway were detected, implying the increasing NADPH requirements.

TOMÀS-GAMISANS; FERRER; ALBIOL (2016) compared the previous models of *Kg. phaffi* and provided a general version applicable to both strains (GS115 and DSMZ 70382) (Table 2). This version was constructed by analyzing the pathways from the three previous models and literature data. They unified the metabolite charges and identifiers and constructed a consensus model draft by first merging PpaMBEL1254 and iPP668, then including iLC915. Although they used ModelBorgifier (SAULS; BUESCHER, 2014) in the first merger, most steps were manually conducted due to structural differences in the models. The COBRA and RAVEN toolboxes were used to ensure pathway connectivity and biomass formation. Two important features in the new model (iMT1026) were the upgrades in glycosylation pathways; and the oxidative phosphorylation electron transport chain, essential for recombinant protein production. The biomass reaction of this model was also curated for each carbon source tested, improving model accuracy.

In 2017, two research groups reconstructed novel GEMs of *K. phaffi* (YE et al., 2017; CANKORUR-CETINKAYA; DIKICIOGLU; OLIVER, 2017). YE et al. (2017) upgraded the iPP668, PpaMBEL1254, iL915, and iMT1026 models, described above. First, they again annotated all metabolites using KEGG and BiGG (KING et al., 2016) (Table 2). Then, each reaction was manually verified and corrected when necessary. These revised models were merged and updated with the most recent annotations of *K. phaffi* from KEGG, UniProt (CONSORTIUM, 2023), and IMG (https://img.jgi.doe.gov/). To find dead-end reactions, they applied the GapFind function from COBRA (HEIRENDT et al., 2019) and included essential

reactions from KEGG to solve the gaps. The final model, iRY1243, included 1243 annotated genes (an increase of more than 200 genes compared to iMT1026). Some new features comprised cell components, such as folate, thiamin, porphyrin, biotin metabolism, membrane transport reactions, and non-growth associated maintenance (NGAM) fitting based on chemostat data. Genes were verified with RNA-seq data obtained in the same study for *K. phaffi* G1HL under glucose chemostat cultivations, covering about 74% of all genes expressed when glucose was the sole carbon source. The model displayed a better capacity to predict the growth on 30 carbon and 21 nitrogen sources compared to iMT1026 and consistent flux distributions in the pentose phosphate pathway (PPP), Embden-Meyerhof-Parnas (EMP) pathway, and tricarboxylic acid (TCA) cycle compared to ^{13}C flux analysis. Finally, they validated the model for strain engineering based on the production of *S*-adenosyl-L-methionine.

CANKORUR-CETINKAYA; DIKICIOGLU; OLIVER (2017) standardized the model of *K. phaffii* (Kp.1.0), especially metabolite and reaction identifiers and gene-reaction rules (Table 2). First, they reconstructed a core model from the PpaMBEL1254, iPP668, and ILC915 models. They used KEGG and MetaCyc to resolve reaction directionalities, metabolite representation, and compartmentalization. Gap-filling was performed via manual curation. Notably, they provided 12 different biomass equations (carbohydrates, proteins, lipids, RNA, DNA, and GAM) to improve the predictive capability, which were based on the available literature accounting for various recombinant protein productions [wild-type strain (control), antibody-fragment (Fab), and *Rhizopus oryzae* lipase (ROL)], aeration levels (normoxic, oxygen-limited, and hypoxic), glycerol:methanol co-feeding ratios [80:20, 60:40, and 40:60 (w:w)], and dilution rates (0.05 and 0.16 h^{-1}). However, they failed to demonstrate the effects of these biomass differences on flux distributions under varied cultivation conditions since the reactions that presented a higher variation on the flux variability analysis (FVA) (BURGARD; VAIDYARAMAN; MARANAS, 2001) did not change with the biomass representation. Nevertheless, when the FSEOF approach was applied, the biomass equation affected the prediction of potential targets for recombinant protein production, demonstrating the importance of varied biomass equations suitable for the intended conditions.

Finally, TOMÀS-GAMISANS; FERRER; ALBIOL (2018) updated the iMT1026 model, focusing on both methanol and glycerol growth using chemostat data with a wide range of dilution rates (0.035, 0.050, 0.065, 0.100, 0.130, and 0.160 h^{-1} for glycerol and 0.035, 0.050, 0.065, 0.080, and 0.100 h^{-1} for methanol), which provided accurate physiological data, biomass composition, and elemental balance (Table 2). Besides the biomass and elemental composition,

the chemostat data was used to estimate the GAM and NGAM for these carbon sources. The updated model (iMT1026 v3.0) accurately predicted the growth parameters (growth, substrate and oxygen uptake, and CO_2 production rates) with errors below 5%. These results highlighted the importance of continuous improvement and curation of GEMs based on accurate experimental data (rates and biomass fractions), pivotal to reflect *in vivo* flux distributions.

1.3.2. Yarrowia lipolytica

After *Komagataella phaffi*, *Y. lipolytica* is the non-conventional yeast with more GEMs available (Figure 2A). *Y. lipolytica* has emerged as a model for the oleaginous phenotype [capacity of accumulating at least 20% (w/w) of dry biomass as lipids] and citric acid production (CAVALLO et al., 2017). It has the "generally regarded as safe" (GRAS) status and uses different substrates as carbon sources (e.g., hydrocarbons, sugars, fatty acids, alcohols, and organic acids), comprising a potential workhorse for producing biofuels and oleochemicals (GROENEWALD et al., 2014; ABDEL-MAWGOUD et al., 2018).

The first GEM of *Y.* lipolytica (iNL895) was reconstructed by LOIRA et al. (2012) using an automatic approach, Pantograph (LOIRA; ZHUKOVA; SHERMAN, 2015), based on the orthologous genes between the target (*Y. lipolytica*) and the scaffolds (*S. cerevisiae* iMM904, iIN800, and the yeast consensus network 4.36) (Table 3). The biomass reaction of the model was also based on *S. cerevisiae* with a few adjustments (e.g., nucleotide distribution). Most model additions and curations were focused on fatty acid and hydrocarbon metabolisms due to their importance for *Y. lipolytica* applications. For example, they introduced the ω-oxidation pathway for alkane conversion into alcohols, then into fatty acids, as well as alkane transport. Based on previous physiological data, they removed reactions for sucrose assimilation (absence of invertase) and ethanol production and consumption. The model neither presented anaerobic growth nor ethanol production, which is consistent with the fact that *Y. lipolytica* is a strictly aerobic yeast. The model had a parity with 64 cases (39 true positives and 25 true negatives) from 98 studies, including gene knockouts and different media and carbon sources. However, the model was not quantitatively validated (growth rate and flux data comparisons, for example).

Figure 2 – Distribution of genome-scale metabolic models (GEMs) of non-conventional yeasts. (A) Number of GEMs for each species (Black indicates Ascomycota; Dark gray indicates Basidiomycota). **(B)** Number of GEMs for each phylum and **(C)** Number of GEM per year based on **(A)**.

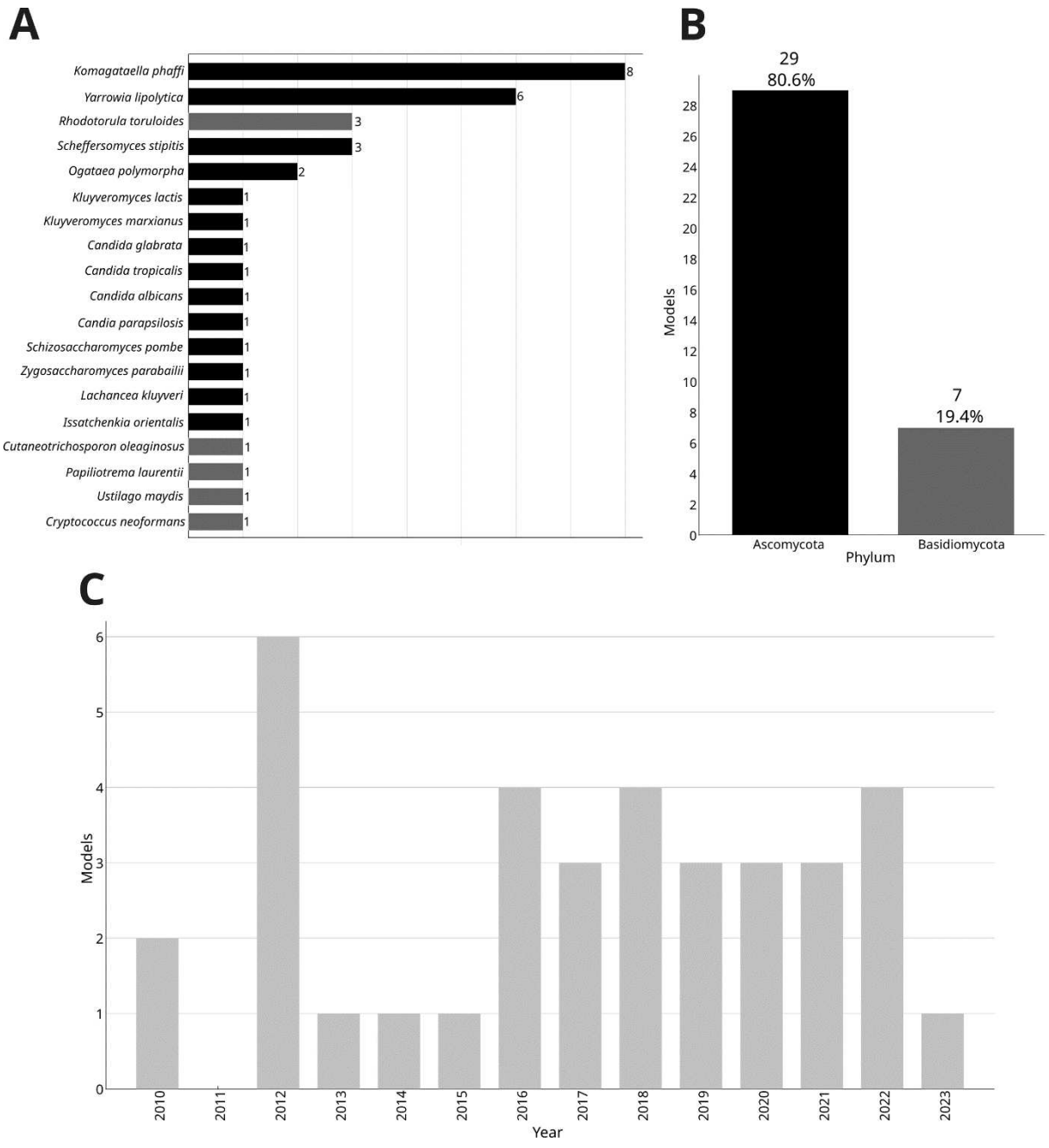


Table 3 – Summary of genome-scale metabolic models (GEMs) of *Yarrowia lipolytica*.

GEM	Genes	Metabolites	Reactions	Draft reconstruction strategy	Validation	Main insight(s)	Reference
iNL895	895	1847	2002	Homology-based (*S. cerevisiae* iMM904, iIN800, and the consensus 4.36) using Pantograph.	Qualitative: growth/ no growth compared to 64 conditions and gene knockouts.	Absence of anaerobic growth and ethanol production.	(LOIRA et al., 2012)
iYL619_PCP	619	843	1142	Database-retrieved (KEGG, BRENDA, BiGG).	Gene essentiality. Qualitative: growth on 29 carbon and nitrogen sources. Quantitative: growth on glucose minimal medium.	Prediction of targets to enhance lipid production using FVA, such as acetyl-CoA carboxylase for lipid production and the conversion of 2-phosphoglycerate to pyruvate for serine production.	(PAN; HUA, 2012)
iMK735	735	1111	1336	Homology-based (*S. cerevisiae* iND750).	Qualitative: growth on 25 carbon sources. Quantitative: Growth rates on glucose and glycerol; biomass yields and growth profile on glucose and glycerol by dFBA.	Use of different fatty acid compositions and lipid contents to differentiate growth and lipid accumulation phases. Prediction of targets to enhance lipid production during the accumulation phase by FVA. Effects of different NADPH sources on lipid synthesis rate. Effects of oxygen and limitation and citrate production rates on	(KAVŠCEK et al., 2015a)

						nitrogen-limited fed-batch cultivations.	
iYali4	901	1683	1985	Homology-based (*S. cerevisiae* consensus Yeast 7.1 and *Y. lipolytica* iNL895) using RAVEN Toolbox.	Quantitative: Growth rates on glucose and glycerol.	Improved lipid metabolism description for FAME integration. Inclusion of novel enzymes based on RNA-seq data: homogentisate 1,2-dioxygenase, endo-1,3(4)-β-glucanase, β-galactosidase, fumarylacetoacetase, orotidine-5'-P decarboxylase (Ura3), and argininosuccinate synthase (Arg1).	(KERKHOVEN et al., 2016)
iYL_2.0	645	1083	1471	Homology-based (*Y. lipolytica* iYL619_PCP, and iYali4).	Gene essentiality. Qualitative: growth on 29 carbon sources. Quantitative: Growth rate on glucose.	Engineering targets to enhance lipid production: OptGeneKnock, IdealKnock, and APGC. Some of the best targets were carbonate hydrolase, d-glyceraldehyde-3-phosphate glyceronetransferase, O-acyltransferase, fatty acid biosynthesis, amino acid metabolism and glycerolipid metabolism.	(WEI et al., 2017)
iYLI647	647	1152	1347	Homology-based (*Y. lipolytica* iMK735).	Quantitative: Growth rates on glucose and glycerol.	Inclusion of ω-oxidation pathway to convert fatty acids to DCAs. Prediction of targets to improve DCA production: GDLS, tSOT, and CMA. Cytochrome P450	(MISHRA et al., 2018)

			hydroxylase, fatty acid oxidase, acetyl-CoA carboxylase, cytosolic and mitochondrial malate dehydrogenases, NAD-dependent malic enzyme, and glutamate dehydrogenase were suggested for overexpression and stearoyl-CoA desaturase was indicated for knockout. Effect of branched-chain amino acid supplementation on DCA production.

PAN and HUA (2012) reconstructed a GEM of *Y. lipolytica* (iYL619_PC) using biochemical knowledge from various databases, including KEGG, BRENDA, and BiGG (Table 3). Transport and exchange reactions were included considering the reactions in the *S. cerevisiae* model iMM904 and the available literature. The biomass equation was integrated using experimental data from *Y. lipolytica,* while the GAM and NGAM were estimated using continuous culture data on glycerol. The predicted growth in minimal media was consistent with the experimental data. Regarding substrate assimilation, the growth on 24 of 29 (83%) substrates was correctly predicted. They also performed gene deletion analysis considering glucose minimal media and FVA to determine targets to enhance lipid production. The genes with higher variations were considered possible targets, such as acetyl-CoA carboxylase (highest variation).

KAVŠCEK et al. (2015b) reconstructed a GEM of *Y. lipolytica* to provide novel strategies to increase neutral lipid production while decreasing citrate excretion (Table 3). They used the *S. cerevisiae* GEM iND750 as a scaffold due to its accurate description of the lipid metabolism compared to the orthologs of *Y. lipolytica* in the KEGG database. A literature search defined the metabolites that the model can assimilate and produce. The final model, iMK735, comprised 735 genes, 1336 reactions, and 1111 metabolites. The fatty acid profile was differentiated between growth and lipid accumulation phases, experimentally determined in glycerol and glucose media. However, the composition of amino acids, nucleotides, and other metabolites remained the same as in iND750. During predictions, they varied the neutral lipid composition from 0.4 to 60%. Besides FBA, they used dynamic FBA (dFBA) (MAHADEVAN; EDWARDS; DOYLE, 2002) to simulate growth phases and calculate biomass yields. The final biomass differed from experimental data by only 6.6% for glucose and 2.2% for glycerol as carbon sources. They also used FVA to evaluate flux ranges during the lipid accumulation phase. Interestingly, they separately limited the sources of nicotinamide adenine dinucleotide phosphate (NADPH) to the PPP, cytosolic isocitrate dehydrogenase, malic enzyme, mannitol dehydrogenase, tetrahydrofolate synthase, or succinate semialdehyde dehydrogenase to evaluate their effects on lipid synthesis rates. Additionally, they modeled fed-batch cultivations using the specific citrate production rate from a nitrogen-limited stationary phase and simulated the effects of oxygen limitation.

KERKHOVEN et al. (2016) used the RAVEN toolbox and the *S. cerevisiae* consensus GEM Yeast 7.1 as a template to reconstruct a model for *Y. lipolytica* (iYali4) (Table 3). They modified the lipid metabolism description in the template to facilitate FAME integration and transferred

the curations from *Y. lipolytica* iNL895 and reactions to their new reconstruction. Model gap-filling and curation were performed manually. Notably, they used the RNA-seq assembly and annotation from *Y. lipolytica* W29 to identify six novel enzymes [homogentisate 1,2-dioxygenase, endo-1,3(4)-β-glucanase, β-galactosidase, fumarylacetoacetase, orotidine-5'-P decarboxylase (Ura3), and argininosuccinate synthase (Arg1)] and included them in the model. They adjusted the GAM and NGAM by fitting glucose uptake and growth data from previous *Y. lipolytica* continuous cultures with the ATP maintenance reaction as the objective function for simulations. Although the model performed well in predicting growth rates on glucose, it predicted considerably higher growth rates for glycerol compared to experimental results.

Furthermore, WEI et al. (2017) used two GEMs of *Y. lipolytica*, iYL619_PCP and iYali4, as templates to reconstruct a novel GEM (iYL_2.0). They added absent reactions in KEGG and standardized the annotation of metabolites and genes using UniProt, ChEBI (https://www.ebi.ac.uk/chebi/), and PubChem (https://pubchem.ncbi.nlm.nih.gov/) (Table 3). They used thermodynamic data and BioPath (REITZ et al., 2004) to determine reaction reversibility. Gaps were identified using the available literature and the GapFind function in the COBRA toolbox. The growth rate predictions in minimum media were better than previous models (iYL619_PCP, iNL895, and iYali4), as well as the predicted gene essentiality in different *in silico* media. Finally, they predicted gene targets to improve triacylglycerol (TAG) production in minimal media using two knockout algorithms: OptGeneKnock (ZHANG et al., 2015) and IdealKnock (GU et al., 2016), and one overexpression algorithm: analysis of production and growth coupling (APGC) (JIAN et al., 2016). The TAG synthesis reaction was the objective function at a fixed growth rate. OptGeneKnock indicated carbonate hydrolyase as the best single deletion and a double deletion with d-glyceraldehyde-3-phosphate glyceronetransferase. IdealKnock identified a quadruple knockout target: O-acyltransferase and three reactions in fatty acid synthesis. The APGC algorithm suggested six genes for overexpression related to amino acids metabolism, fatty acid degradation, glycerolipid metabolism, and biomass formation.

The most recently published GEM reconstruction of *Y. lipolytica* focused on the design of di-carboxylic acid (DCA)-overproducing strains (MISHRA et al., 2018). First, the authors compared the annotations of the four previous *Y. lipolytica* GEMs (iNL895, iYL619, iMK735, and iYali4) (Table 3). They selected iMK735 as the scaffold, then added the ω-oxidation pathway to convert fatty acids to DCAs, β-oxidation pathways, and the branched-chain amino acid (BCAA) pathway. The mass and elemental balance and directionality of reactions were

manually verified and corrected, and the biomass equation was redefined for nitrogen and carbon limiting conditions. The GAM and NGAM were adjusted using previous glycerol data. This model (iYLI647) was more accurate in predicting growth rates using glycerol and glucose. Notably, they used three approaches to predict targets to optimize DCA production, with glucose as the carbon source and dodecanedioic acid (DDDA) as the DCA representative. First, they applied genetic design by local search (GDLS) (LUN et al., 2009) to find growth-coupled solutions by knocking out reactions. Next, they used flux activity analysis to identify metabolism bottlenecks and potential overexpression targets. Then, RNA-seq data were used to implement the transcriptomic-based strain optimization tool (tSOT) (KIM et al., 2016). Interestingly, they also performed a cofactor modification analysis (CMA) (LAKSHMANAN et al., 2013) to determine targets to improve the pool of cofactors required to catalyze the engineered reactions. The overexpression targets by flux activity analysis included cytochrome P450 hydroxylase, fatty acid oxidase, and acetyl-CoA carboxylase. The tSOT indicated cytosolic and mitochondrial malate dehydrogenases, the NAD-dependent malic enzyme, and glutamate dehydrogenase as targets. Meanwhile, the GDLS suggested the simultaneous knockout of stearoyl-CoA desaturase, increasing the pool of free fatty acids and myoinositol-1-phosphate synthase. Finally, according to the CMA, increasing the supply of NADPH to the ω-oxidation pathway would also increase DCA production.

For a more in-depth comparison of *Y. lipolytica* GEMs, please refer to Xu et al. (XU; HOLIC; HUA, 2020).

1.3.3. Candida

The first GEM of *Candida* sp. yeasts was reconstructed for *C. glabrata* (XU et al., 2013), a multi-vitamin (pyridoxine, thiamine, nicotinamide, and biotin) auxotrophic yeast used for industrial pyruvate production. The first reconstruction step was based on three approaches: KGML (KEGG markup language) network retrieval, reaction list by KEGG Automatic Annotation Server (KAAS), and comparative genomics with *S. cerevisiae*, *Komagataella phaffi,* and *Aspergillus niger*. They integrated the reaction list obtained using the three methods comprising the draft model (Table 4). Next, they used GapFind (SATISH KUMAR; DASIKA; MARANAS, 2007) and the COBRA Toolbox to find gaps in the network. Reaction compartments were assigned using CELLO (YU et al., 2014) and WoLF PSORT (HORTON et al., 2007), and transport reactions were determined using TransportDB and Transporter

Classification Database (TCDB) (SAIER et al., 2021). The DNA fraction of the biomass equation was adjusted based on the GC content in the *C. glabrata* genome; lipids, carbohydrates, fatty acids, and cell-wall components were corrected using data from previous studies, and the amino acid distribution was determined from their batch cultures on glucose. Similar to other yeast GEMs, the model iNX804 was validated using FBA predictions regarding growth on 40 carbon and 20 nitrogen sources in minimal media, achieving a 95% match after the curation steps. The model was also able to predict the growth rate correctly. Since *C. glabrata* is commonly used to produce pyruvate, they evaluated two phases (growth and pyruvate production) to determine metabolic engineering targets and compare flux distributions. Finally, they used FSEOF to determine possible targets for fumarate accumulation and acetoin biosynthesis. Increased fumarate accumulation was identified when glycolytic and cytosolic reductive pathway fluxes increased. Notably, they constructed a strain of *C. glabrata* with increased fumarate transport and overexpression of malate dehydrogenase to validate this prediction. For acetoin, they introduced a third acetoin pathway to the model to allow growth-coupled production, which does not occur in wide-type strains. This strain was also validated *in vivo*.

The second GEM of the genus *Candida* was reconstructed for *C. tropicalis,* an oleaginous yeast that can use various carbon sources, such as sugars, alkanes, and fatty acids (MISHRA et al., 2016). Notably, it has the ω-oxidation pathway to produce DCAs, as mentioned above (see *Yarrowia lipolytica* section). First, the authors reconstructed a draft model using KEGG and BRENDA databases (Table 4). The reactions collected were manually checked for elemental and charge balances and assigned to subcellular compartments using CELLO. The connectivity of the draft network was verified with GapFind. The amino acid composition of the biomass equation was described based on their experimental data, and the lipid composition was obtained from previous literature. At the same time, total DNA and RNA contents were considered equal to *S. cerevisiae,* and their distribution was described based on *C. tropicalis* GC content. They also experimentally verified the elemental composition (C, N, H, and O) and found that it was consistent with the biomass composition in the model. They validated the model iCT646 using chemostat data in glucose minimal media at three dilution rates (0.1, 0.162, and 0.2 h^{-1}), as well as xylose and xylose-formate mixed media data from previous studies. They detected small amounts of ethanol for 0.162 and 0.2 h^{-1} and a CO_2 release/O_2 uptake ratio close to 1.0 in all evaluated conditions, indicating that *C. tropicalis*

might have a close to full respiratory metabolism. Furthermore, they used flux-sum analysis to compare the sum intensity of metabolites involved in lipid accumulation between *C. tropicalis* and *S. cerevisiae*. Most glycolytic intermediates presented similar flux-sums in both yeasts, but *C. tropicalis* had significantly higher fatty acid, acetyl-CoA, and cofactor turnover rates than *S. cerevisiae*. Finally, they evaluated the cellular behavior during DCA production using glycerol as the carbon source and a strain with disrupted β-oxidation. The simulations indicated that besides disrupting β-oxidation, the flux of α-monocarboxylic acid to *de novo* fatty acid synthesis should also be reduced to improve DCA yield.

Although *C. albicans* is considered the main etiological agent of candidiasis and the most common cause of invasive fungemia in humans (BERMAN, 2012), its first validated GEM was only released in 2020 by VIANA et al. (2020). The model was reconstructed using *merlin 4.0.2* (CAPELA et al., 2022), a semi-automatic reconstruction tool with a user-friendly interface, and was curated and validated with OptFlux 3.0 (ROCHA et al., 2010) (Table 4). This software also predicts transport reactions using the TCDB and TranSyt tool (LAGOA et al., 2021) and compartmentalization with WoLF PSORT. The biomass equation was described by experimental and literature data on *C. albicans*. Interestingly, the amino acid distribution was estimated based on the codon usage predicted from the genome sequence with the e-BiomassX tool (SANTOS; ROCHA, 2013). The theoretical P/O ratio was retrieved from the *S. cerevisiae* iMM904 model. The GAM was adjusted for the specific biomass composition of *C. albicans*, while the NGAM was retrieved from the *C. tropicalis* iCT646 model. The final model, iRV781, was qualitatively validated for carbon and nitrogen source utilization and quantitatively using previous data on glucose batch growth. When the oleic and nicotinic acid uptakes were unconstrained, the model predicted anaerobic growth, consistent with previous experimental data. Furthermore, the authors used a set of previously determined essential *C. albicans* genes to verify the gene essentiality predictions and determine drug targets for *C. albicans* considering growth in a medium replicating human serum conditions [Roswell Park Memorial Institute (RPMI) medium]. For example, the model predicted 11 ERG genes as essential in the RPMI medium, including the target of azole (*ERG11*).

Table 4 – Summary of genome-scale metabolic models (GEMs) of *Candida*.

GEM	Genes	Metabolites	Reactions	Draft reconstruction strategy	Validation	Main insight(s)	Reference
Candida glabrata							
iNX804	804	1025	1287	Database-retrieved (KEGG and KAAS) and homology-based (*S. cerevisiae*, *Komagataella phaffi,* and *Aspergillus niger*).	Gene essentiality. Qualitative: viability on 40 carbon and 20 nitrogen sources. Quantitative: growth rate on glucose batch cultivations. Validation *in vivo* of some targets for fumarate and acetoin production.	Identification of chitin synthesis, 1-6-Beta-glucan synthesis, dihydrofolate reductase, dihydropteroate synthase, ergosterol, cytochrome P450 51, squalene monooxygenase, thymidylate synthase, tryptophan-upregulated aromatic aminotransferase, CoA synthase, and catalase were predicted as drug targets. Prediction of overexpression targets to increase fumarate production using FSEOF: glucose-6-phosphate isomerase, phosphofructokinase, glyceraldehyde3-phosphate dehydrogenase, phosphoglycerate mutase, phosphopyruvate hydratase, pyruvate kinase, pyruvate carboxylase, malate dehydrogenase, and fumarase.	(XU et al., 2013)
Candida tropicalis							

	iCT646	646	1147	1420	Database-retrieved (KEGG and BRENDA).	Quantitative: elemental (C, N, H, and O) composition. Chemostat growth on glucose and batches with xylose and mixed xylose-formate.	Indicative that *C. tropicalis* has a close to full respiratory metabolism. β-oxidation disruption and preventing α-monocarboxylic acid from entering de novo fatty acid biosynthesis can improve DCA production using glycerol as the carbon source.	(MISHRA et al., 2016)
Candida albicans								
	iRV781	781	927	1221	Database-retrieved (KEGG and BRENDA) using *merlin 4.0.2*.	Gene essentiality. Qualitative: viability on 40 carbon and 15 nitrogen sources. Quantitative: aerobic and anaerobic batch growth on glucose.	Proposal of drug targets based on gene essentiality on human serum conditions (RPMI medium): squalene epoxidase (ERG1), dihydroorotate dehydrogenase (URA1), lanosterol synthase (ERG7), beta subunit of fatty acid synthetase (FAS1), thioredoxin reductase (TRR1), dihydrofolate reductase (DFR1), folic acid biosynthesis (FOL1), 1,3-beta-D-glucan synthesis (GSC1), inositol 1,3,4,5,6-pentakisphosphate 2-kinase (GSL1, GSL2), alpha subunit of fatty acid synthetase (FAS2), and lanosterol 14-alpha-demethylase (ERG11).	(VIANA et al., 2020)
Candida parapsilosis								

iDC1003	1003	1278	1804	Database-retrieved (KEGG and BRENDA) using *merlin 4.0.2*.	Gene essentiality. Qualitative: viability on 50 carbon and 17 nitrogen sources. Quantitative: batch growth on glucose.	Proposal of novel drug targets for the *Candida* genus based on gene essentiality on human serum conditions (RPMI medium) similar to the targets described above for *C. albicans*	(VIANA et al., 2022)

The most recent GEM from the *Candida* genus was reconstructed for *C. parapsilosis* (iDC1003) (VIANA et al., 2022) using a similar strategy to *C. albicans* described above (VIANA et al., 2020) (Table 4). *C. parapsilosis* is an emerging human pathogen and has become the second most common *Candida* spp. etiological agent. They also identified drug targets based on gene essentiality considering growth in the RPMI medium. In this case, 147 enzymes were predicted as essential. Interestingly, they crossed the essential enzymes in the *C. albicans*, *C. glabrata*, and *C. parapsilosis* models to determine general targets for *Candida* sp. and found 56 common enzymes. Besides the common targets of azoles and echinocandins (Erg11 and Fks11, respectively), they also identified other targets without orthologs in humans, including Abz1/2, Erg4, and Ura1.

1.3.4. Rhodotorula (Rhodosporidium) toruloides

The basidiomycete oleaginous yeast *Rhodotorula (Rhodosporidium) toruloides* is characterized by carotenoid production. It has become a promising host for metabolic engineering for fatty-acid-derived products, especially due to its capacity to accumulate high lipid amounts, sustain growth in lignocellulosic hydrolysates, and use different substrates. Three GEMs of *R. toruloides* are currently available (TIUKOVA et al., 2019; DINH et al., 2019; KIM et al., 2021) (Figure 2A).

TIUKOVA et al. (2019) used two well-curated GEMs as templates, one of *S. cerevisiae* (Yeast 8.2.0) and one of *Y. lipolytica* (iYali 4.1.1), to conduct an automatic reconstruction with RAVEN 2.2.1 (WANG et al., 2018) (Table 5). The model gap-filling was conducted using Meneco (PRIGENT et al., 2017) to avoid self-producing loops and stoichiometric inconsistencies with Yeast 8.2.0 as the standard. The lipid metabolism was represented with the SLIMEr formalism (SÁNCHEZ et al., 2019), which describes lipids as measurable entities. Only fatty acids with >5% of the overall acyl chain distribution were modeled based on experimental data, including 16:0, 18:0, 18:1, 18:2, and 18:3. The GAM and NGAM were determined with glucose uptake rates from continuous culture data. Seven reactions for torulene and carotene biosynthesis and 67 for mitochondrial β-oxidation were manually included. The annotation and consistency of the final model, *rhto*-GEM, were verified using MEMOTE (LIEVEN et al., 2020), presenting a relatively low score (62%), especially due to the effects of the SLIMEr formalism. The model was validated by comparing predictions with glucose, glycerol, and xylose *in vivo* growth. The gene essentiality prediction was also validated with

data generated with T-DNA insertion in more than 6000 genes. Furthermore, they evaluated the reactions that most affected the oleaginous phenotype, and the main differences between carbon sources were related to carbon assimilation. For example, the PPP was more important for xylose than for glycerol. Besides, acetyl-CoA carboxylase was identified as essential for all carbon sources. Finally, they employed FSEOF to optimize the production of TAGs, linolenic acid, and carotenoids. Besides glycolytic enzymes and the PPP under xylose, the components of the pyruvate decarboxylase (PDC)-aldehyde dehydrogenase (ALD)-acetyl-CoA synthetase (ACS) pathway were predicted to improve TAG production, as well as acetyl-CoA carboxylase, diacylglycerol O-acyltransferase, and glycerol-3-phosphate dehydrogenase. Regarding linolenic acid, apart from the targets identified to improve TAG production, oleoyl-CoA and linoleoyl-CoA desaturases were predicted for overexpression. For carotenoid improvement, torularhodin was used as the representative, and most enzymes on the mevalonate pathway were predicted for overexpression, as well as phytoene synthase and dehydrogenase (carotenoid biosynthetic enzymes).

DINH et al. (2019) reconstructed the draft model for *R. toruloides* using Yeast 7.6 (*S. cerevisiae*) as the scaffold, further extended using information from the "Build Fungal Model" application from KBase (ARKIN et al., 2018) (Table 5). The biomass equation was defined using Yeast 7.6 and data for *R. toruloides* generated on carbon and nitrogen limiting conditions, especially the macromolecular composition. The NGAM and GAM were estimated using chemostat data. Manually included reactions comprised genes in the carotenoid and D-arabitol synthesis pathways, as well as the lipid metabolism in oleaginous yeasts (e.g., ATP citrate lyase). The *iRhto*1108's annotation and consistency were also checked with MEMOTE, presenting a score of 87%, superior to *rhto*-GEM. Notably, the MEMOTE test was also used to correct unbounded fluxes that could generate infeasible cycles. In this case, the directionality of reactions and/or transporters was corrected. They also compared gene essentiality predictions to the previously published T-DNA insertion data. Next, they explored the metabolic responses of *R. toruloides* under nitrogen starvation. The model predicted increased TAG yield under both nitrogen and phosphate-limiting conditions. Next, they compared the predictions with previously published transcriptomic data. They found 12 upregulated (e.g., ATP synthase subunit, complexes I and III from the electron transport chain, and sterol and sphingolipids synthesis) and 11 downregulated (e.g., phospholipid synthesis, cell wall biosynthesis genes under nitrogen limitation related to flux predictions. Finally, they predicted targets to improve TAG production using OptForce (RANGANATHAN; SUTHERS; MARANAS, 2010). For

example, the first fatty acid elongation reaction, acetyl-CoA synthetase, ATP citrate lyase, and NADPH production via PPP, malic enzyme, and glycerol dehydrogenase were identified as targets for upregulation.

Moreover, KIM et al. (2021) used the iYL_2.0 from *Y. lipolytica* and reactions gathered from curated models on BiGG to reconstruct a novel *R. toruloides* GEM (Table 5). Protein function and localization were determined with MycoCosm (https://mycocosm.jgi.doe.gov/mycocosm/home) and WoLF PSORT, and peroxisomal protein localization was predicted using the PTS1 and PTS2 motifs with FIMO (GRANT; BAILEY; NOBLE, 2011). The reconstruction, validation, and simulations were performed in COBRApy (EBRAHIM et al., 2013), and the biomass equation was modified with BOFdat (LACHANCE et al., 2019), a tool to generate biomass objective functions based on experimental data. Interestingly, they used a segmented linear regression to determine lipid composition, considering fatty acids that increased from low to high lipid conditions. This curated draft of the model comprised 1106 genes, 1934 reactions, and 2010 metabolites in nine compartments. Notably, they used the model to evaluate the growth of carbon sources found in lignocellulosic hydrolysates (glucose, xylose, arabinose, *p*-coumarate) and compared to multi-omics (transcriptomics, metabolomics, proteomics) data obtained in synthetic defined media. Interestingly, simulations indicated that *R. toruloides* might use an alternative pathway for xylose and arabinose utilization via D-arabinitol and D-ribulose generating ribulose-5-phosphate instead D-xylulose-5-phosphate. Additionally, the *p*-coumarate assimilation pathway was elucidated. Subsequently, they used Biolog Phenotype Microarrays (PM1, PM2, PM3B, and PM4A) in 384 conditions, corrected the inconsistencies between the Biolog results and the model, and expanded the simulations to 213 conditions with an accuracy of 78.4%. Then, they used fitness scores from RB-TDNAseq to evaluate gene essentiality predictions. The final version of the model, which was refined after the phenotype screening and essentiality prediction, had 1142 genes, 2398 reactions, and 2051 metabolites.

Table 5 – Summary of genome-scale metabolic models (GEMs) of *Rhodotorula toruloides*.

GEM	Genes	Metabolites	Reactions	Draft reconstruction strategy	Validation	Main insight(s)	Reference
rhto-GEM	852	2277	2731	Homology-based (*S. cerevisiae* Yeast 8.2.0 and Y. *lipolytica* iYali 4.1.1) using RAVEN 2.2.1.	Annotation and consistency verified with MEMOTE. Gene essentiality compared to T-DNA insertion data. Quantitative: growth rates on glucose, glycerol, and xylose; maximum TAG yield.	Alterations in the oleaginous phenotype promoted by different carbon sources. Engineering targets to improve TAG, linolenic acid, and carotenoid (torularhodin) production by FSEOF. Acetyl-CoA carboxylase, 6-phospogluconolactonase, glucose-6-phosphate dehydrogenase, phosphogluconate dehydrogenase, ribulose 5-phosphate 3-epimerase, Stearoyl-CoA desaturase, PA phosphatase were common targets for TAG and linolenic acid improvement; mevalonate kinase, hydroxymethylglutaryl CoA reductase, mevalonate pyrophoshate decarboxylase, phosphomevalonate kinase, farnesyltranstransferase, phytoene synthase, phytoene dehydrogenase, dimethylallyltranstransferase were some of the targets	(TIUKOVA et al., 2019)

						predicted to improve torularhodin production.	
*iRhto*1108	1108	1985	2204	Homology-based (*S. cerevisiae* Yeast 7.6) and automatic with the "Build Fungal Model" application in KBase.	Annotation and consistency verified with MEMOTE. Gene essentiality compared to T-DNA insertion data. Auxotrophies. Qualitative: viability on 18 carbon sources. Quantitative: growth rate on xylose and glucose, and arabitol production.	Effects of nitrogen and phosphate limitation on growth and TAG production with comparisons of flux predictions to RNA-seq data. Engineering targets to improve TAG production using OptForce: glucose 6-phosphate dehydrogenase, glycerol dehydrogenase (NADP-dependent), malic enzyme (NADP), diacylglycerol acyltransferase, stearoyl-CoA desaturase, acyl-CoA dehydrogenase, ACP S-acetyltransferase, acetyl-CoA carboxylase were identified for overexpression; ferrocytochrome-c:oxygen oxidoreductase, citrate to cis-aconitate(3-), oxoglutarate dehydrogenase were identified for downregulation; lipid degradation and beta-oxidation were suggested for knockout.	(DINH et al., 2019)

Rt_IFO0880	1142	2051	2398	Homology-based (*Y. lipolytica* iYL_2.0) and database-retrieved (BiGG).	Gene essentiality compared to RB-TDNAseq. Qualitative: Biolog Phenotype Microarrays with 384 conditions. Quantitative: growth on glucose, xylose, arabinose, and *p*-coumarate.	Segmented linear regression to determine fatty acid composition from low to high lipid conditions. Description of xylose, arabinose, and *p*-coumarate assimilation pathways based on multi-omics data.	(KIM et al., 2021)

1.3.5. Scheffersomyces stipitis

Previously known as *Pichia stipitis*, *Scheffersomyces stipitis* can assimilate xylose and convert it into ethanol, as such, it has been considered a promising yeast for producing second-generation ethanol. Three GEMs of *S. stipitis* have been reconstructed (CASPETA et al., 2012; BALAGURUNATHAN et al., 2012; HILLIARD et al., 2018).

BALAGURUNATHAN et al. (2012) first compiled the reactions in KEGG and MetaCyc databases to reconstruct the draft network (Table 6). Then, they organized these reactions in three compartments using protein localization predictors and adjusted the directionality, charge, and elemental balance based on the literature, previous GEMs, and BRENDA, KEGG, and BiGG databases. The biomass composition was experimentally determined for carbohydrates, proteins, lipids, and nucleic acids. The growth and sugar uptake rates from their and previous experiments were used to estimate the GAM and NGAM. The model was validated and expanded using the Biolog Phenotype MicroArray (190 carbon, 95 nitrogen, 59 phosphorus, and 35 sulfur sources). This reconstruction led to the iBB814 model, which was further applied to predict the requirements for anaerobic growth, a known drawback for *S. stipitis* industrial applications. Six reactions were identified using KEGG and glucose as the carbon source, four related to phospholipids and two involved in converting dihydroorotate to orotate. They also evaluated the impacts of NADPH and nicotinamide adenine dinucleotide (NADH) ratios on xylose utilization and found low xylitol production under all conditions and that oxygen availability constrained its uptake, especially when xylose reductase was dependent on NADPH.

CASPETA et al. (2012) reconstructed a GEM for *S. stipitis* using the iIN800 model of *Saccharomyces cerevisiae* as the reference network (Table 6). The identification of homologs between the two yeasts and other steps of the reconstruction process was conducted using RAVEN. The biomass equation was described using previous studies, and the ATP requirements were estimated using the available continuous cultivation data. The predictions using the iSS884 model showed a dependence between oxygen levels and xylose utilization, as well as the anaerobic growth impairment and the absence of xylitol production. Interestingly, they simulated a three-reactor series with glucose-xylose mixtures found in lignocellulosic hydrolysates: (i) glucose uptake with high oxygen availability to stimulate biomass production with ethanol production; (ii) the stream of the last reactor, with final depletion of glucose and

initial use of xylose in lower oxygen availability to favor ethanol production; (iii) conversion of xylose to ethanol with oxygen 10 times lower than in the first reactor.

These two models of *Scheffersomyces stipitis* were used to validate a system identification (SID)-based framework to evaluate metabolic networks (DAMIANI et al., 2015), which was further used by HILLIARD et al. (2018) to update the IBB814 model (Table 6). This framework does not rely on point validations (e.g., comparisons to experimental on single conditions) but in multivariate analysis, such as Principal Component Analysis (PCA). To improve iBB814 and generate iDH814, incorrect model predictions were first identified. For example, the oxidative PPP was not activated under aerobic conditions, which is not expected for *S. stipitis* and many other yeasts. Then, the desired behavior was forced by applying additional constraints to identify key reactions (e.g., forcing increasing fluxes in the oxidative PPP with increasing oxygen uptakes). The reactions that mostly responded to the forced behavior were identified with PCA using the Principal Component (PC) magnitude. The cause of the inactive oxidative PPP was the malate oxidation reaction, which would not have been quickly identified without the SID-based framework. Finally, these reactions were manually examined for elemental and charge balance and directionality. Furthermore, they included a constraint considering the flux ratio between NADPH and NADH-dependent xylose reductase (XR), which was tuned to better agree with experimental data. This ratio should be increased while oxygen availability decreases, shifting the preference of this enzyme from NADPH to NADH and compensating NADH production by the XR-xylitol dehydrogenase (XDH) reaction pair. After the improvements, iDH814 predicted higher growth rates, less CO_2 production, a narrow ethanol production phase, and broader xylitol production, which are closer to experimental results.

Table 6 – Summary of genome-scale metabolic models (GEMs) of *Scheffersomyces stipitis*.

GEM	Genes	Metabolites	Reactions	Draft reconstruction strategy	Validation	Main insight(s)	Reference
iBB814	814	971	1371	Database-retrieved (KEGG, MetaCyc, BRENDA, and BiGG).	Qualitative: Biolog Phenotype Microarray (190 carbon, 95 nitrogen, 59 phosphorus, and 35 sulfur sources)	Identification of requirements for anaerobic growth, such as those involved in the direct formation of O_2 or from H_2O_2 (via glutathione or pyridoxine), and conversion of dihydroorotate to orotate. Impacts of NADPH and NADH ratios and oxygen availability on xylose utilization.	(BALAGURU NATHAN et al., 2012)
iSS884	884	922	1332	Homology-based (*S. cerevisiae* iIN800) using RAVEN.	Gene essentiality. Qualitative: viability on different carbon sources. Quantitative: growth on glucose and xylose with varied oxygen uptake rates	Dependence between oxygen levels and xylose utilization. Simulation of a three-reactor series to improve ethanol production from a glucose-xylose mixture. Glucose was quickly metabolized to ethanol and biomass (1st reactor), then glucose was depleted and half of the xylose available was converted to ethanol (2nd reactor), and the remaining xylose was converted to ethanol (3rd reactor).	(CASPETA et al., 2012)

iDH814	814	972	1380	Homology-based (*Scheffersomyces stipitis* iBB814).	Quantitative: growth rates on glucose and xylose, and CO_2, ethanol, and xylitol production rates.	Demonstration of the system identification (SID)-based framework to evaluate and improve GEMs. Inclusion of a constraint that considers the flux ratio between NADPH and NADH-dependent xylose reductase (XR) correlated to oxygen levels.	(HILLIARD et al., 2018)

1.3.6. Kluyveromyces

Two GEMs are currently available for the *Kluyveromyces* genus, one for *Kluyveromyces lactis* (DIAS et al., 2014) and one for *Kluyveromyces marxianus* (MARCIŠAUSKAS; JI; NIELSEN, 2019). These two yeasts can use diverse carbon sources, such as xylose, xylitol, lactose, arabinose, and cellobiose, to produce enzymes (such as β-galactosidase), ethanol, and volatile compounds. Although *K. lactis* has been studied more in terms of physiology and genetics, *K. marxianus* has gained increased attention due to its thermotolerance and fast growth capacity (SPOHNER et al., 2016; KARIM; GERLIANI; AÏDER, 2020).

DIAS et al. (2014) reconstructed and validated the GEM of *K. lactis* (iOD907) using *merlin 2.0*, based on the iMM904 model of *Saccharomyces cerevisiae* model, data from Biolog Phenotype MicroArrays and literature (Table 7). First, the genome annotation was updated, and the identified KEGG pathways were retrieved. Then, they manually verified and corrected the reversibility of reactions and identified transport reactions by searching for transmembrane helixes, followed by comparisons with TCDB. Only facilitated diffusion reactions, such as H_2O, CO_2, and NH_3, were included without gene associations. To determine the biomass equation, they assumed the composition was similar to *S. cerevisiae* (iMM904), except for the nucleotide, amino acid, and polysaccharide distribution. The GAM was also retrieved from iMM904, and the NGAM was adjusted using chemostat data. The gap-filling procedure was manually conducted with iMM904 and KEGG as references. Although the model predicted growth under anaerobic conditions, *K. lactis* does not grow under anoxic conditions, even in the presence of sterols and unsaturated fatty acids. This result indicated that the absence of this phenotype *in vivo* is unrelated to a metabolic deficiency or sterol uptake and might be related to other regulatory mechanisms. Besides, they used FVA to evaluate the profile of fermentation byproducts (e.g., glycerol and acetate) under low oxygen availability and performed gene essentiality analysis.

The GEM of *K. marxianus* was reconstructed by Marcišauskas et al. (MARCIŠAUSKAS; JI; NIELSEN, 2019) using RAVEN and information from multiple databases (KEGG, MetaCyc, TransportDB, and BRENDA). The first draft was reconstructed using the iOD907 model of *K. lactis* as a template, and the second using reactions from KEGG (Table 7). Then, the unique reactions in the KEGG model were incorporated into the first draft, and compartment discrepancies were resolved considering the second model. When available, the biomass equation was retrieved from iOD907 and adjusted with literature data for *K.*

marxianus. The GAM, NGAM, and P/O ratio remained the same as in iOD907. The *fill-Gaps* function in RAVEN was used to correct gaps with iOD907, Yeast 7.6, and KEGG as references. The overall consistency of the iSM996 model was validated with MEMOTE. The model was qualitatively (growth/non-growth) and quantitatively (growth and uptake rates) validated for different carbon and nitrogen sources. Notably, they integrated transcriptomic data to generate condition-specific models considering growth on glucose with different oxygen availabilities and temperatures (30 and 45 °C) by deactivating genes. These models were further used to predict the production capacity of biomass components. They found auxotrophies for L-lysine and L-isoleucine (previously reported), riboflavin, ferroheme, L-alanine, L- phenylalanine, and L-tyrosine at 45 °C, which would contribute to increased carbon and energy availability in this condition.

Table 7 – Summary of genome-scale metabolic models (GEMs) of *Kluyveromyces*.

GEM	Genes	Metabolites	Reactions	Draft reconstruction strategy	Validation	Main insight(s)	Reference
Kluyveromyces lactis							
iOD907	907	1476	1867	Database-retrieved (KEGG) using *merlin 2.0* and homology-based (*S. cerevisiae* iMM904).	Gene essentiality. Qualitative: Biolog Phenotype Microarray with 21 carbon sources. Quantitative: batch growth on glucose.	Absence of anaerobic growth is unrelated to metabolic deficiencies or sterol uptake. Profile of fermentation byproducts (e.g., glycerol, acetate) under low oxygen availability by FVA.	(DIAS et al., 2014)
Kluyveromyces marxianus							
iSM996	996	1531	1913	Database-retrieved (KEGG and MetaCyc) and homology-based (*Kluyveromyces lactis* iOD907) using RAVEN.	Annotation and consistency verified with MEMOTE. Qualitative: viability on media with different carbon and nitrogen sources. Quantitative: Growth rate on glucose, galactose, lactose, fructose, xylose, and sucrose.	Integration of transcriptomic data to generate condition-specific models. Presence of auxotrophies for L-lysine and L-isoleucine, riboflavin, ferroheme, L-alanine, L- phenylalanine, and L-tyrosine at 45 °C.	(MARCIŠAUSKAS; JI; NIELSEN, 2019)

1.3.7. Ogataea polymorpha

Also known as *Hansenula polymorpha* and *Pichia angusta*, *Ogataea polymorpha* is a methylotrophic yeast closely related to *K. phaffi*. Likewise, *K. phaffi*, *O. polymorpha* has been widely applied for recombinant protein production. Nevertheless, *O. polymorpha* can ferment xylose to ethanol, grow in temperatures up to 50 ºC, assimilate nitrate, and its AOX promoter has high activity with glycerol as the sole carbon source, different from *K. phaffi*. Two GEMs of *O. polymorpha* have been recently reconstructed (LIEBAL et al., 2021; ZORRILLA; KERKHOVEN, 2022).

LIEBAL et al. (2021) reconstructed iUL909 using the iMT1026v3 model of *K. phaffi* as a template due to their phylogenetic proximity (Table 8). They added 114 reactions (39 new genes) from the iRY1243 model of *K. phaffi* identified as necessary using the Biolog Phenotype Microarray and/or shake flask cultivations. The GAM, NGAM, and biomass equation of iUL909 were retrieved from iMT1026v3, which led to growth predictions close to its template. The iUL909 model presented a MEMOTE score of 45%, representing an improvement from iMT1026v3 (only 24%). Finally, they applied the FSEOF approach to predict targets to overproduce lactate and succinate using methanol or glucose as carbon sources. They found that stimulating the glyoxalase system would enhance lactate production, while lower glycolysis enhancement (glyceraldehyde-3-phosphate dehydrogenase, pyruvate kinase, and pyruvate carboxylase) would improve succinate production.

ZORRILLA; KERKHOVEN (2022) reconstructed a draft GEM of *O. polymorpha* (*hanpo*-GEM) to describe a protocol for reconstructing networks using high-quality models of phylogenetically-related organisms as templates and the RAVEN Toolbox (Table 8). In this case, they used Yeast 8.3.0 and *rhto*-GEM v.1.1.2 as templates for the draft reconstruction. The biomass macromolecular content was modified using previous experimental data. The DNA and RNA distribution was estimated based on the genome sequence, the constitutive amino acid ratio was estimated using the protein FASTA file, carbohydrate distribution was based on experimental evidence, and lipids were described using the SLIMEr formalism. The gap-filling was conducted using the *fillGaps* function in RAVEN with Yeast 8.3.0 as the reaction source. Besides updating the gene-reaction rules, the pathway required for methanol assimilation was manually inserted. The main purpose of the published book chapter was to exemplify how to initiate the reconstruction of a GEM with RAVEN, and the reconstructed *hanpo*-GEM should

therefore be regarded as an intermediate draft GEM and not be used directly due to the lack of deeper curation and validation.

Table 8 – Summary of genome-scale metabolic models (GEMs) of *Ogataea polymorpha*.

GEM	Genes	Metabolites	Reactions	Draft reconstruction strategy	Validation	Main insight(s)	Reference
iUL909	909	1639	2263	Homology-based (*Komagataella phaffi* iMT1026v3 and iRY1243).	Annotation and consistency verified with MEMOTE. Qualitative: Biolog Phenotype Microarray with 190 carbon sources. Quantitative: growth rate on glycerol, methanol, and glucose.	Engineering targets to improve the production of lactate and succinate with glucose or methanol as carbon sources by FSEOF: improved conversion of dihydroxyacetone phosphate to methylglyoxal to lactoyl-GSH to lactate can improve lactate production; improved conversion of glyceraldehyde-3-phosphate to pyruvate to oxalacetate to succinate might increase succinate production.	(LIEBAL et al., 2021)
hanpo-GEM	984	2118	2370	Homology-based (*S. cerevisiae* Yeast 8.3.0 and *R. toruloides rhto*-GEM 1.1.2).	-	Draft model used only to exemplify the protocol for GEM reconstruction using RAVEN. It lacks validation and curation for biological simulations.	(ZORRILLA; KERKHOVEN, 2022)

1.3.8. Schizosaccharomyces pombe

The model yeast *Schizosaccharomyces pombe* has only one GEM available. Sohn et al. (SOHN et al., 2012) reconstructed SpoMBEL1693 to demonstrate the Reconciliation of *In silico*/*in vivo* mutaNt Growth (RING) framework designed to validate and improve GEMs based on single-gene knockout mutant libraries (Table 9). This framework comprises an iterative process in which the growth phenotypes of single knockout mutants are compared to *in silico* single knockouts, characterized as lethal or viable. The draft model was built using the annotations in KEGG, NCBI, and GeneDB. Compartmentalization was described considering only experimental evidence of protein localization and transport reactions was collected from TransportDB. Gaps and reaction balance, and directionality were manually corrected. To describe the biomass equation, they determined the amino acid and fatty acid distribution, calculated the DNA distribution using the genome annotation, and retrieved the carbohydrate profile from the available literature. Interestingly, they considered a molar distribution of 5% mRNA, 75% rRNA, and 20% tRNA. The mRNA composition was determined considering the genome annotation. The rRNA was calculated with 16S, 23S, and 5S subunit sequences, and tRNA was estimated with leucine and glycine-transporting RNAs. The final network, adjusted using RING, included 605 genes, 1693 reactions, and 1744 metabolites divided into eight compartments. Besides RING, SpoMBEL1693 was validated both qualitatively (growth/non-growth on various carbon sources) and quantitatively (ethanol production at increasing dilution rates via FVA).

Table 9 – Summary of genome-scale metabolic models (GEMs) of other non-conventional yeasts.

GEM	Genes	Metabolites	Reactions	Draft reconstruction strategy	Validation	Main insight(s)	Reference
Schizosaccharomyces pombe							
SpoMBEL1693	605	1744	1693	Database-retrieved (KEGG).	Gene essentiality. Qualitative: viability on media with different carbon sources. Quantitative: ethanol production at different dilution rates by FVA.	Demonstration of the RING framework to validate and improve GEMs.	(SOHN et al., 2012)
Zygosaccharomyces parabailii							
ZyPa1	807	2091	3096	Database-retrieved (KEGG) and homology-based (*Kluyveromyces lactis* iOD907 and *S. cerevisiae* Yeast 7).	Gene essentiality. Quantitative: biomass yields on chemostat on cultivations glucose.	Fluxes ranges and main pathways of acetate consumption in the presence of glucose by FVA. Acetate contributed to both cytosolic and mitochondrial acetyl-CoA pool, and catabolism of this organic can contribute to detoxification.	(FILIPPO et al., 2018)
Lachancea kluyveri							

	iPN730	730	1179	1235	Homology-based (13 fungal templates) using the "Build Fungal Model" function in KBase.	Qualitative: aerobic and anaerobic growth on seven carbon sources. Quantitative: growth rates and respiration quotients.	Ethyl acetate and ethanol production profile in aerobic, semi-aerobic, and hypoxic conditions with glucose as the carbon source by dFBA. Context-specific models (GIMME algorithm) using RNA-seq data on growth with uracil or ammonium as nitrogen sources. On uracil, the most affected pathways were purine, histidine riboflavin metabolism, uracil degradation, phospholipid metabolism, transport.	(NANDA et al., 2020)
Issatchenkia orientalis								
	*iIsor*850	850	1702	1826	Homology-based: *S. cerevisiae* Yeast 7.6; "Build Fungal Model" function in KBase.	Annotation and consistency verified with MEMOTE. Gene essentiality: validation of genes via CRISPR-Cas9 editing. Quantitative: growth rate on glucose, glycerol, fructose, succinic acid, lactic acid, ethanol, and citric acid.	Engineering targets to improve succinic acid production using glucose as the carbon source by OptKnock with validation of some targets *in vivo*: pyruvate decarboxylase, aspartate transaminase, valine transaminase, glucose-6-phosphate dehydrogenase, and homoserine dehydrogenase.	(SUTHERS et al., 2020)
Cutaneotrichosporon oleaginosus								
	*i*NP636_*Coleaginosus*_ATCC20509	636	1373	1553	Homology-based (*Y. lipolytica* iNL895).	Quantitative: comparison of growth and lipid production with	Effects of using a dynamic biomass equation with various C/N ratios for lipid production on	(PHAM et al., 2021)

					surface response method for varied C/N ratios.	glucose, xylose, fructose, sucrose, and glycerol. Acetyl-CoA sources for lipid biosynthesis by FVA. The model suggested that ATP:citrate lyase is the main source of acetyl-CoA.	
Papiliotrema laurentii							
papla-GEM	796	2127	2465	Homology-based (*R. toruloides rhto*-GEM 1.3.0 and *Y. lipolytica* iYali4.1.2) using RAVEN 2.5.3.	Annotation and consistency verified with MEMOTE. Qualitative: growth on 10 carbon sources. Quantitative: growth and carbon uptake rates on glucose and xylose batch and chemostat cultivations.	Engineering targets to improve lipid production on glucose or xylose by FSEOF and eMOMA. Sugar transport, acetyl-CoA carboxilase, adenylate kinase, fatty-acyl-CoA synthase, NADH and O_2 transport were the best targets for both glucose and xylose. Ferrocytochrome-c:oxygen oxidoreductase and ubiquinol:ferricytochrome c reductase were also highlighted for xylose.	(VENTORIM et al., 2022)
Ustilago maydis							
iUma22	814	1233	1856	Database-retrieved: Pathway Tools (PathoLogic component)	Annotation and consistency verified with MEMOTE. Qualitative: Biolog Carbon Microarray. Quantitative: batch growth with different glucose concentrations.	Pan-GEM to identify common and specific reactions between different strains.	(LIEBAL et al., 2022)

Cryptococcus neoformans							
iCryptococcus	649	1143	1270	Database-retrieved (KEGG and ModelSEED).	Gene essentiality. Qualitative: growth on seven carbon sources.	Response to nutrient limitations simulating the macrophage environment during the initial infection stages. Prediction of drug targets based on gene essentiality that do not have an ortholog in the human model. The potential target genes are related to pathways such as steroid biosynthesis, oxidative phosphorylation, arginine biosynthesis, purine, pyrimidine, alanine, aspartate, glutamate, glycine, lysine, and glycerophospholipid metabolism, phenylalanine, tyrosine, tryptophan biosynthesis, and transport.	(TEZCAN et al., 2023)

1.3.9. Zygosaccharomyces parabailii

Zygosaccharomyces parabailii is one of the three yeasts in the *Z. bailii sensu lato* clade (*Z. bailii*, *Z. parabailii,* and *Z. pseudobailii*), characterized by high tolerance to low pH and weak acids. Besides its importance as food spoilage, this tolerance might also be beneficial in the production of metabolites, such as lactic acid (PALMA; SÁ-CORREIA, 2019).

The GEM of *Z. parabailii* was reconstructed by FILIPPO et al. (2018) (Table 9). The first draft model was assembled by retrieving the reactions in KEGG. This draft was refined (compartmentalization and reaction balance and directionality) using iOD907 (*K. lactis*) and Yeast7 (*Saccharomyces cerevisiae*), which were also the sources of the biomass equation and ATP requirements. The gap-filling was manually performed using KEGG as the reference. The total and detailed content of proteins, lipids, carbohydrates, amino acids, and fatty acids were experimentally determined from chemostat cultivations with glucose as the carbon source at two dilution rates (0.1 and 0.3 h^{-1}). After validating the ZyPa1 model by comparing *in vivo* and *in silico* biomass yields and gene deletion analysis, they evaluated the pathways involved in acetate consumption in the presence of glucose. They found that acetate contributes to the cytosolic and mitochondrial acetyl-CoA pools, mainly for ATP generation and as a backbone for membrane remodeling.

1.3.10. Lachancea kluyveri

Previously known as *Saccharomyces kluyveri*, the weak Crabtree positive yeast *L. kluyveri* has been the focus of many physiological studies due to its metabolic capabilities, such as using purines and pyrimidines as sole nitrogen sources and respiro-fermentative metabolism. NANDA et al. (2020) used the “Build Fungal Model” in KBase to generate the draft model for *L. kluyveri* with 13 fungal templates (Table 9). The iMM904 model (*S. cerevisiae*) was the standard to correct reactions and retrieve the biomass equation. Gap-filling was manually conducted using KEGG and iMM904 as references. The reconstructed network, iPN730, was validated by comparing *in silico* and *in vivo* growth rates and respiration quotients (CO_2 production/O_2 consumption), as well as qualitative comparisons for anaerobic and aerobic growth on glucose, galactose, maltose, sucrose, melibiose, ethanol, and glycerol. They also applied dFBA for aerobic, semi-aerobic, and hypoxic growth on glucose, considering ethanol and ethyl acetate production. In aerobic conditions, the model neither produced ethanol nor ethyl acetate, while both compounds were produced with oxygen reduction. Consistent with

previous experimental data, ethyl acetate was the major overflow product under hypoxic conditions. However, only ethanol was produced in hypoxia, which might be related to the absence of acetate without oxygen.

Notably, they also reconstructed context-specific models with the Gene Inactivity Moderated by Metabolism and Expression (GIMME) algorithm using expression data on growth with uracil or ammonium as sole nitrogen sources. In GIMME, reactions are inactivated for each condition based on an expression threshold. After the reduction, 78 unique reactions were detected for uracil and 88 for ammonium. Then, they applied FVA and the Jaccard index to determine the differential fluxes between the two conditions. Most reactions with differential flux were related to purine, histidine, riboflavin, and pyrimidine metabolisms and uracil degradation.

1.3.11. Issatchenkia orientalis

Also known as *Pichia kudriavzevii, Candida glycerinogenes,* and *Candida krusei,* the acid and ethanol-tolerant yeast *Issatchenkia orientalis* has been proposed as a platform for organic acid production using low-cost substrates, such as lignocellulosic biomasses. To reconstruct the GEM of *I. orientalis,* SUTHERS et al. (2020) used Yeast 7.6 (*S. cerevisiae*) as a template, then expanded using the KBase function "Build Fungal Model" (Table 9). When necessary, protein localization was assigned with DeepLoc (ALMAGRO ARMENTEROS et al., 2017). Reaction directionality and balance were verified and corrected using MetaCyc and ModelSEED (SEAVER et al., 2021). The model was also updated with information from Yeast 8.3.4 (*S. cerevisiae*). The biomass equation coefficients were determined with chemostat cultivations with glucose as the carbon source and literature data. The GAM and NGAM were estimated using ministat at various dilution rates (0.11-0.65 h^{-1}). The model was evaluated using MEMOTE, and the indicated gaps, thermodynamically infeasible cycles, and inconsistencies were corrected. After corrections, the model, *iIsor*850, presented a score of 84%.

The growth rate and carbon utilization *in silico* were compared to experimental data on minimum media to validate the model. Interestingly, although *I. orientalis* has all genes necessary for xylose utilization and the model predicted growth, it could only be consumed in mixed cultures with glucose at low uptake rates *in vivo*. The gene essentiality predicted in aerobic glucose growth was verified with gene knockouts (such as adenylyl-sulfate kinase and phosphatidylserine synthase) via Clustered Regularly Interspaced Short Palindromic Repeats

(CRISPR)-Cas9 editing. Finally, as a proof of concept, they applied the OptKnock (BURGARD; PHARKYA; MARANAS, 2003) framework to determine targets for succinic acid production under aerobic glucose growth. Succinate dehydrogenase depletion increased the production in all simulated scenarios. However, this modification led to only slight succinate increases *in vivo*. Additionally, they applied this algorithm to a strain with fumarate reductase cultivated in microaerophilic conditions. The better solutions included the simultaneous knockout of pyruvate decarboxylase, aspartate transaminase, valine transaminase, glucose-6-phosphate dehydrogenase, and homoserine dehydrogenase.

1.3.12. Cutaneotrichosporon oleaginosus

The oleaginous yeast *Cutaneotrichosporon oleaginosus* (also known as *Cryptococcus curvatus*) can grow on various byproducts and wastes, including municipal and food industry streams, whey permeates, yeast lysate from the beer industry, crude glycerol from biodiesel processing, and lignocellulosic biomasses (DI FIDIO et al., 2021). PHAM et al. (2021) reconstructed the GEM of *Cutaneotrichosporon oleaginosus* to explore its potential for biofuel production. They used the iNL895 model of *Y. lipolytica* as a scaffold to build the first draft. Lipid pathway curation was based on KEGG and RNA-seq data, the central metabolism was manually curated, the GAM was the same as in iNL895, and the NGAM was retrieved from the literature (Table 9). The correction of gaps and infeasible cycles was manually conducted. Interestingly, they developed a condition-specific biomass equation with experimentally determined composition for proteins, carbohydrates, nucleic acids, and lipids. The biomass equation varied with the carbon/nitrogen (C/N) ratio in the simulated media, considering carbohydrates and nucleic acid constants, while the lipid and protein fractions varied. The C/N ratio was modeled using quadratic regression and previous multi-factorial-guided media optimization. The *i*NP636_*Coleaginosus*_ATCC20509 model was used to evaluate the effects of using a constant or dynamic biomass equation. The dynamic equation led to *in silico* results closer to *in vivo* ones. They also used this equation to estimate the best C/N ratio for glucose, xylose, fructose, sucrose, and glycerol toward growth and lipid production. Finally, they applied FVA to determine the best acetyl-CoA source for lipid production and lipid metabolism regulation, considering different C/N ratios.

1.3.13. Papiliotrema laurentii

Papiliotrema laurentii, previously known as *Cryptococcus laurentii*, has broad biotechnological applications, such as the production of enzymes, surfactants, lipids, biocontrol of phytopathogenic fungi, growth promotion of plants, polyester degradation, and bioremediation. *P. laurentii* grows on various carbon sources, including mono and oligosaccharides, organic acids, glycerol, and fatty acids (DE ALMEIDA et al., 2022). The model of *P. laurentii* was reconstructed by VENTORIM et al. (2022) using the RAVEN toolbox 2.5.3 (Table 9). They used two well-curated models, *rhto*-GEM 1.3.0 (*R. toruloides*) and iYali4.1.2 (*Y. lipolytica*), as templates to build the homology-based draft and imported non-gene-associated reactions from *rhto*-GEM. The biomass equation (carbohydrate and lipid profile and protein content) was experimentally determined using glucose and nitrogen-limited chemostats or retrieved from the literature. Lipid reactions were described using the SLIMEr formalism considering 16:0, 18:0, 18:1, and 18:2 fatty acids (> 5% of the total pool). Gap-filling was conducted using Meneco and the *fillGaps* function in RAVEN with *rhto*-GEM as the reference. The GAM was estimated using growth and glucose uptake rates from bioreactor cultivations, and the NGAM remained the same form *rhto*-GEM. The final network, *papla*-GEM, similar to other models with the SLIMEr formalism, presented a low MEMOTE score (53%). Meanwhile, the model was validated with quantitative measurements considering growth on glucose and xylose batch and chemostat (carbon and nitrogen-limiting) cultures, as well as qualitative evidence on arabinose, mannose, galactose, sucrose, acetate, and lactate. Furthermore, they evaluated the oleaginous phenotype with the environmental version of minimization of metabolic adjustment (eMOMA) approach, designed to represent nitrogen-limiting conditions compared to a reference state (non-limiting) (KIM et al., 2019; SEGRÈ; VITKUP; CHURCH, 2002). Then, the eMOMA-adjusted model was used to predict metabolic engineering targets to improve lipid production via FSEOF. The main targets for glucose and xylose were well-described central metabolism and lipid synthesis genes.

1.3.14. Ustilago maydis

The plant pathogen *Ustilago maydis* is responsible for the corn smut disease and has become an important model for cell biology and physiology, especially long-distance intra/extracellular transport, mitosis, dimorphism, and microtubule organization. *U. maydis* also produces a wide range of metabolites of industrial interest, including organic acids (such as succinate, malate,

and itaconate), erythritol, mannitol, and glycolipids (OLICÓN-HERNÁNDEZ et al., 2019). LIEBAL et al. (2022) used Pathway Tools (KARP et al., 2021) to reconstruct the first GEM of *U. maydis* (Table 9). They applied the PathoLogic component to build a pathway/genome database, which was considered the draft network. The biomass composition was described based on the genome sequence (DNA, RNA, and amino acids) or retrieved from the literature (cell wall and lipid components). The GAM was estimated via optimization of the sum of the squared errors of growth data, and the NGAM was determined with the lowest but not infeasible glucose uptake rate. The iUma22 model presented a MEMOTE score of 57%. The *in silico* performance was qualitatively validated with the Biolog Carbon Microarray, and 96% of *in vivo* growth was correctly predicted. Quantitatively, simulations were compared to previous and new batch growth data with different initial glucose concentrations (50-216 g/L). The growth and substrate uptake rates continuously decreased with increasing glucose concentrations. Interestingly, they assembled a pangenome of five *U. maydis* strains to identify strain-specific differences, which presented 7839 coding genes and 1458 enzymes. The pangenome and iUma22 shared 513 genes. Most unique genes in iUma22 were related to oxidative phosphorylation, while pangenome-unique genes mainly comprised different carbon metabolic pathways.

1.3.15. Cryptococcus neoformans

Cryptococcus neoformans is one the most common opportunistic pathogenic yeasts in humans, especially for those with compromised immune systems. Due to its importance, *C. neoformans* has also become a model system, and several molecular tools have been developed to understand its physiological, morphological, and infectious behavior, from sexual reproduction and dimorphic life cycle to signal transduction and polysaccharide synthesis (RATHORE et al., 2022; SRIKANTA; SANTIAGO-TIRADO; DOERING, 2014).

The GEM of *C. neoformans* was only released in 2023 by TEZCAN et al. (2023). First, they collected the reactions available in KEGG and verified their directionality using ModelSEED (Table 9). Compartmentalization was based on the iMM904 and SpoMBEL1693 models and various databases (e.g., KEGG, UniProt), and the gap-filling was performed manually and focused on dead-end metabolites. The biomass reaction was based on iMM904, and the detailed cell wall and capsule composition were updated. The GAM was not modified from iMM904, and the NGAM lower bound was fixed using an estimated value for pathogenic

organisms (HENSON et al., 2019). The *iCryptococcus* model was validated considering normal glucose growth and nutrient-limiting conditions to simulate the macrophage environment during the initial infection stages. In this case, glucose uptake was blocked, and other carbon sources, such as acetate, fumarate, and succinate, were allowed. Then, they evaluated the impacts of virulence factors, such as using urea as a nitrogen source, on *C. neoformans* growth. For example, they showed that with increased urease activity, ATP production increased, consistent with experimental evidence. Finally, they applied gene essentiality analysis to determine drug targets and found potential genes without human orthologs. Based on DrugBank data (https://go.drugbank.com/), many targets with already used drugs for other diseases or pathogenic fungi infections were assigned, including known drugs directed to ergosterol, chorismate, and 1,3-β-glucan synthesis.

1.3.16. Simultaneous reconstructions of multiple yeast draft models

Some effort has been directed to the simultaneous reconstruction of draft metabolic models for multiple yeast species. PITKÄNEN et al. (2014) developed the comparative metabolic reconstruction framework (CoReCo) to generate multiple GEMs of related species. This approach requires a phylogenetic tree and protein-coding sequences for each species of interest. CoReCo was demonstrated by constructing gapless GEMs for 49 fungal species based on homology, the KEGG database, and network comparisons. One of the advantages of this method is the generation of networks for poorly annotated or evolutionary distant (ancestral) species. Moreover, CORREIA; MAHADEVAN (2020) built a pan-network for 33 fungi with 1553 orthologs, 2759 reactions, and 2251 metabolites using the models of *Schizosaccharomyces pombe*, *Aspergillus niger, Y. lipolytica*, *K. phaffii, Scheffersomyces stipitis, Eremothecium gossypii*, and *Saccharomyces cerevisiae* as scaffolds. This pan-network can be derived into specific GEMs at different taxonomic levels, such as phylum, families, genres, species, or strains.

The biggest simultaneous reconstruction of yeast draft GEMs was conducted by LU et al. (2021)/LU; KERKHOVEN; NIELSEN (2022) with the RAVEN toolbox and included 332 budding yeast genomes (subphylum Saccharomycotina) annotated by SHEN et al. (2018). The pan-GEM was reconstructed using Yeast8 (*S. cerevisiae*), MetaCyc, and KEGG models as templates. During this process, 562 new reactions were added to Yeast8, including all species selected. Then, the pan-GEM was used to generate species-specific GEMs by removing the

unmapped reactions in each species. Finally, the biomass equation from Yeast8 was adjusted with lipid, carbon, protein, DNA, and RNA contents of four main phenotype groups (normal, heat-tolerant, oleaginous, and pathogenic) based on previously published GEMs of representative species form these functional clades.

1.4. Challenges and Perspectives

Although the number of GEMs for non-*Saccharomyces* yeasts has recently grown (Figure 2C), many challenges related to their reconstruction, continuous curation, validation, and incorporation of omics data remain. Few yeasts (e.g., *K. phaffi, Y. lipolytica*, *R. toruloides, Scheffersomyces stipitis*) have models that have been refined over time or novel reconstructions based on previous GEMs of the same species, a process that has been well-established for *Saccharomyces cerevisiae* [see LOPES; ROCHA (2017) and CHEN; LI; NIELSEN (2022)]. Additionally, important physiological data directly affecting the simulation capacity and ranges have not been deeply explored, such as detailed biomass compositions for different growth conditions; specific GAM and NGAM estimates; growth, substrate, oxygen, and product exchange rates over multiple conditions; ^{13}C analysis for internal fluxes. In most reconstructions, biomass parameters and ATP requirements were considered the same as *S. cerevisiae*, incompletely modified, or retrieved from the template models. The absence of species-specific data might hamper GEM applications, such as predicting the best cultivation conditions for producing a specific metabolite, comparing metabolic capabilities between different organisms, and describing metabolic functions. This lack of basic data is also represented by the gap in gene essentiality analysis for each species *in vivo*, which are crucial not only for model validation but also for determining possible engineering targets, trustful drug targets for pathogens, and a better understanding of yeast physiology. Moreover, most yeast GEMs have been reconstructed for Ascomycetes (including the pan-GEM) (Figure 2B), leaving a gap for the reconstruction of Basidiomycetes.

In contrast to *S. cerevisiae*, the integration of omics data and generation of other model paradigms has also been scarce for non-conventional yeast GEMs (CHEN; LI; NIELSEN, 2022). For example, no thermodynamic-constrained model is available for non-*Saccharomyces*, partly due to the lack of data for parametrization in these organisms. Besides, although omics studies have been conducted for non-conventional yeasts, little has been done regarding data integration into GEMs. In the reconstructions described in this review, only SUTHERS et al.

(2020) (*I. orientalis*) and MARCIŠAUSKAS; JI; NIELSEN (2019) (*K. marxianus*) incorporated RNA-seq data to generate context-specific models. Regarding proteomics, the GECKO toolbox was applied to reconstruct enzymatic-constrained GEMs (ecGEMs) of only *Y. lipolytica* and *K. marxianus* to incorporate the proteome under standard and stressful conditions (DOMENZAIN et al., 2022). Thus, although GEMs have been proposed as great frameworks for integrating omics data, this strategy has been poorly explored for non-conventional yeasts.

For the next few years, the availability of GEMs should increase even more due to the rising number of genome annotations and the development and improvement of tools to facilitate network generation, curation, and validation, such as COBRA, RAVEN, *merlin*, KBase, MEMOTE, and ModelSEED. These tools, which are being continuously improved, can also be used in developing and curating exiting GEMs, which must be systematically established for non-conventional organisms. As such, it is expected to expand our understanding of yeast metabolic diversity and regulation of yeast metabolism. Importantly, the reconstruction of novel GEMs might be facilitated not only by automatic tools but also by the works of PITKÄNEN et al. (2014), CORREIA; MAHADEVAN (2020), and LU et al. (2021)/LU; KERKHOVEN; NIELSEN (2022), whose models can be used as scaffolds/templates to generate well-curated models. Model quality highly depends on *in vivo* and *in vitro* data. Thus, works focusing on yeast physiology, such as growth kinetics on different substrates and biomass composition, should be encouraged to allow the reconstruction of high-quality GEMs and their use in systems metabolic engineering approaches. Hence, non-conventional yeasts are expected to be employed as new biotechnological, medical, and agronomical platforms.

Moreover, GEMs of yeasts that have different models, such as K. phaffii and Y. lipolytica, came out in the same year, indicating a communication impairment between groups that work with GEM reconstruction and/or the same organism. This might delay improvements in yeast models or even prevent them and lead to unnecessary double efforts. The yeast consensus model (yeast-GEM) for S. cerevisiae is a good example of how collaborative work between multiple research groups studying the same organism can improve, develop, and boost GEMs. Hence, besides reconstructing novel yeast models, efforts should also be directed to create or strengthen partnerships for future consensus/collaborative non-conventional yeast GEMs.

1.5. Abbreviations

GEMs: Genome-scale metabolic models; COBRA Toolbox: Constraint-Based Reconstruction and Analysis Toolbox; ATP: adenosine triphosphate; FBA: flux balance analysis; SCP: single-cell protein; HAS: human serum albumin; hSOD: human superoxide dismutase; Basic Local Alignment Search Tool (BLAST); FVA: flux variability analysis; dFBA: dynamic flux balance analysis; FSEOF: Flux Scanning based on Enforced Objective Function; KEGG: Kyoto Encyclopedia of Genes and Genomes; GAM: growth-associated maintenance; NGAM: non-growth associated maintenance; FAB: Human monoclonal antibody 3H6 Fab fragment; PPP: pentose phosphate pathway; TCA cycle: Tricarboxylic acid cycle; EMP pathway: Embden-Meyerhof-Parnas pathway; TAG: triacylglycerol; RNA-seq: RNA sequencing; GRAS: generally regarded as safe; NADPH: nicotinamide adenine dinucleotide phosphate; DCA: di-carboxylic acid; BCAA: branched-chain amino acid; DDDA: dodecanedioic acid; CMA: cofactor modification analysis; tSOT: transcriptomic-based strain optimization tool; GDLS: genetic design by local search; KGML: KEGG markup language; KAAS: KEGG Automatic Annotation Server; TCDB: Transporter Classification Database; RPMI: Roswell Park Memorial Institute; SLIMEr: Split Lipids Into Measurable Entities; PDC: pyruvate decarboxylase; ALD: aldehyde dehydrogenase; ACS: acetyl-CoA synthetase; NADH: nicotinamide adenine dinucleotide; SID: system identification; PCA: Principal Component Analysis; XR: xylose reductase; XDH: xylitol dehydrogenase; P/O ratio: Phosphorus/Oxygen ratio; RING: Reconciliation of *In silico*/*in vivo* mutaNt Growth; GIMME: Gene Inactivity Moderated by Metabolism and Expression; APGC: Analysis of production and growth coupling; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; C/N ratio: carbon/nitrogen ratio; eMOMA: Environmental version of minimization of metabolic adjustment; CoReCo: Comparative metabolic reconstruction framework; ecGEM: enzymatic-constrained genome-scale metabolic model.

1.6. ORCID

E.L.M. Almeida: 0000-0002-3526-0975

E.J. Kerkhoven: 0000-0002-3593-5792

W.B. Silveira: 0000-0001-7869-8144

1.7. Authorship contribution statement

E.L.M. Almeida: Conceptualization, Writing - original draft. **E.J. Kerkhoven**: Writing - review & editing **W.B. Silveira**: Conceptualization, Writing - review & editing.

1.8. Acknowledgements

The authors thank the Universidade Federal de Viçosa (UFV), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq Finance Code 140538/2021-6), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) for their financial support for the research and the scholarship. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) —Finance Code 001.

1.9. Declaration of Competing Interest

The authors declare that they have no competing interests.

1.10. References

ABDEL-MAWGOUD, A. M. et al. Metabolic engineering in the host Yarrowia lipolytica. **Metabolic Engineering**, v. 50, n. July, p. 192–208, 2018.

AGREN, R. et al. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. **PLoS Computational Biology**, v. 9, n. 3, 2013.

ALMAGRO ARMENTEROS, J. J. et al. DeepLoc: prediction of protein subcellular localization using deep learning. **Bioinformatics (Oxford, England)**, v. 33, n. 21, p. 3387–3395, 2017.

ARKIN, A. P. et al. KBase: The United States department of energy systems biology knowledgebase. **Nature Biotechnology**, v. 36, n. 7, p. 566–569, 2018.

BALAGURUNATHAN, B. et al. Reconstruction and analysis of a genome-scale metabolic model for Scheffersomyces stipitis. **Microbial Cell Factories**, v. 11, n. 1, p. 27, 23 dez. 2012.

BECKER, S. A. et al. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. **Nature Protocols**, v. 2, n. 3, p. 727–738, 2007.

BERMAN, J. Candida albicans. **Current Biology**, v. 22, n. 16, p. R620–R622, 2012.

BERNAUER, L. et al. Komagataella phaffii as Emerging Model Organism in Fundamental Research. **Frontiers in Microbiology**, v. 11, n. January, p. 1–16, 2021.

BOEKHOUT, T. et al. **Trends in yeast diversity discovery**. [s.l.] Springer Netherlands, 2022. v. 114

BURGARD, A. P.; PHARKYA, P.; MARANAS, C. D. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. **Biotechnology and Bioengineering**, v. 84, n. 6, p. 647–657, 2003.

BURGARD, A. P.; VAIDYARAMAN, S.; MARANAS, C. D. Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments. **Biotechnology Progress**, v. 17, n. 5, p. 791–797, 2001.

CANKORUR-CETINKAYA, A.; DIKICIOGLU, D.; OLIVER, S. G. Metabolic modeling to identify engineering targets for Komagataella phaffii: The effect of biomass composition on gene target identification. **Biotechnology and Bioengineering**, v. 114, n. 11, p. 2605–2615, 2017.

CAPELA, J. et al. merlin , an improved framework for the reconstruction of high-quality genome-scale metabolic models. **Nucleic Acids Research**, v. 50, n. 11, p. 1–15, 24 jun. 2022.

CASPETA, L. et al. Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. **BMC Systems Biology**, v. 6, 2012.

CAVALLO, E. et al. Yarrowia lipolytica: A model yeast for citric acid production. **FEMS Yeast Research**, v. 17, n. 8, p. 1–16, 2017.

CHANG, A. et al. BRENDA, the ELIXIR core data resource in 2021: New developments and updates. **Nucleic Acids Research**, v. 49, n. D1, p. D498–D508, 2021.

CHEN, Y.; LI, F.; NIELSEN, J. Genome-scale modeling of yeast metabolism: retrospectives and perspectives. **FEMS Yeast Research**, v. 22, n. 1, p. 1–9, 29 jan. 2022.

CHOI, H. S. et al. In silico identification of gene amplification targets for improvement of lycopene production. **Applied and Environmental Microbiology**, v. 76, n. 10, p. 3097–3105, 2010.

CHUNG, B. K. S. et al. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. **Microbial Cell Factories**, v. 9, p. 1–15, 2010.

CONSORTIUM, T. U. UniProt: the Universal Protein Knowledgebase in 2023. **Nucleic acids research**, v. 51, n. D1, p. D523–D531, 2023.

CORREIA, K.; MAHADEVAN, R. Pan-Genome-Scale Network Reconstruction: Harnessing Phylogenomics Increases the Quantity and Quality of Metabolic Models. **Biotechnology Journal**, v. 15, n. 10, p. 1–10, 2020.

DAMIANI, A. L. et al. Comprehensive evaluation of two genome-scale metabolic network models for Scheffersomyces stipitis. **Biotechnology and Bioengineering**, v. 112, n. 6, p. 1250–1262, 2015.

DE ALMEIDA, E. L. M. et al. *Papiliotrema laurentii:* general features and biotechnological applications. **Applied Microbiology and Biotechnology**, v. 106, n. 21, p. 6963–6976, 5 out. 2022.

DI FIDIO, N. et al. Cutaneotrichosporon oleaginosus: A versatile whole-cell biocatalyst for the production of single-cell oil from agro-industrial wastes. **Catalysts**, v. 11, n. 11, 2021.

DIAS, O. et al. iOD907, the first genome-scale metabolic model for the milk yeast Kluyveromyces lactis. **Biotechnology Journal**, v. 9, n. 6, p. 776–790, 2014.

DINH, H. V. et al. A comprehensive genome-scale model for Rhodosporidium toruloides IFO0880 accounting for functional genomics and phenotypic data. **Metabolic Engineering Communications**, v. 9, n. August, p. e00101, 2019.

DOMENZAIN, I. et al. Evaluating accessibility, usability and interoperability of genome-scale metabolic models for diverse yeasts species. **FEMS Yeast Research**, v. 21, n. 1, p. 1–9, 2021.

DOMENZAIN, I. et al. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. **Nature Communications**, v. 13, n. 1, p. 1–13, 2022.

EBRAHIM, A. et al. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. **BMC Systems Biology**, v. 7, n. 1, p. 74, 2013.

FANG, X.; LLOYD, C. J.; PALSSON, B. O. Reconstructing organisms in silico: genome-scale models and their emerging applications. **Nature Reviews Microbiology**, v. 18, n. December, p. 23–26, 2020.

FILIPPO, M. D. et al. Genome-scale metabolic reconstruction of the stress-tolerant hybrid yeast Zygosaccharomyces parabailii. **bioRxiv**, p. 373621, 2018.

GEIJER, C.; LEDESMA-AMARO, R.; TOMAS-PEJO, E. Unraveling the potential of non-conventional yeasts in biotechnology. **FEMS Yeast Research**, v. 22, n. 1, p. 1–6, 2022.

GRANT, C. E.; BAILEY, T. L.; NOBLE, W. S. FIMO: Scanning for occurrences of a given motif. **Bioinformatics**, v. 27, n. 7, p. 1017–1018, 2011.

GROENEWALD, M. et al. Yarrowia lipolytica: Safety assessment of an oleaginous yeast with a great industrial potential. **Critical Reviews in Microbiology**, v. 40, n. 3, p. 187–206, 2014.

GU, D. et al. IdealKnock: A framework for efficiently identifying knockout strategies leading to targeted overproduction. **Computational Biology and Chemistry**, v. 61, p. 229–237, 2016.

HEIRENDT, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. **Nature Protocols**, v. 14, n. 3, p. 639–702, 2019.

HENSON, M. A. et al. Metabolic Modeling of Cystic Fibrosis Airway Communities Predicts Mechanisms of Pathogen Dominance. **mSystems**, v. 4, n. 2, 2019.

HILLIARD, M. et al. Elucidating redox balance shift in Scheffersomyces stipitis' fermentative metabolism using a modified genome-scale metabolic model. **Microbial Cell Factories**, v. 17, n. 1, p. 1–15, 2018.

HORTON, P. et al. WoLF PSORT: Protein localization predictor. **Nucleic Acids Research**, v. 35, n. SUPPL.2, p. 585–587, 2007.

HOU, B. K. et al. BioSilico: An integrated metabolic database system. **Bioinformatics**, v. 20, n. 17, p. 3270–3272, 2004.

IRANI, Z. A. et al. Genome-scale metabolic model of Pichia pastoris with native and humanized glycosylation of recombinant proteins. **Biotechnology and Bioengineering**, v. 113, n. 5, p. 961–969, 2016.

JIAN, X. et al. In silico identification of gene amplification targets based on analysis of production and growth coupling. **BioSystems**, v. 145, p. 1–8, 2016.

KANEHISA, M. ;; GOTO, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27–30, 1 jan. 2000.

KARIM, A.; GERLIANI, N.; AÏDER, M. Kluyveromyces marxianus: An emerging yeast cell factory for applications in food and biotechnology. **International Journal of Food Microbiology**, v. 333, n. May, p. 108818, 2020.

KARP, P. D. et al. The BioCyc collection of microbial genomes and metabolic pathways. **Briefings in Bioinformatics**, v. 20, n. 4, p. 1085–1093, 2018.

KARP, P. D. et al. Pathway Tools version 23.0 update: Software for pathway/genome informatics and systems biology. **Briefings in Bioinformatics**, v. 22, n. 1, p. 109–126, 2021.

KAVŠCEK, M. et al. Optimization of lipid production with a genome-scale model of Yarrowia lipolytica. **BMC Systems Biology**, v. 9, n. 1, p. 1–13, 2015a.

KAVŠCEK, M. et al. Optimization of lipid production with a genome-scale model of Yarrowia lipolytica. **BMC Systems Biology**, v. 9, n. 1, p. 1–13, 2015b.

KERKHOVEN, E. J. et al. Regulation of amino-acid metabolism controls flux to lipid accumulation in yarrowia lipolytica. **npj Systems Biology and Applications**, v. 2, n. November 2015, p. 1–7, 2016.

KIM, J. et al. Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in Rhodosporidium toruloides. **Frontiers in Bioengineering and Biotechnology**, v. 8, n. January, 8 jan. 2021.

KIM, M. et al. Transcriptomics-based strain optimization tool for designing secondary metabolite overproducing strains of Streptomyces coelicolor. **Biotechnology and Bioengineering**, v. 113, n. 3, p. 651–660, 2016.

KIM, M. et al. In silico identification of metabolic engineering strategies for improved lipid production in yarrowia lipolytica by genome-scale metabolic modeling. **Biotechnology for Biofuels**, v. 12, n. 1, p. 1–14, 2019.

KING, Z. A. et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. **Nucleic Acids Research**, v. 44, n. D1, p. D515–D522, 2016.

LACHANCE, J. C. et al. BOFDAT: Generating biomass objective functions for genome-scale metabolic models from experimental data. **PLoS Computational Biology**, v. 15, n. 4, p. 1–20, 2019.

LAGOA, D. et al. TranSyT, an innovative framework for identifying transport systems. **bioRxiv doi: 10.1101/2021.04.29.441738**, n. 7, 2021.

LAKSHMANAN, M. et al. Cofactor modification analysis: A computational framework to identify cofactor specificity engineering targets for strain improvement. **Journal of Bioinformatics and Computational Biology**, v. 11, n. 6, p. 1–19, 2013.

LEWIS, N. E.; NAGARAJAN, H.; PALSSON, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. **Nature Reviews Microbiology**, v. 10, n. 4, p. 291–305, 2012.

LIEBAL, U. W. et al. Genome-scale model reconstruction of the methylotrophic yeast Ogataea polymorpha. **BMC Biotechnology**, v. 21, n. 1, p. 23, 15 dez. 2021.

LIEBAL, U. W. et al. Ustilago maydis Metabolic Characterization and Growth Quantification with a Genome-Scale Metabolic Model. **Journal of Fungi**, v. 8, n. 5, p. 524, 20 maio 2022.

LIEVEN, C. et al. MEMOTE for standardized genome-scale metabolic model testing. **Nature Biotechnology**, v. 38, n. 3, p. 272–276, 2020.

LOIRA, N. et al. A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. **BMC Systems Biology**, v. 6, n. 1, p. 35, 2012.

LOIRA, N.; ZHUKOVA, A.; SHERMAN, D. J. Pantograph: A template-based method for genome-scale metabolic model reconstruction. **Journal of Bioinformatics and Computational Biology**, v. 13, n. 2, p. 1–19, 2015.

LOPES, H.; ROCHA, I. Genome-scale modeling of yeast: chronology, applications and critical perspectives. **FEMS yeast research**, v. 17, n. 5, p. 1–14, 2017.

LU, H. et al. A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. **Nature Communications**, v. 10, n. 1, p. 3586, 2019.

LU, H. et al. Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection. **Molecular Systems Biology**, v. 17, n. 10, p. 1–23, out. 2021.

LU, H.; KERKHOVEN, E. J.; NIELSEN, J. A Pan-Draft Metabolic Model Reflects Evolutionary Diversity across 332 Yeast Species. **Biomolecules**, v. 12, n. 11, p. 1–15, 2022.

LUN, D. S. et al. Large-scale identification of genetic design strategies using local search. **Molecular Systems Biology**, v. 5, n. 296, p. 1–8, 2009.

MAHADEVAN, R.; EDWARDS, J. S.; DOYLE, F. J. Dynamic Flux Balance Analysis of diauxic growth in Escherichia coli. **Biophysical Journal**, v. 83, n. 3, p. 1331–1340, 2002.

MARCIŠAUSKAS, S.; JI, B.; NIELSEN, J. Reconstruction and analysis of a Kluyveromyces marxianus genome-scale metabolic model. **BMC Bioinformatics**, v. 20, n. 1, p. 1–9, 2019.

MISHRA, P. et al. Genome-scale metabolic modeling and in silico analysis of lipid accumulating yeast Candida tropicalis for dicarboxylic acid production. **Biotechnology and Bioengineering**, v. 113, n. 9, p. 1993–2004, 2016.

MISHRA, P. et al. Genome-scale model-driven strain design for dicarboxylic acid production in Yarrowia lipolytica. **BMC Systems Biology**, v. 12, n. Suppl 2, 2018.

NANDA, P. et al. Reconstruction and analysis of genome-scale metabolic model of weak Crabtree positive yeast Lachancea kluyveri. **Scientific Reports**, v. 10, n. 1, p. 1–17, 2020.

NAVARRETE, C.; L. MARTÍNEZ, J. Non-conventional yeasts as superior production platforms for sustainable fermentation based bio-manufacturing processes. **AIMS Bioengineering**, v. 7, n. 4, p. 289–305, 2020.

OLICÓN-HERNÁNDEZ, D. R. et al. New Insights of Ustilago maydis as Yeast Model for Genetic and Biotechnological Research: A Review. **Current Microbiology**, v. 76, n. 8, p. 917–926, 2019.

ORTH, J. D.; THIELE, I.; PALSSON, B. Ø. O. What is flux balance analysis? **Nature Biotechnology**, v. 28, n. 3, p. 245–248, mar. 2010.

PALMA, M.; SÁ-CORREIA, I. Physiological Genomics of the Highly Weak-Acid-Tolerant Food Spoilage Yeasts of Zygosaccharomyces bailii sensu lato. **Progress in molecular and subcellular biology**, v. 58, p. 85–109, 2019.

PAN, P.; HUA, Q. Reconstruction and In Silico Analysis of Metabolic Network for an Oleaginous Yeast, Yarrowia lipolytica. **PLoS ONE**, v. 7, n. 12, p. 1–11, 2012.

PHAM, N. et al. Genome-scale metabolic modeling underscores the potential of Cutaneotrichosporon oleaginosus ATCC 20509 as a cell factory for biofuel production. **Biotechnology for Biofuels**, v. 14, n. 1, p. 1–17, 2021.

PITKÄNEN, E. et al. Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. **PLoS Computational Biology**, v. 10, n. 2, 2014.

PRIGENT, S. et al. Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks. **PLOS Computational Biology**, v. 13, n. 1, p. e1005276, 27 jan. 2017.

RANGANATHAN, S.; SUTHERS, P. F.; MARANAS, C. D. OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. **PLoS Computational Biology**, v. 6, n. 4, 2010.

RATHORE, S. S. et al. A holistic review on Cryptococcus neoformans. **Microbial Pathogenesis**, v. 166, n. October 2021, p. 105521, 2022.

REITZ, M. et al. Enabling the exploration of biochemical pathways. **Organic and Biomolecular Chemistry**, v. 2, n. 22, p. 3226–3237, 2004.

REN, Q.; KANG, K. H.; PAULSEN, I. T. TransportDB: A relational database of cellular membrane transport systems. **Nucleic Acids Research**, v. 32, n. DATABASE ISS., p. 284–288, 2004.

ROCHA, I. et al. OptFlux: an open-source software platform for. **BMC systems biology**, v. 4, n. 45, 2010.

SAIER, M. H. et al. The transporter classification database (TCDB): 2021 update. **Nucleic Acids Research**, v. 49, n. D1, p. D461–D467, 2021.

SÁNCHEZ, B. J. et al. SLIMEr: Probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. **BMC Systems Biology**, v. 13, n. 1, p. 1–9, 2019.

SANTOS, S.; ROCHA, I. Development of computational methods for determination of biomass composition and evaluation of its impact in genome-scale models prediction. n. October, 2013.

SATISH KUMAR, V.; DASIKA, M. S.; MARANAS, C. D. Optimization based automated curation of metabolic reconstructions. **BMC Bioinformatics**, v. 8, p. 1–16, 2007.

SAULS, J. T.; BUESCHER, J. M. Assimilating genome-scale metabolic reconstructions with modelBorgifier. **Bioinformatics**, v. 30, n. 7, p. 1036–1038, 2014.

SEAVER, S. M. D. et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. **Nucleic Acids Research**, v. 49, n. D1, p. D575–D588, 8 jan. 2021.

SEGRÈ, D.; VITKUP, D.; CHURCH, G. M. Analysis of optimality in natural and perturbed metabolic networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 23, p. 15112–15117, 2002.

SHEN, X. X. et al. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. **Cell**, v. 175, n. 6, p. 1533- 1545.e20, 2018.

SOHN, S. B. et al. Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. **Biotechnology Journal**, v. 5, n. 7, p. 705–715, 2010.

SOHN, S. B. et al. Genome-scale metabolic model of the fission yeast Schizosaccharomyces pombe and the reconciliation of in silico/in vivo mutant growth. **BMC Systems Biology**, v. 6, 2012.

SPOHNER, S. C. et al. Kluyveromyces lactis: An emerging tool in biotechnology. **Journal of Biotechnology**, v. 222, p. 104–116, 2016.

SRIKANTA, D.; SANTIAGO-TIRADO, F. H.; DOERING, T. L. Cryptococcus neoformans : historical curiosity to modern pathogen. **Yeast**, v. 31, p. 47–60, 2014.

SUTHERS, P. F. et al. Genome-scale metabolic reconstruction of the non-model yeast Issatchenkia orientalis SD108 and its application to organic acids production. **Metabolic Engineering Communications**, v. 11, n. June, p. e00148, 1 dez. 2020.

TEZCAN, E. F. et al. Comprehensive genome-scale metabolic model of the human pathogen Cryptococcus neoformans: A platform for understanding pathogen metabolism and identifying new drug targets. **Frontiers in Bioinformatics**, v. 3, n. January, p. 1–16, 2023.

THIELE, I.; PALSSON, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. **Nature Protocols**, v. 5, n. 1, p. 93–121, 2010.

TIUKOVA, I. A. et al. Genome-scale model of Rhodotorula toruloides metabolism. **Biotechnology and Bioengineering**, n. August, p. bit.27162, 22 set. 2019.

TOMÀS-GAMISANS, M.; FERRER, P.; ALBIOL, J. Integration and validation of the genome-scale metabolic models of Pichia pastoris: A comprehensive update of protein glycosylation pathways, lipid and energy metabolism. **PLoS ONE**, v. 11, n. 1, p. 1–24, 2016.

TOMÀS-GAMISANS, M.; FERRER, P.; ALBIOL, J. Fine-tuning the P. pastoris iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources. **Microbial Biotechnology**, v. 11, n. 1, p. 224–237, 2018.

VENTORIM, R. Z. et al. Genome-scale metabolic model of oleaginous yeast Papiliotrema laurentii. **Biochemical Engineering Journal**, v. 180, 1 mar. 2022.

VIANA, R. et al. Genome-Scale Metabolic Model of the Human Pathogen Candida albicans: A Promising Platform for Drug Target Prediction. **Journal of Fungi**, v. 6, n. 3, p. 171, 11 set. 2020.

VIANA, R. et al. A Genome-Scale Metabolic Model for the Human Pathogen Candida Parapsilosis and Early Identification of Putative Novel Antifungal Drug Targets. **Genes**, v. 13, n. 2, p. 1–13, 2022.

WANG, H. et al. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. **PLoS Computational Biology**, v. 14, n. 10, p. 1–17, 2018.

WEI, S. et al. Reconstruction of genome-scale metabolic model of Yarrowia lipolytica and its application in overproduction of triacylglycerol. **Bioresources and Bioprocessing**, v. 4, n. 1, 2017.

XU, N. et al. Reconstruction and analysis of the genome-scale metabolic network of Candida glabrata. **Mol. BioSyst.**, v. 9, n. 2, p. 205–216, 2013.

XU, Y.; HOLIC, R.; HUA, Q. Comparison and Analysis of Published Genome-scale Metabolic Models of Yarrowia lipolytica. **Biotechnology and Bioprocess Engineering**, v. 25, n. 1, p. 53–61, 2020.

YE, R. et al. Comprehensive reconstruction and evaluation of Pichia pastoris genome-scale metabolic model that accounts for 1243 ORFs. **Bioresources and Bioprocessing**, v. 4, n. 1, 2017.

YU, C. S. et al. CELLO2GO: A web server for protein subCELlular lOcalization prediction with functional gene ontology annotation. **PLoS ONE**, v. 9, n. 6, 2014.

ZHANG, C. et al. Logical transformation of genome-scale metabolic models for gene level applications and analysis. **Bioinformatics**, v. 31, n. 14, p. 2324–2331, 2015.

ZORRILLA, F.; KERKHOVEN, E. J. Reconstruction of Genome-Scale Metabolic Model for Hansenula polymorpha Using RAVEN. Em: **Methods in Molecular Biology**. [s.l: s.n.]. v. 2513p. 271–290.

CHAPTER 2 - *lista*-GEM: THE GENOME-SCALE METABOLIC RECONSTRUCTION OF *Lipomyces starkeyi*

Published as a pre-print: https://doi.org/10.1101/2023.09.25.559328

2.1. Abstract

Oleaginous yeasts cultivation in low-cost substrates is an alternative for more sustainable production of lipids and oleochemicals. *Lipomyces starkeyi* accumulates high amounts of lipids from different carbon sources, such as glycerol, and glucose and xylose (lignocellulosic sugars). Systems metabolic engineering approaches can further enhance its capabilities for lipid production, but no genome-scale metabolic networks have been reconstructed and curated for *L. starkeyi*. Herein, we propose *lista*-GEM, the first genome-scale metabolic model of *L. starkeyi*. We reconstructed the model using two high-quality models of oleaginous yeasts as templates and further curated the model to reflect the metabolism of *L. starkeyi*. We simulated phenotypes and predicted flux distributions in good accordance with experimental data. We also predicted targets to improve lipid production in glucose, xylose, and glycerol. The phase plane analysis indicated that the carbon availability affected lipid production more than oxygen availability. We found that the maximum lipid production in glucose and xylose required more oxygen than glycerol. Enzymes related to lipid synthesis in the endoplasmic reticulum were the main targets to improve lipid production: stearoyl-CoA desaturase, fatty-acyl-CoA synthase, diacylglycerol acyltransferase, and glycerol-3-phosphate acyltransferase. The glycolytic genes encoding pyruvate kinase, enolase, phosphoglycerate mutase, glyceraldehyde-3-phosphate dehydrogenase, and phosphoglycerate kinase were predicted as targets for overexpression. Pyruvate decarboxylase, acetaldehyde dehydrogenase, acetyl-CoA synthetase, adenylate kinase, inorganic diphosphatase, and triose-phosphate isomerase were predicted only when glycerol was the carbon source. Therefore, we demonstrated that *lista*-GEM provides multiple metabolic engineering targets to improve lipid production by *L. starkeyi* using carbon sources from agricultural and industrial wastes.

Keywords: Metabolic modeling; Oleaginous yeasts; Metabolic engineering; Lipid production; Circular economy

2.1. Introduction

The optimization of bioprocesses allied with advances in genetic and metabolic engineering have enabled considerable advances in the production of lipids using oleaginous yeasts (SAHA; MUKHOPADHYAY, 2021; XU et al., 2016). These yeasts can accumulate at least 20% of their dry biomass as lipids, especially triacylglycerols (TAGs) (SALVADOR LÓPEZ et al., 2022). The requirement for sustainable sources of lipids for the production of oleochemicals, biodiesel, and human nutrition has boosted research with oleaginous yeasts able to use agricultural, industrial, and urban wastes as substrates for the development of bioprocesses (ABELN; CHUCK, 2021; SPAGNUOLO et al., 2019).

Lipomyces starkeyi is an oleaginous yeast capable of growing and producing lipids, ranging from 26 to 55% of its dry biomass, using a diverse range of carbon sources, such as glucose, galactose, arabinose, xylose, glycerol, mannose, cellobiose, and sucrose (SMITH; KURTZMAN, 2011; ZHANG et al., 2022). Its growth and lipid production have been demonstrated in lignocellulosic biomasses, including corn stover (POMRANING et al., 2019), wheat straw (YU et al., 2011), lignin derivatives (PUTRA et al., 2023); glycerol (MARUYAMA et al., 2018; LIU et al., 2017); and sewage sludge (ANGERBAUER et al., 2008). Importantly, *L. starkeyi* also tolerates inhibitors found in lignocellulosic hydrolysates, such as hydroxymethylfurfural, furfural, and phenolic compounds in synthetic media (PUTRA et al., 2023; RAHMAN et al., 2017) and detoxified wheat straw hydrolysate with high acetic acid concentration (4.2 g/L) (YU et al., 2011). Besides, *L. starkeyi* can use levoglucosan, a major product from lignocellulose pyrolysis, as a carbon source (NING et al., 2008).

Both lipid production and accumulation by *L. starkeyi* take place under nitrogen limitation conditions. This leads to an accumulation of citrate inside the cell, which is then converted to acetyl-CoA, kick-starting the fatty acid biosynthesis (TAKAKU et al., 2020). In contrast to other oleaginous yeasts, the malic enzyme of *L. starkeyi* can use both NAD^+ and $NADP^+$ as cofactors but prefers NAD^+ (TANG et al., 2010). Thus, there is an additional requirement for $NADP^+$ from the pentose phosphate pathway to drive fatty acid biosynthesis.

Efforts to engineer *L. starkeyi* strains can benefit from systems biology approaches related to genome-scale metabolic models (GEMs). These models aim to reconstruct and summarize the metabolic network of an organism based on its genome annotation, which is useful for both helping understand its physiology and for metabolic engineering endeavors (YE et al., 2022). GEMs are mathematically formalized as an optimization problem, where a

metabolic objective is maximized or minimized given the assumption of steady-state metabolism and constraints on the uptake of substrates and excretion of products (ORTH et al., 2010). There have been many GEM reconstructions for oleaginous yeasts, such as *Yarrowia lipolytica* (KAVŠČEK et al., 2015; KERKHOVEN et al., 2016; LOIRA et al., 2012; MISHRA et al., 2018; PAN; HUA, 2012; Wei et al., 2017), *Rhodotorula toruloides* (DINH et al., 2019; KIM et al., 2021; TIUKOVA et al., 2019), *Papiliotrema laurentii* (VENTORIM et al., 2022), and *Cutaneotrichosporon oleaginosus* (PHAM et al., 2021). These models have been applied to predict essential genes and the use of different carbon and nitrogen sources, as well as better cultivation strategies and metabolic engineering targets. Currently, only a small-scale metabolic model of *L. starkeyi* is available, which limits the integration of omics data and the prediction of metabolic engineering targets that are not included in the central metabolism (ZHOU et al., 2021). To the best of our knowledge, a curated GEM for *L. starkeyi* was still not reconstructed. Here, we present the first genome-scale reconstruction of *L. starkeyi*, termed *lista*-GEM. We show that *lista*-GEM successfully captures the growth and lipid-producing phenotype of *L. starkeyi* and, therefore, is a useful platform for *in silico* metabolic engineering of this yeast.

2.3 Material and Methods

2.3.1 Draft reconstruction and lipid metabolism

For the first draft of the genome-scale metabolic reconstruction of *L. starkeyi,* denominated *lista*-GEM, we used two well-curated GEMs as templates: *Y. lipolytica* iYali 4.1.2 (KERKHOVEN et al., 2016) and *R. toruloides rhto*-GEM 1.3.0 (TIUKOVA et al., 2019). First, we identified the reactions from orthologs between the *L. starkeyi* genome NRRL Y-11557 (NCBI ID: 10576) and the *Y. lipolytica* or *R. toruloides* using bidirectional BLASTp (MADDEN, 2013). We considered as orthologs the genes with e-value $< 1 \times 10^{-20}$, identity > 35%, and alignment length > 150 bp. We excluded the reactions in iYali that were already present in *rhto*-GEM or that simplified lipid metabolism. Then, we retrieved the pseudo-reactions (e.g., biomass formation and exchange reactions) from *rhto*-GEM. We performed the reconstruction steps using the RAVEN Toolbox 2.7.9 (WANG et al., 2018) in MATLAB (The MathWorks Inc., Natick, Massachusetts).

To represent the lipid metabolism in *lista*-GEM, we used the Split Lipids Into Measurable Entities (SLIMEr) formalism (SÁNCHEZ et al., 2019), which describes lipids by splitting them into their basic components, such as pseudo-reactions that describe both the lipid

classes and the acyl chain distributions. Here, we incorporated the following acyl chains of biotechnological importance: 16:0, 16:1, 18:0, 18:1, 18:2, 18.3.

2.3.2. Biomass composition

From the total content of lipids, proteins, carbohydrates, RNA, and DNA retrieved from experimental measurements (ANSCHAU et al., 2014; MATSUZAWA et al., 2018; PROBST; VADLANI, 2015), we updated the biomass equation of the *rhto*-GEM template and used it for *lista*-GEM. We also updated the biomass composition using data from glucose continuous cultures at a dilution rate of 0.06 h^{-1} (ANSCHAU et al., 2014). We calculated the distribution of deoxyribonucleotides based on the GC content (47%) of *L. starkeyi* genome, as well as the sum of mRNAs and ncRNAs. For the amino acid distribution, we calculated it from the amino acid composition of translated coding sequences. We collected the contribution of triacylglycerols (TAGs), sterols, free FAs, phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol (PI), phosphatidylglycerol (PG), phosphatidylserine (PS), cardiolipin, and diacylglycerols (DAGs) from PROBST; VADLANI (2015) and UZUKA et al. (1974). Considering data from CALVEY et al. (2016), MATSUZAWA et al. (2018), and TAKAKU et al. (2020), we adjusted the FA profile for the chains 16:0, 16:1, 18:0, 18:1, 18:2, and 18.3. The calculation procedures used to define the stoichiometric coefficients are provided in the *lista*-GEM documentation biomassCalculations.xlsx file available in the GitHub repository and Zenodo archive (See Data availability).

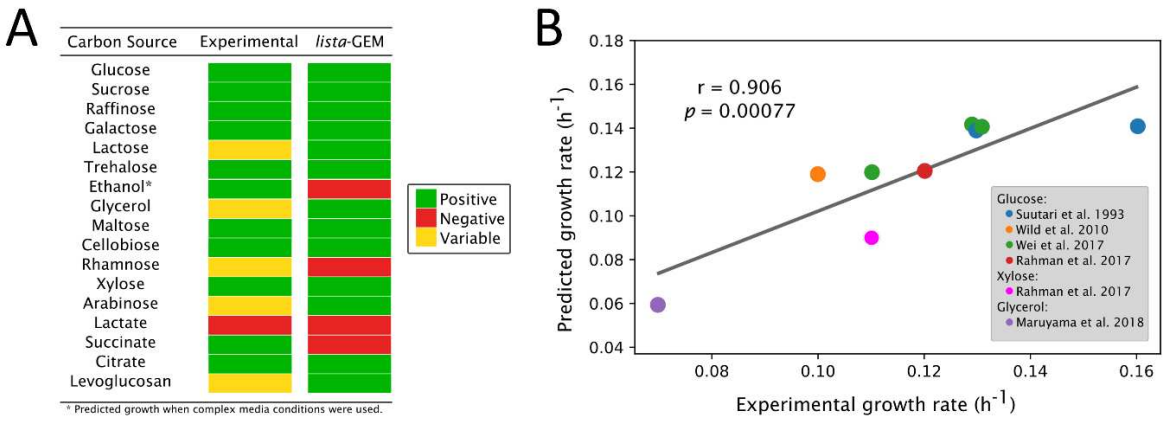
2.3.3. Gap-filling, manual curation, and quality assessment

The gap-filling of *lista*-GEM was conducted in two steps. In the first step, we used Meneco (PRIGENT et al., 2017) to identify the reactions required for the biosynthesis of biomass components (target compounds) based on a list of available metabolites (seeds). The reactions identified by Meneco were retrieved from *rhto*-GEM. However, after Meneco was applied, we noticed that the model could still not sustain growth (i.e. produce biomass). Thus, in the second step, we used the "*fillGaps*" function from the RAVEN Toolbox. We considered growth on glucose (1 mmol/gDW h) at a biomass production rate of 0.01 h^{-1}. The reactions required to sustain biomass formation were then retrieved from *rhto*-GEM and iYali templates. Finally, we

noted that three reactions included from iYali ('y300065', 'y300066', 'y200008') were not required for growth and led to water and H^+ overproduction in rich media simulations and removed them.

After the gap-filling step, we included the specific reactions required by *L. starkeyi* to sustain growth on the specified carbon sources and to meet cofactor requirements. In contrast to other oleaginous yeasts, the malic enzyme of *L. starkeyi* preferably uses NAD^+ instead of $NADP^+$ as a cofactor (TANG et al., 2010). Thus, we removed the malic enzyme reaction that used $NADP^+$ and maintained only the one that uses NAD^+. Additionally, we manually included the reactions necessary for L-rhamnose, lactose, cellobiose, and levoglucosan utilization. Finally, we updated the gene-reaction rules (grRules field), replacing the genes in the model that still contained the identification from the template (*R. toruloides*) with *L. starkeyi* homologs. The non-growth associated maintenance reaction remained the same as in *rhto*-GEM due to the lack of available data for *L. starkeyi*. We assessed the quality of the final reconstruction using MEMOTE (LIEVEN et al., 2020) (Figure 1A).

Figure 1 - **(A)** Viability predicted by *lista*-GEM for different carbon sources compared to experimental data (Smith and Kurtzman, 2011). **(B)** Correlation between experimental and predicted growth rates on glucose, xylose and glycerol (Maruyama et al., 2018; Rahman et al., 2017; Suutari et al., 1993; Wild et al., 2010).



2.3.4. Simulations and validation

To quantitatively assess the growth of *L. starkeyi* on different carbon sources (glucose, acetate, arabinose, cellobiose, citrate, ethanol, galactose, lactose, levoglucosan, xylose, rhamnose, R-lactate, S-lactate, mannose, trehalose; see Figure 1B) in minimal medium, we first constrained the lower bound of exchange reactions to zero and left only the oxygen, ammonium, H+, iron, phosphate, potassium, and sulfate exchange reactions unconstrained. Then, we allowed the uptake of each carbon source at -3 mmol/[g dry weight (DW) h] and optimized the formation of biomass via Flux Balance Analysis (FBA). To simulate growth on rich media, we also set the uptake of amino acids to -1 mmol/(g DW h).

To quantitatively assess model performance, we compared the experimental growth rate gathered from the literature with our predictions. Data were available for glucose, xylose and glycerol. When available, we set the carbon uptake rate as described in the manuscript. If not, we assumed a value of -3 mmol/(g DW h). The media (minimal or rich) was also adjusted based on the source manuscript description and the biomass formation was optimized using FBA. The correlation between *in vivo* and *in silico* growth data was determined by the Pearson's correlation coefficient (Figure 1B).

Furthermore, we conducted phase plane analysis in two different scenarios to determine conditions that would favor growth and lipid production in three carbon sources found in agro-industrial wastes (glucose and xylose from lignocellulosic biomasses and glycerol from biodiesel production). In the first scenario, we varied the carbon source uptake [from 0 to -10 mmol/(g DW h)] and oxygen uptake [from 0 to -50 mmol/(g DW h)] rates, while maintaining the other components of the minimal media as described above unconstrained. In the second scenario, instead of constraining carbon and oxygen uptake, we constrained nitrogen [from 0 to -9 mmol/(g DW h)] and oxygen [from 0 to -27 mmol/(g DW h)] uptake rates, with a fixed carbon uptake rate of -3 mmol/(g DW h). We used FBA to optimize growth considering the biomass formation equation as described above and lipid production considering a pseudoreaction for TAG (1-16:0, 2-18:1, 3-18:1) exchange.

Moreover, we evaluated the main reactions related to lipid accumulation in nitrogen-limiting conditions using the environmental version of minimization of metabolic adjustment (eMOMA) (KIM et al., 2019). First, the same pseudoreaction described above to represent TAG exchange was added to the model, and the lower bound of the non-growth associated maintenance reaction (NGAM) was set to a low value [0.5 mmol/(g DW h)] to represent

stationary growth. Then, we blocked the exchange reactions for ethanol, trehalose, butanediol, pyruvate, fumarate, 2-oxoglutarate, malate, oxaloacetate, glyoxylate, and acetate since we did not find evidence regarding the excretion of these metabolites for *L. starkeyi* under nitrogen-limiting conditions. We also blocked the exchange of decanoate, palmitate, palmitoleate, oleate, 14-demethyllanosterol, episterol, ergosterol, fecosterol, lanosterol, zymosterol, and ergosta-5,7,22,24(28)-tetraen-3beta-ol to promote TAG accumulation. Then, we set the growth as objective and performed FBA to obtain the flux distribution under non-restricted conditions (minimal media). Next, we blocked nitrogen exchange to simulate nitrogen restriction and confirmed that the model could not predict the growth and conducted the traditional MOMA between the model with and without nitrogen restriction. To test the reactions that affect lipid accumulation via knockout or overexpression, we removed reactions with zero flux in both conditions. Thereafter, we performed the eMOMA by knocking out or overexpressing (2x higher flux) the remaining reactions. We kept reactions where at least 2% increase in TAG exchange compared to the nitrogen-restricted reference and at least 90% growth remained compared to the nitrogen-abundant reference condition. We conducted eMOMA simulations for glucose, xylose, and glycerol at a fixed carbon uptake of -3 mmol/(g DW h).

Finally, we predicted overexpression targets to improve lipid production using glucose, xylose, and glycerol as carbon sources via flux scanning based on enforced objective flux (FSEOF) analysis (CHOI et al., 2010). We performed the simulations considering minimal media, set the NGAM to 0 mmol/(g DW h), the TAG exchange pseudoreaction as the target, and the carbon uptake to -3 mmol/(g DW h). We conducted all simulations using the RAVEN Toolbox (v. 2.7.9) and/or the COBRA Toolbox (v. 3.4) (HEIRENDT et al., 2019) in MATLAB (The MathWorks Inc., Natick, Massachusetts) using Gurobi® (v. 10.0) as the solver.

2.4. Results and Discussion

2.4.1. Properties of the lista-GEM reconstruction

Herein, we reconstructed the first genome-scale metabolic model of the oleaginous yeast *Lipomyces starkeyi*. We applied a stepwise reconstruction strategy using the RAVEN toolbox based on the pipelines described by TIUKOVA et al. (2019) and VENTORIM et al. (2022). Most genes (907 of 935) in the model were recovered in the first step (Homology draft; Table 1) of the reconstruction via bidirectional BLAST with the *R. toruloides* and *Y. lipolytica*, and their respective GEMs (*rhto*-GEM 1.3.0 and iYali 4.1.2). The next steps of the

reconstruction focused mainly on adding pseudo and lipid metabolism (SLIMEr) reactions, gap-filling and manual curation of the model (see Material and Methods). The final version of the model *lista*-GEM 1.0.0 presented a final MEMOTE score of 52%. This low stoichiometry consistency is related to the fact that the model was penalized for stoichiometric consistency and annotation, a common phenomenon for models that included the SLIMEr formalism, as lipid species are normalized by their weight for direct integration of lipid measurements (SÁNCHEZ et al., 2019). Consistently, the GEMs *rhto*-GEM (TIUKOVA et al., 2019) and *papla*-GEM (VENTORIM et al., 2022), which included the SLIMEr formalism, have a score similar to *lista*-GEM.

Table 1 - Genes, metabolites, and reactions of *lista*-GEM during the reconstruction steps.

Reconstruction step	Genes	Metabolites	Reactions
Homology draft	907	1820	1537
Add pseudoreactions	907	2420	2132
Lipid curation	960	2811	2326
Biomass definition	960	2811	2326
Gap-filling	925	2831	2328
Manual curation	935	2844	2337

2.4.2. lista-GEM accurately represents the metabolism of L. starkeyi

The model presented a good performance for qualitative and quantitative growth prediction compared to experimental data. For 14 of 17 carbon sources, the model correctly predicted the growth/non-growth profile (Figure 1A) (SMITH; KURTZMAN, 2011). For ethanol, growth was predicted only in complex media simulations with uptake of essential amino acids. However, the model could not predict growth in none of the conditions tested for succinate. Furthermore, the predicted growth rates in three carbon sources of biotechnological interest (glucose, xylose, and glycerol) presented a good Pearson correlation with *in vivo* measurements ($r = 0.906$, $p = 0.00077$) (MARUYAMA et al., 2018; RAHMAN et al., 2017; SUUTARI et al., 1993; WILD et al., 2010) (Figure 1B).

The phase plane analysis for growth and TAG production in glucose, xylose and glycerol as carbon sources indicated a higher dependence on carbon uptake than oxygen uptake (Figures 2-3, S1-4). Although, a minimum oxygen uptake [15-20 mmol/(g DW h) for glucose and xylose and 5-10 mmol/(g DW h) for glycerol] was required by *L. starkeyi* to reach

maximum biomass production, further increases in oxygen availability, in contrast to carbon availability, did not increased biomass production.

Figure 2 - Effects of varying the glucose and oxygen uptake rates on growth rate.

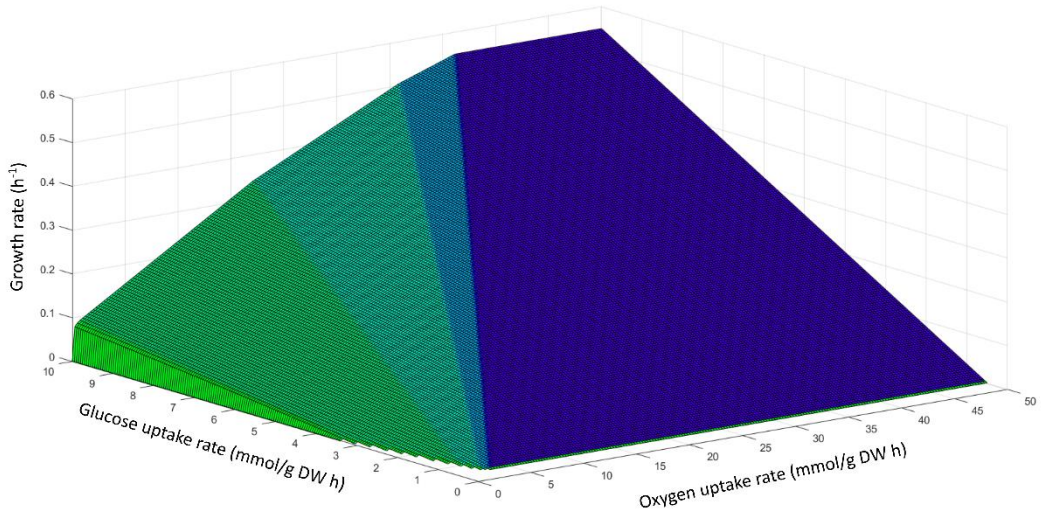
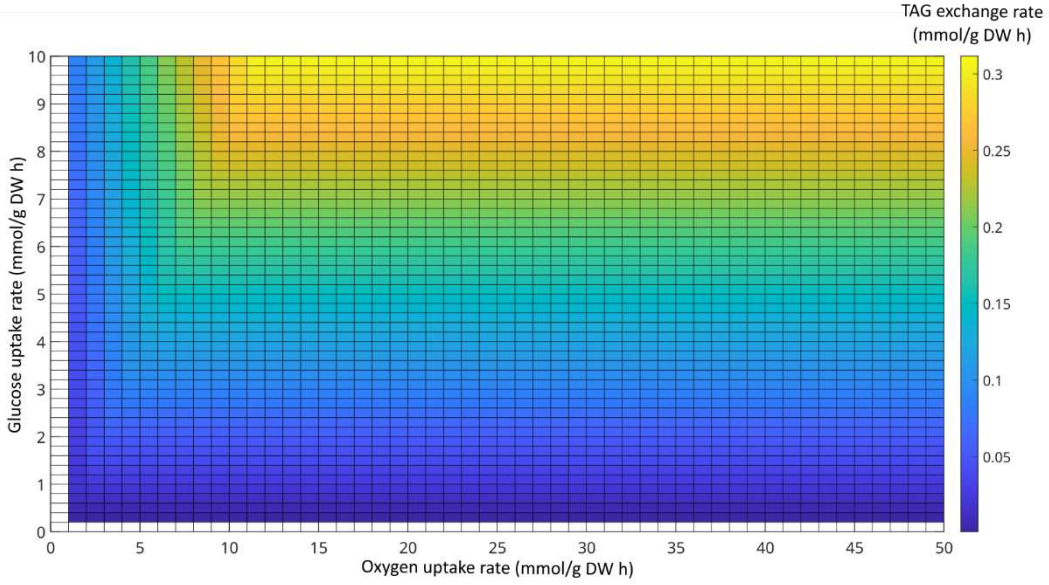


Figure 3 - Effects of varying the glucose and oxygen uptake rates on TAG exchange.



Notably, the oxygen requirement for maximum TAG production was lower than those for growth on glucose and xylose (Figures 2-3 and S1-4). This is likely associated with the fact that lipid accumulation in oleaginous yeasts starts from the late-exponential phase (Ratledge, 2008), where the oxygen availability is lower than the exponential phase. Importantly, from a bioprocess development point of view, the low oxygen requirement for TAG accumulation in *L. starkeyi* is advantageous, as the dissolved oxygen availability, which in turn affects the oxygen transfer rate, is a limiting factor for aerobic processes. Since the fatty acid synthesis has a high demand for reducing power, even compared to other biomass components, this reduced requirement of oxygen during the lipid accumulation phase might not be restricted to *L. starkeyi*, but might also represent the behavior of other oleaginous microorganisms.

Meanwhile, the oxygen requirement for maximum growth and TAG production was the same on glycerol (Figures S3-4), highlighting important differences between glycolytic and gluconeogenic carbon sources regarding the lipid production by *L. starkeyi*. For the phase plane analysis performed varying the nitrogen source uptake rate and oxygen uptake rate, we noticed that the key determinant for achieving growth and producing TAG was the availability of oxygen (Figures S5-7).

2.4.3. Predicted targets for enhancing the production of TAGs by metabolic engineering strategies

To improve the production of lipids using metabolic engineering, we identified gene targets for knockout or overexpression using a combination of minimal adjustment of fluxes (eMOMA) and flux scanning (FSEOF). We simulated growth conditions where nitrogen was limited and the NGAM was set to a low value, which represents stationary growth, and used a TAG representative to simulate lipid production (see Material and Methods).

The eMOMA approach is useful to predict the flux distribution for a changed environment. Similar to MOMA, eMOMA is implemented as a linear or quadratic problem to minimize the L1 or L2-norm distances, respectively, between the reference and alternative flux distributions. However, while the MOMA approach is tailored to minimize the difference between a wild-type and a mutant strain based on the principle of minimal metabolic adjustment, the eMOMA approach expands the MOMA implementation by considering an additional constraint, where the flux through the uptake reaction of a growth-limiting nutrient is equal to zero. To predict the important reactions for lipid production, we constrained the

uptake of exchange reactions relative to metabolites from the TCA cycle, sterols, and various lipids such as decanoate, palmitate, palmitoleate and oleate. This approach allowed the identification of the main reactions associated with lipid accumulation. It is important to point out that many of the identified reactions were the same in glucose, xylose, or glycerol as the carbon sources, being primarily involved in the exchange reaction of the carbon source, the pentose phosphate pathway, tricarboxylic acid (TCA) cycle, and lipid metabolism (Tables 2-4).

Table 2 - Targets identified by eMOMA on *lista*-GEM using glucose as carbon source.

		Effects of knockout		Effects of overexpression		
Reaction ID	**Reaction name**	**Growth**	**TAG exchange**	**Growth**	**TAG exchange**	**Best result**
r_1714	D-glucose exchange	0.0000	0.0000	2.0119	2.0219	2.0219
r_0958	pyruvate carboxylase	0.9323	0.9932	0.9983	1.1031	1.1031
r_0016	2-aceto-2-hydroxybutanoate synthase	-0.0000	0.0000	0.9931	1.0718	1.0718
r_0353	dihydroxy-acid dehydratase (2,3-dihydroxy-3-methylpentanoate)	-0.0000	0.0000	0.9931	1.0718	1.0718
r_0669	ketol-acid reductoisomerase (2-aceto-2-hydroxybutanoate)	-0.0000	0.0000	0.9931	1.0718	1.0718

Note: The values represent the factor related to the value obtained for the condition without knockout or overxpression.

Table 3 - Targets identified by eMOMA on *lista*-GEM using xylose as carbon source.

		Effects of knockout		Effects of overexpression		
Reaction ID	**Reaction name**	**Growth**	**TAG exchange**	**Growth**	**TAG exchange**	**Best result**
r_1718	D-xylose exchange	0.0000	0.0000	2.0145	2.0308	2.0308
r_0016	2-aceto-2-hydroxybutanoate synthase	-0.0000	0.0055	0.9931	1.0804	1.0804
r_0353	dihydroxy-acid dehydratase (2,3-dihydroxy-3-methylpentanoate)	-0.0000	0.0055	0.9931	1.0804	1.0804
r_0669	ketol-acid reductoisomerase (2-aceto-2-hydroxybutanoate)	-0.0000	0.0055	0.9931	1.0804	1.0804
r_0958	pyruvate carboxylase	0.9323	0.9505	0.9983	1.0795	1.0795

Note: The values represent the factor related to the value obtained for the condition without knockout or overxpression.

Table 4 - Targets identified by eMOMA on *lista*-GEM using glycerol as carbon source.

		Effects of knockout		Effects of overexpression		
Reaction ID	**Reaction name**	**Growth**	**TAG exchange**	**Growth**	**TAG exchange**	**Best result**
r_1808	glycerol exchange	0.0000	0.0000	2.0249	2.0245	2.0245
r_0958	pyruvate carboxylase	0.9384	1.0884	0.9980	1.1180	1.1180
r_0353	dihydroxy-acid dehydratase (2,3-dihydroxy-3-methylpentanoate)	-0.0000	0.0000	0.9930	1.1058	1.1058
r_0669	ketol-acid reductoisomerase (2-aceto-2-hydroxybutanoate)	-0.0000	0.0000	0.9930	1.1058	1.1058
r_0016	2-aceto-2-hydroxybutanoate synthase	-0.0000	0.0000	0.9930	1.1058	1.1058
r_0219	aspartate-semialdehyde dehydrogenase	-0.0000	0.0000	0.9919	1.0668	1.0668
r_0215	aspartate kinase	-0.0000	0.0000	0.9919	1.0668	1.0668
r_1832	H+ exchange	0.9861	1.0198	1.0000	1.0604	1.0604
r_1794	formate transport	0.9999	1.0476	1.0000	1.0471	1.0476
r_2045	serine transport	0.9999	1.0476	1.0000	1.0472	1.0476
r_0503	glycine hydroxymethyltransferase	0.9999	1.0476	1.0000	1.0472	1.0476
r_0724	methenyltetrahydrifikate cyclohydrolase	0.9999	1.0476	1.0000	1.0471	1.0476
r_0733	methylenetetrahydrofolate dehydrogenase (NADP)	0.9999	1.0476	1.0000	1.0430	1.0476
r_0447	formate-tetrahydrofolate ligase	0.9999	1.0476	1.0000	1.0471	1.0476
r_1811	glycine transport	0.9999	1.0473	1.0000	1.0463	1.0473
r_2141	fatty-acyl-CoA synthase (n-C18:0CoA)	1.0000	0.9860	0.9913	1.0444	1.0444
r_0725	methenyltetrahydrofolate cyclohydrolase	0.9999	1.0438	0.9977	0.0000	1.0438
r_0732	methylenetetrahydrofolate dehydrogenase (NADP)	1.0000	1.0421	0.2523	0.2539	1.0421

Note: The values represent the factor related to the value obtained for the condition without knockout or overxpression.

The FSEOF approach relies on enforcing the flux on product formation by searching for candidate reactions whose increase in flux also increases the flux on product formation. This ensures that all identified reactions contribute to enhancing the formation of the desired product. The maximum theoretical value for product formation predicted by conventional FBA is biologically unrealistic since the formation of biomass becomes negligible. To circumvent this, FSEOF sets as the objective function the biomass formation and identifies the intracellular fluxes that increase when the maximum theoretical value for product formation is applied as a constraint. This makes it possible to achieve a product formation flux close to the maximum theoretical value while respecting biological feasibility. The FSEOF analysis identified similar reactions to eMOMA, such as those involved in pyruvate metabolism, the pentose phosphate pathway, and TCA cycle. Additionally, we also identified reactions such as acetyl-CoA carboxylase, acetyl-CoA synthetase, and fatty-acyl-CoA synthase, which are key reactions involved in the biosynthesis of fatty acids (Table 5).

Table 5 - Targets identified by FSEOF to improve lipid production using *lista*-GEM.

TAG production				
Glucose	**Xylose**	**Glycerol**	**Reaction name**	**Genes**
18.21	18.17	19.26	pyruvate kinase	ODQ76269.1
17.99	17.96	18.67	acetyl-CoA carboxylase	ODQ75673.1 and ODQ72018.1
17.76	17.73	18.88	enolase	ODQ74427.1
17.76	17.73	18.88	phosphoglycerate mutase	ODQ73127.1 or ODQ76545.1
17.28	17.24	18.46	glyceraldehyde-3-phosphate dehydrogenase	ODQ75822.1
17.28	17.24	18.46	phosphoglycerate kinase	ODQ69690.1
16.75	16.72	17.60	bicarbonate formation	ODQ71959.1
-	-	21.09	pyruvate decarboxylase	ODQ73413.1
-	-	21.09	acetaldehyde dehydrogenase	ODQ69191.1 or ODQ69510.1 or ODQ70419.1 or ODQ70788.1 or ODQ71777.1 or ODQ72007.1 or ODQ73381.1 or ODQ75322.1
-	-	20.98	acetyl-CoA synthetase	ODQ74542.1
-	-	18.90	adenylate kinase	ODQ72719.1 or ODQ74535.1
-	-	17.24	inorganic diphosphatase	ODQ71884.1
-	-	16.45	triose-phosphate isomerase	ODQ69158.1 or ODQ69323.1

1.67	1.67	1.72	stearoyl-CoA desaturase (n-C18:0CoA -> n-C18:1CoA), ER membrane	ODQ73471.1
1.65	1.65	1.69	fatty-acyl-CoA synthase (n-C18:0CoA)	ODQ70131.1 or (ODQ75304.1 and ODQ70130.1) or ODQ70131.1 or ODQ75303.1 or ODQ75304.1
0.89	0.89	0.92	diacylglycerol acyltransferase (1-16:0, 2-18:1, 3-18:1), endoplasmic reticulum membrane	ODQ70106.1 or ODQ75508.1
0.82	0.82	0.85	PA phosphatase (1-16:0, 2-18:1), endoplasmic reticulum membrane	ODQ71343.1
0.82	0.82	0.85	1-acyl-sn-glycerol-3-phosphate acyltransferase (1-16:0, 2-18:1), endoplasmic reticulum membrane	ODQ69760.1 or ODQ75695.1
0.79	0.79	0.82	glycerol-3-phosphate acyltransferase (16:0), endoplasmic reticulum membrane	ODQ74991.1
0.69	0.69	0.73	fatty-acyl-CoA synthase (n-C16:0CoA)	ODQ70131.1 or (ODQ75304.1 and ODQ70130.1) or ODQ70131.1 or ODQ75303.1 or ODQ75304.1
-	-	0.79	glycerol kinase	ODQ76330.1
-	0.76	-	glycerol-3-phosphate dehydrogenase (NAD)	ODQ74980.1

Note: Slopes derived from FSEOF, indicated of whether gene expression should be increased to direct flux from growth toward production. The higher the slope, the higher the theoretical increase in TAG production.

Similar to other metabolic engineering strategies reported for other oleaginous yeasts, we could identify common targets for overexpression and knockout. The eMOMA and FSEOF analyses performed on the *rhto*-GEM and *i*MK735 (KIM et al., 2019) models also predicted many of the same targets shown on Tables 2-5. Additionally, many of the target genes predicted using *lista*-GEM are experimentally validated for *L. starkeyi* using the three tested sugars, such as the acetyl-CoA carboxylase, fatty acid synthetase and glycerol-3-phosphate acyltransferase (ZHANG et al., 2022). The overexpression of these genes has been performed by Jeffries et al. 2017 (JEFFRIES et al., 2017) to improve lipid production using stillage and its derivatives. They reported an increase of 85% in the lipid titer. In another study, the overexpression of the gene coding for acetyl-CoA carboxylase led to an increase in the production of malonyl-CoA (LU et al., 2008). Further, by doubling the copy number of the genes encoding subunits of the enzyme fatty-acid synthetase, CHEN et al. (2020) reported an increase of 60% in the lipid content. Taken together, our results highlight the accuracy of the *lista*-GEM reconstruction to

identify important reactions for lipid production and to develop suitable metabolic engineering strategies to enhance lipid production.

2.5. Conclusions

Herein, we present the first genome-scale metabolic reconstruction and curation of *L. starkeyi*, termed *lista*-GEM. The model was based on two high-quality reconstructions and further curated using experimental data to represent the metabolic specificities of *L. starkeyi*. The growth conditions simulated using FBA were in good agreement with experimental growth data, underscoring the usefulness of *lista*-GEM in predicting phenotypes. Further, its usefulness for metabolic engineering was demonstrated by the prediction of gene targets in line with experimental results. Although a genome-scale reconstruction only describes current knowledge and is never finished (ANTON et al., 2023), the open nature of GEMs allows for its continuous development by all members of the scientific community. While the *lista*-GEM already proves itself useful for the study of lipid metabolism and for biotechnological applications, enhancements such as enzyme constraints can further push it to new horizons.

2.6. Data availability

The scripts used for the reconstruction and simulations as well as the *lista*-GEM model are available in the GitHub repository at https://github.com/LabFisUFV/lista-GEM or through Zenodo at https://doi.org/10.5281/zenodo.8367982. The model is provided according to the standard-GEM template (ANTON et al., 2023) and in different formats (TXT, SMBL, XLSX, and MAT).

2.8. Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) [Finance Code 001] and Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq) [Finance Code 140538/2021-6].

2.9. Conflicts of interest

The authors declare no competing interests.

2.10. References

ABELN, F., CHUCK, C.J. The history, state of the art and future prospects for oleaginous yeast research. **Microb. Cell Factories** 20, 221. 2021. https://doi.org/10.1186/s12934-021-01712-1

ANGERBAUER, C., et al. Conversion of sewage sludge into lipids by *Lipomyces starkeyi* for biodiesel production. 2008 **Bioresour. Technol.** 99, 3051–3056. https://doi.org/10.1016/j.biortech.2007.06.045

ANSCHAU, A. et al. Effect of feeding strategies on lipid production by *Lipomyces starkeyi*. **Bioresour. Technol.** 157, 214–222. 2014. https://doi.org/10.1016/j.biortech.2014.01.104

ANTON, M., et al. standard-GEM: standardization of open-source genome-scale metabolic models (preprint). **Systems Biology.** 2023. https://doi.org/10.1101/2023.03.21.512712

CALVEY, C.H. et al. Nitrogen limitation, oxygen limitation, and lipid accumulation in *Lipomyces starkeyi*. **Bioresour. Technol.** 200, 780–788. 2016. https://doi.org/10.1016/j.biortech.2015.10.104

CHEN, L., et al. Cellular lipid production by the fatty acid synthase-duplicated *Lipomyces kononenkoae* BF1S57 strain for biodiesel making. **Renew. Energy** 151, 707–714. 2020. https://doi.org/10.1016/j.renene.2019.11.074

CHOI, H.S., et al. *In Silico* Identification of Gene Amplification Targets for Improvement of Lycopene Production. **Appl. Environ. Microbiol.** 76, 3097–3105. 2010. https://doi.org/10.1128/AEM.00115-10

DINH, H.V. et al. A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. **Metab. Eng. Commun.** 9, e00101. 2019. https://doi.org/10.1016/j.mec.2019.e00101

HEIRENDT, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. **Nat. Protoc.** 14, 639–702. 2019. https://doi.org/10.1038/s41596-018-0098-2

JEFFRIES, T., MOKRY, D., CALVEY, C.H., Composition and methods for producing lipids and other biomaterials from grain ethanol stillage and stillage derivatives. 2017. US 10,662,448 B2.

KAVŠČEK, M., et al. Optimization of lipid production with a genome-scale model of *Yarrowia lipolytica*. **BMC Syst. Biol.** 9, 72. 2015. https://doi.org/10.1186/s12918-015-0217-4.

KERKHOVEN, E.J., et al. 2016. Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. **Npj Syst. Biol. Appl.** 2, 16005. https://doi.org/10.1038/npjsba.2016.5

KIM, J., et al. Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in *Rhodosporidium toruloides*. **Front. Bioeng. Biotechnol.** 8, 612832. 2021. https://doi.org/10.3389/fbioe.2020.612832

KIM, M., et al. In silico identification of metabolic engineering strategies for improved lipid production in *Yarrowia lipolytica* by genome-scale metabolic modeling. **Biotechnol. Biofuels** 12, 187. 2019. https://doi.org/10.1186/s13068-019-1518-4

LIEVEN, C., et al. MEMOTE for standardized genome-scale metabolic model testing. **Nat. Biotechnol.** 38, 272–276. 2020. https://doi.org/10.1038/s41587-020-0446-y

LIU, L., et al. Efficient microbial oil production on crude glycerol by *Lipomyces starkeyi* AS 2.1560 and its kinetics. **Process Biochem.** 58, 230–238. 2017. https://doi.org/10.1016/j.procbio.2017.03.024

LOIRA, N., et al. A genome-scale metabolic model of the lipid-accumulating yeast *Yarrowia lipolytica*. **BMC Syst. Biol.** 6, 35. 2012. https://doi.org/10.1186/1752-0509-6-35

LU, X., VORA, H., KHOSLA, C. Overproduction of free fatty acids in *E. coli:* Implications for biodiesel production. **Metab. Eng.** 10, 333–339. 2008. https://doi.org/10.1016/j.ymben.2008.08.006

MADDEN, T., 2013. The BLAST Sequence Analysis Tool. **NCBI Handbook**.

MARUYAMA, Y et al. 2018. Characterization of oil-producing yeast *Lipomyces starkeyi* on glycerol carbon source based on metabolomics and 13C-labeling. **Appl. Microbiol. Biotechnol.** 102, 8909–8920. https://doi.org/10.1007/s00253-018-9261-5

MATSUZAWA, T., et al. Identification and characterization of Δ12 and Δ12/Δ15 bifunctional fatty acid desaturases in the oleaginous yeast *Lipomyces starkeyi*. **Appl. Microbiol. Biotechnol.** 102, 8817–8826. 2018. https://doi.org/10.1007/s00253-018-9345-2

MISHRA, P., et al. Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica*. **BMC Syst. Biol.** 12, 12. 2018. https://doi.org/10.1186/s12918-018-0542-5

NING, J., et al. Purification and characterization of levoglucosan kinase from *Lipomyces starkeyi* YZ-215. **World J. Microbiol. Biotechnol.** 24, 15–22. 2008. https://doi.org/10.1007/s11274-007-9432-5

ORTH, J.D., THIELE, I., PALSSON, B.Ø. What is flux balance analysis? **Nat. Biotechnol.** 28, 245–248. 2010. https://doi.org/10.1038/nbt.1614

PAN, P., HUA, Q., Reconstruction and In Silico Analysis of Metabolic Network for an Oleaginous Yeast, *Yarrowia lipolytica*. **PLoS ONE** 7, e51535. 2012. https://doi.org/10.1371/journal.pone.0051535

PHAM, N. et al. Genome-scale metabolic modeling underscores the potential of *Cutaneotrichosporon oleaginosus* ATCC 20509 as a cell factory for biofuel production. **Biotechnol. Biofuels** 14, 2. 2021. https://doi.org/10.1186/s13068-020-01838-1

POMRANING, K.R., et al. Transcriptomic analysis of the oleaginous yeast *Lipomyces starkeyi* during lipid accumulation on enzymatically treated corn stover hydrolysate. **Biotechnol. Biofuels** 12, 162. 2019. https://doi.org/10.1186/s13068-019-1510-z

PRIGENT, S., et al. Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks. **PLOS Comput. Biol.** 13, e1005276. 2017. https://doi.org/10.1371/journal.pcbi.1005276

PROBST, K.V., VADLANI, P.V. Production of single cell oil from *Lipomyces starkeyi* ATCC 56304 using biorefinery by-products. **Bioresour. Technol.** 198, 268–275. 2015. https://doi.org/10.1016/j.biortech.2015.09.018

PUTRA, F.J.N.,et al. The bioconversion of lignin derivative aldehydes into high-value aromatic alcohols and lipids via *Lipomyces starkeyi.* **Biochem. Eng. J.** 109065. 2023. https://doi.org/10.1016/j.bej.2023.109065

RAHMAN, S. et al. Microbial lipid production from lignocellulosic hydrolyzates: effect of carbohydrate mixtures and acid-hydrolysis byproducts on cell growth and lipid production by *Lipomyces starkeyi*: Microbial lipid production from lignocellulosic hydrolyzates. **J. Chem. Technol. Biotechnol.** 92, 1980–1989. 2017. https://doi.org/10.1002/jctb.5185

RATLEDGE, C., Microbial Lipids, in: Rehm, H.-J., Reed, G. (Eds.), **Biotechnology.** Wiley-VCH Verlag GmbH, Weinheim, Germany, pp. 133–197. 2008. https://doi.org/10.1002/9783527620890.ch4

SALVADOR LÓPEZ, J.M., VANDEPUTTE, M., VAN BOGAERT, I.N.A., Oleaginous yeasts: Time to rethink the definition? **Yeast** 39, 553–606. 2022. https://doi.org/10.1002/yea.3827

SÁNCHEZ, B.J. et al. SLIMEr: probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. **BMC Syst. Biol.** 13, 4. 2019. https://doi.org/10.1186/s12918-018-0673-8

SAHA, R.; MUKHOPADHYAY, M. Prospect of metabolic engineering in enhanced microbial lipid production: review**. Biomass Conversion and Biorefinery**, 3 dez. 2021.

SMITH, M.TH., KURTZMAN, C.P. *Lipomyces* Lodder & Kreger-van Rij (1952), in: The **Yeasts.** Elsevier, pp. 545–560. 2011. https://doi.org/10.1016/B978-0-444-52149-1.00043-4

SPAGNUOLO, M., YAGUCHI, A., BLENNER, M. Oleaginous yeast for biofuel and oleochemical production. **Curr. Opin. Biotechnol.** 57, 73–81. 2019. https://doi.org/10.1016/j.copbio.2019.02.011

SUUTARI, M., PRIHA, P., LAAKSO, S. Temperature shifts in regulation of lipids accumulated by *Lipomyces starkeyi*. **J. Am. Oil Chem. Soc.** 70, 891–894. 1993. https://doi.org/10.1007/BF02545349

TAKAKU, H., et al. Lipid metabolism of the oleaginous yeast *Lipomyces starkeyi*. Appl. Microbiol. Biotechnol. 104, 6141–6148. 2020. https://doi.org/10.1007/s00253-020-10695-9

TANG, W. et al. Molecular Cloning and Characterization of a Malic Enzyme Gene from the Oleaginous Yeast *Lipomyces starkeyi*. **Mol. Biotechnol.** 45, 121–128. 2010. https://doi.org/10.1007/s12033-010-9255-8

TIUKOVA, I.A. Genome-scale model of *Rhodotorula toruloides* metabolism. **Biotechnol. Bioeng.** 116, 3396–3408. 2019. https://doi.org/10.1002/bit.27162

UZUKA, Y. et al. Effect of culture ph on the growth and biotin requirement in a strain of *Lipomyces starkeyi*. **J. Gen. Appl. Microbiol.** 20, 197–206. 1974. https://doi.org/10.2323/jgam.20.197

VENTORIM, R.Z. et al. Genome-scale metabolic model of oleaginous yeast *Papiliotrema laurentii*. **Biochem. Eng. J.** 180, 108353. 2022. https://doi.org/10.1016/j.bej.2022.108353

WANG, H., et al. 2018. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. **PLOS Comput. Biol.** 14, e1006541. https://doi.org/10.1371/journal.pcbi.1006541

WEI, S. et al. Reconstruction of genome-scale metabolic model of *Yarrowia lipolytica* and its application in overproduction of triacylglycerol. **Bioresour. Bioprocess.** 4, 51. 2017. https://doi.org/10.1186/s40643-017-0180-6

WILD, R. et al. Lipids from *Lipomyces starkeyi*. **Food Technol Biotechnol** 48, 329–335. 2010.

YE, C. et al. Genome-scale metabolic network models: from first-generation to next-generation. **Appl. Microbiol. Biotechnol.** 106, 4907–4920. 2022. https://doi.org/10.1007/s00253-022-12066-y

YU, X. et al. Oil production by oleaginous yeasts using the hydrolysate from pretreatment of wheat straw with dilute sulfuric acid. **Bioresour. Technol.** 102, 6134–6140. 2011. https://doi.org/10.1016/j.biortech.2011.02.081

XU, P. et al. Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals. **Proceedings of the National Academy of Sciences**, v. 113, n. 39, p. 10848–10853, 27 set. 2016.

ZHANG, L. et al. Enhancing microbial lipids yield for biodiesel production by oleaginous yeast *Lipomyces starkeyi* fermentation: A review. **Bioresour. Technol.** 344, 126294. 2022. https://doi.org/10.1016/j.biortech.2021.126294

2.11. Supplementary material

Figure S1 - Effects of varying the xylose and oxygen uptake rates on growth.

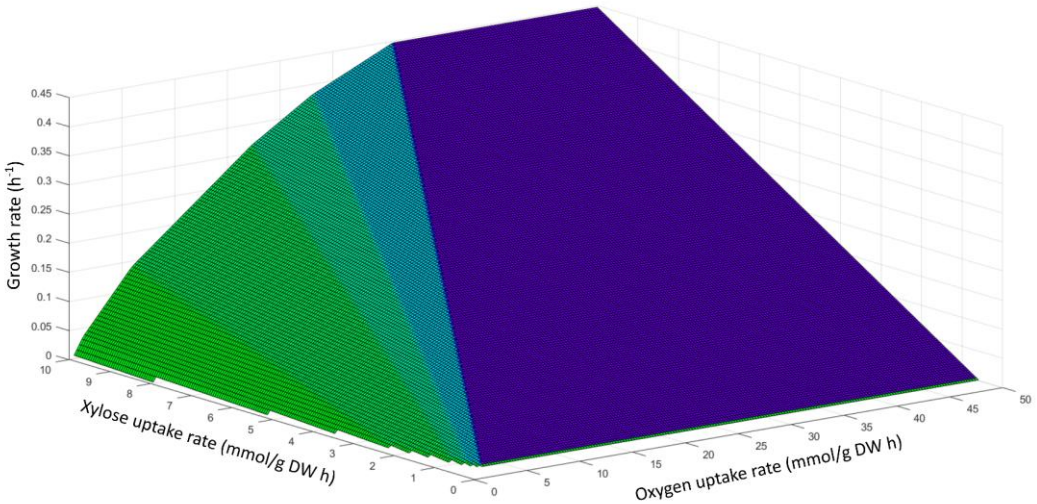


Figure S2 - Effects of varying the xylose and oxygen uptake rates on TAG exchange.

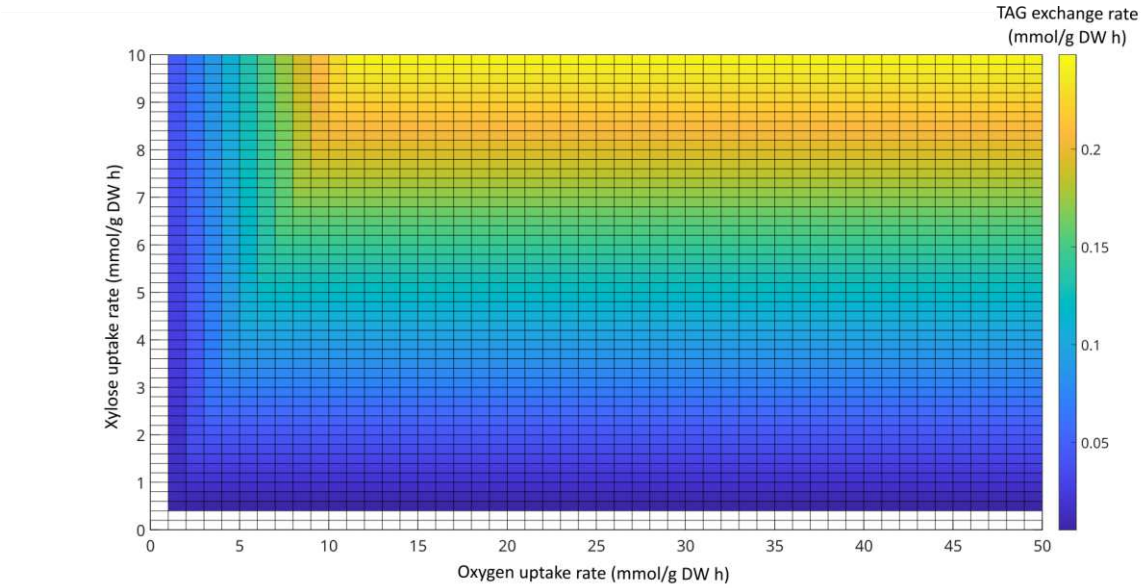


Figure S3 - Effects of varying the glycerol and oxygen uptake rates on growth.

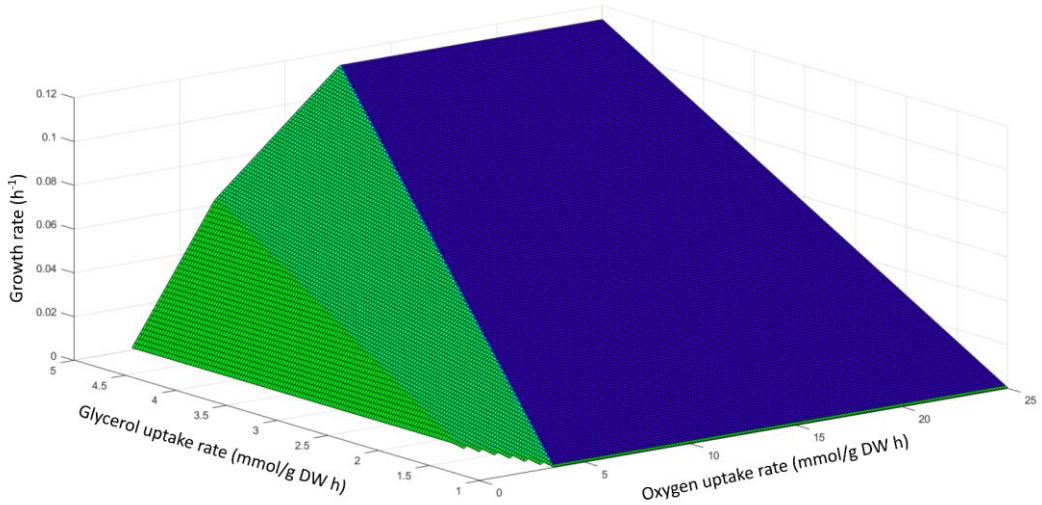


Figure S4 - Effects of varying the glycerol and oxygen uptake rates on TAG exchange.

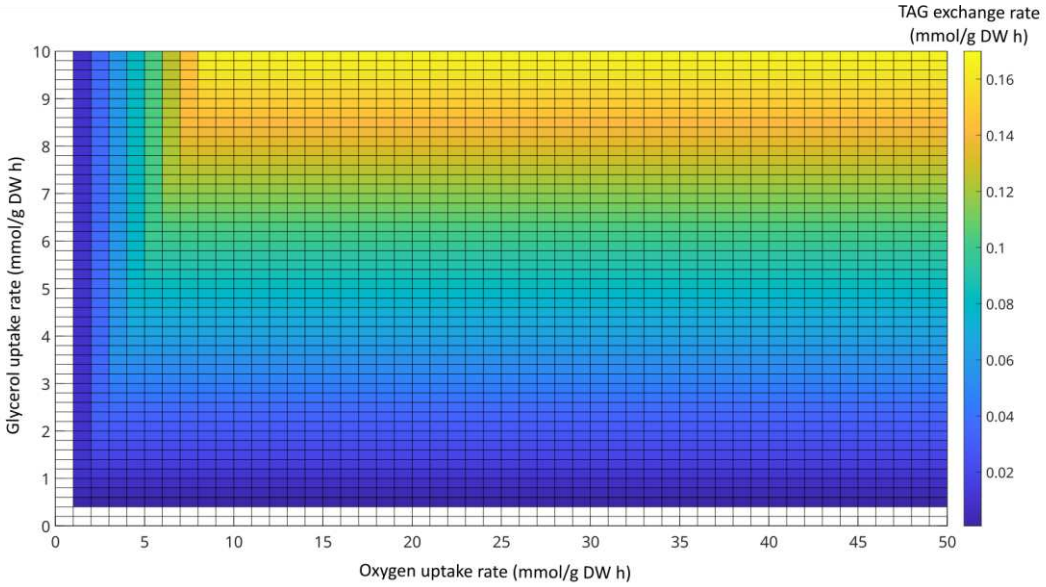


Figure S5 - Effects of varying the nitrogen and oxygen uptake rates on (A) growth and (B) TAG exchange rates fixing glucose as the carbon source.

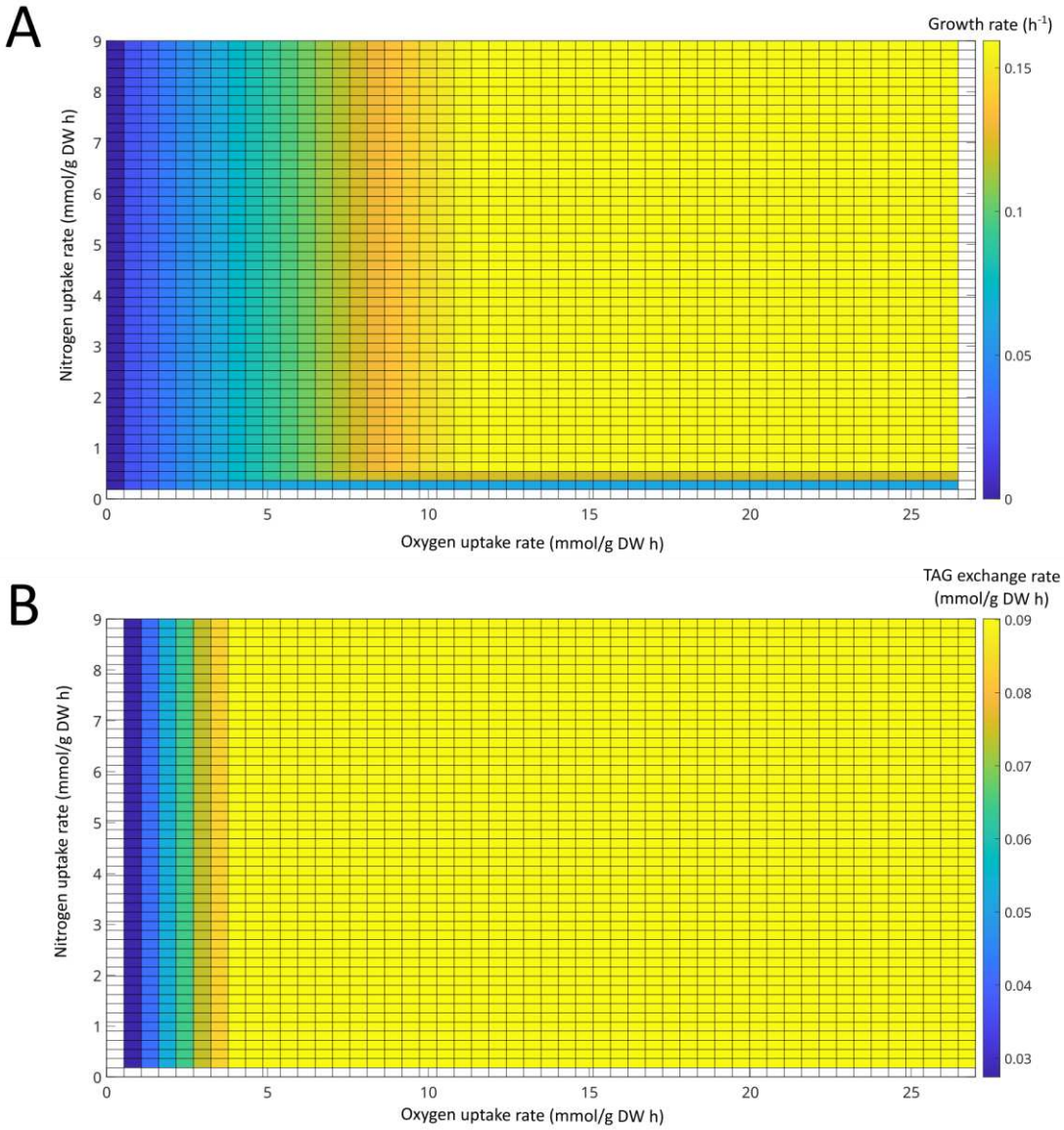


Figure S6 - Effects of varying the nitrogen and oxygen uptake rates on (A) growth and (B) TAG exchange rates fixing xylose as the carbon source

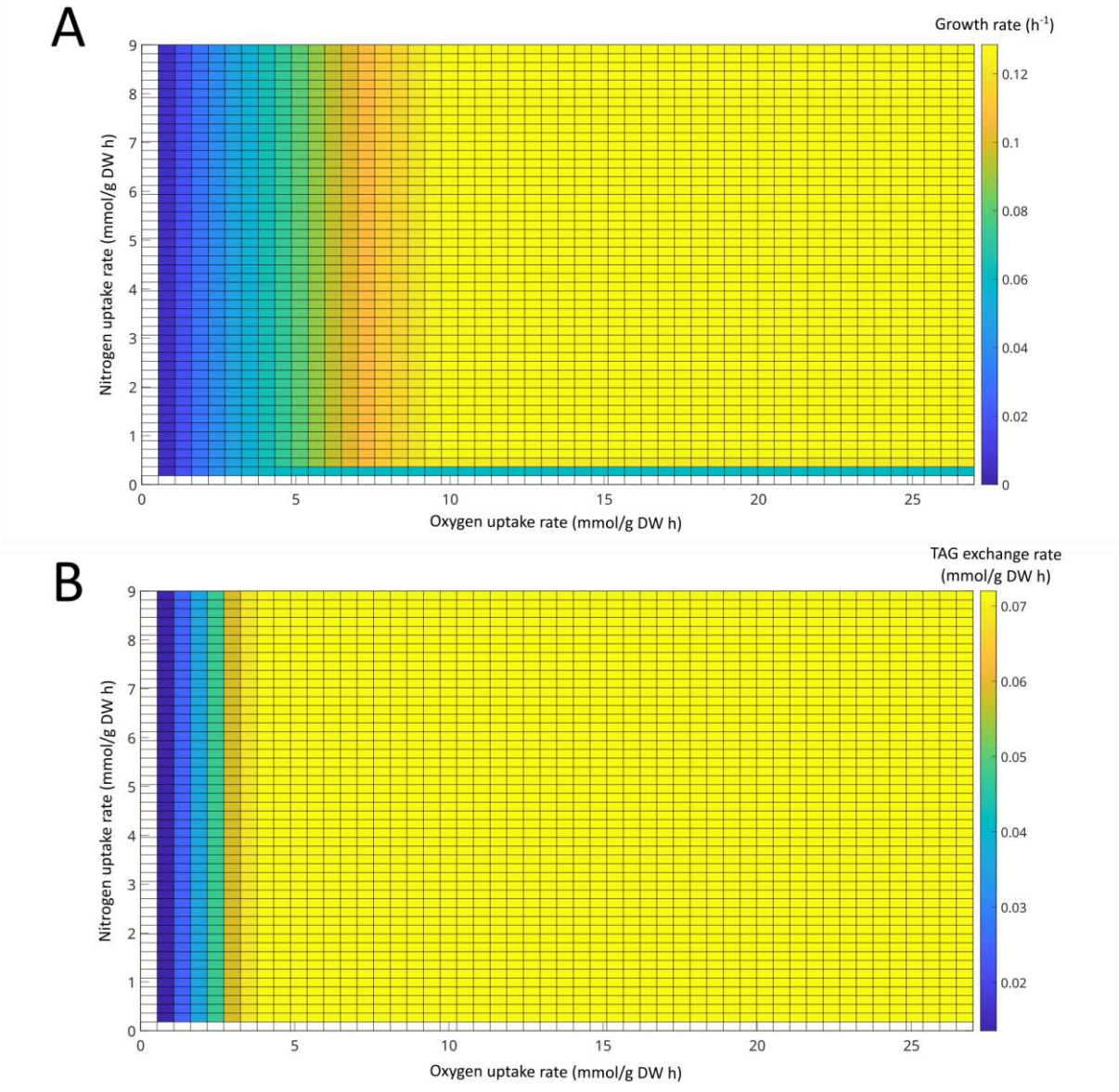
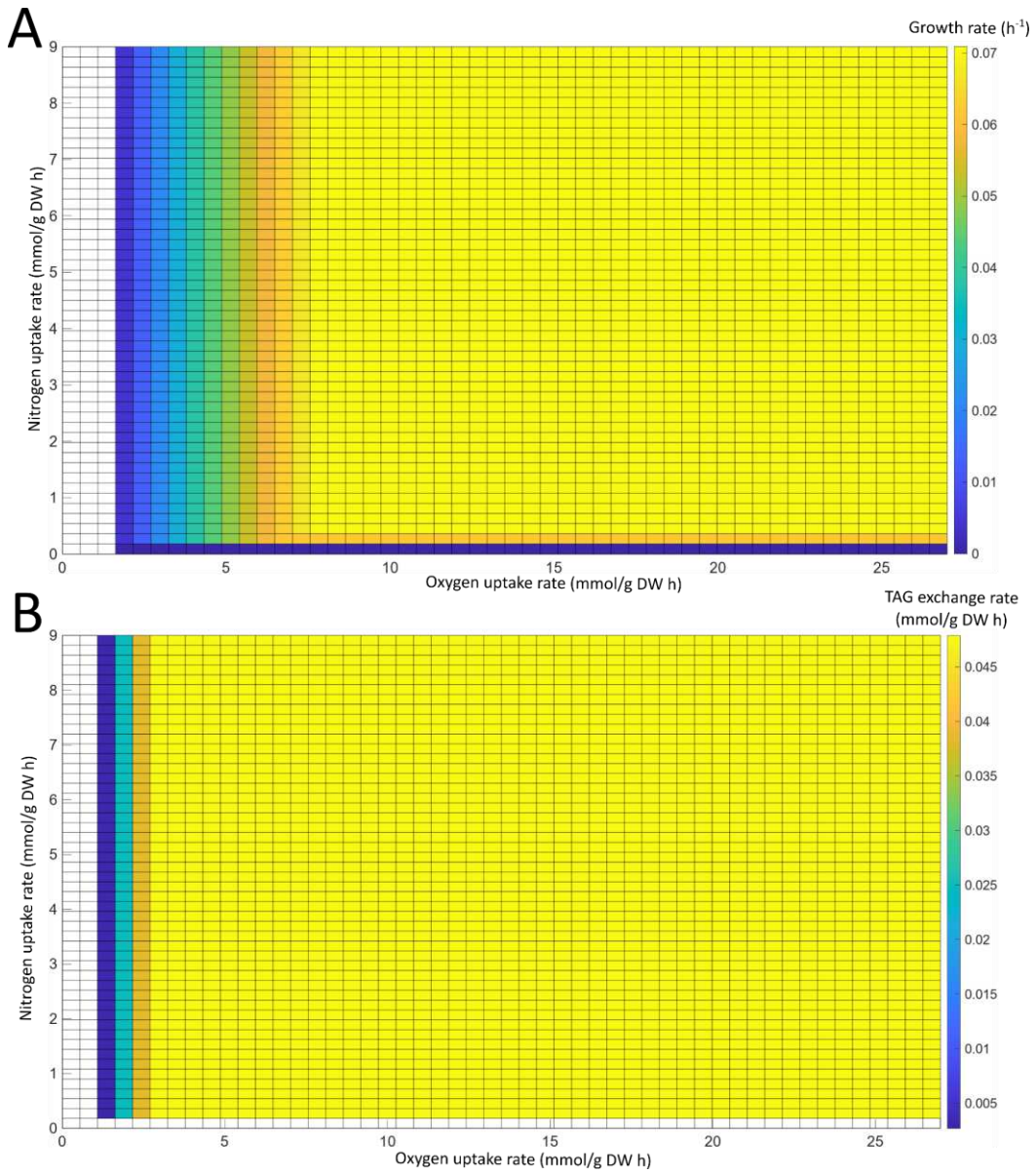


Figure S7 - Effects of varying the nitrogen and oxygen uptake rates on (A) growth and (B) TAG exchange rates fixing glycerol as the carbon source.



CHAPTER 3 - *Papiliotrema laurentii*: GENERAL FEATURES AND BIOTECHNOLOGICAL APPLICATIONS

Eduardo Luís Menezes de Almeida[1*], Rafaela Zandonade Ventorim[1*], Maurício Alexander de Moura Ferreira[1], Wendel Batista da Silveira[1✉]

[1]Department of Microbiology, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

*These authors have contributed equally to this mini-review.

✉**Corresponding author**

Wendel Batista da Silveira

Universidade Federal de Viçosa, Av. P. H. Rolfs, s/n, Viçosa, MG 36570-900, Brazil

Tel: +55-31-3612-2427

wendel.silveira@ufv.br

ORCID: 0000-0001-7869-8144

Published on Applied Microbiology and Biotechnology **-** https://doi.org/10.1007/s00253-022-12208-2

3.1. Abstract

Papiliotrema laurentii, previously classified as *Cryptococcus laurentii*, is an oleaginous yeast that has been isolated from soil, plants, and agricultural and industrial residues. This variety of habitats reflects the diversity of carbon sources that it can metabolize, including monosaccharides, oligosaccharides, glycerol, organic acids, and oils. Compared to other oleaginous yeasts, such as *Yarrowia lipolytica* and *Rhodotorula toruloides*, there is little information regarding its genetic and physiological characteristics. From a biotechnological point of view, *P. laurentii* can produce surfactants, enzymes, and high concentrations of lipids, which can be used as feedstock for fatty acid-derived products. Moreover, it can be applied for the biocontrol of phytopathogenic fungi, contributing to quality maintenance in post- and pre-harvest fruits. It can also improve mycorrhizal colonization, nitrogen nutrition, and plant growth. *P. laurentii* is also capable of degrading polyester and diesel derivatives and acting in the bioremediation of heavy metals. In this review, we present the current knowledge about the basic and applied aspects of *P. laurentii,* underscoring its biotechnological potential and future perspectives.

Keywords: Biotechnological applications, Genetic and physiological characteristics, Oleaginous yeast, *Papiliotrema laurentii*

3.2. Key points

- The physiological characteristics of *P. laurentii* confer a wide range of biotechnological applications
- The regulation of the acetyl-CoA carboxylase in *P. laurentii* is different from most other oleaginous yeasts
- The GEM is a valuable tool to guide the construction of engineered P. *laurentii* strains with improved features for bio-based products

3.3. Introduction

Oleaginous yeasts are a potential source of lipids for the production of biodiesel and other oleochemicals due to their capacity to store large amounts of triacylglycerols (TAGs) in lipid droplets (at least 20% of their dry weight) (Sreeharsha and Mohan 2020; Szczepańska et al. 2021). Lipid accumulation in oleaginous yeasts is favored by high carbon:nitrogen ratios, reaching lipid contents as high as 70% (Vasconcelos et al. 2019). Under these conditions, there is a shift in carbon flux from amino acids production to lipid synthesis (Kim et al. 2019). Importantly, the capacity of lipid accumulation is both yeast strain- and cultivation-dependent (Kamineni and Shaw 2020). Overall, the fatty acid composition of the TAGs accumulated in oleaginous yeasts is similar to plant seed oils used in the production of biodiesel (Vasconcelos et al. 2019). Therefore, yeast oils can meet the requirements related to the quality standard of this biofuel (Adrio 2017; Patel et al. 2017; Szczepańska et al. 2021). The most studied oleaginous yeast species belong to the genera: *Cryptococcus*, *Cutaneotrichosporon*, *Lipomyces*, *Rhodosporidium*, *Rhodotorula*, *Trichosporon, and Yarrowia*. An important characteristic of the oleaginous phenotype is the presence of the ATP:citrate lyase, which is responsible for catalyzing the conversion of citrate (transported from mitochondria to cytosol) into oxaloacetate and acetyl-CoA, increasing the availability of building blocks for fatty acid synthesis (Chaturvedi et al. 2022).

Papiliotrema laurentii, previously known as *Cryptococcus laurentii,* is a non-conventional oleaginous yeast belonging to the Basidiomycota phylum, Tremellomycetes class, and Tremellales order (Liu et al. 2015b). It is a non-motile, encapsulated, and dimorphic yeast (Kurtzman 1973). Moreover, it is capable of assimilating different sugars as carbon sources such as glucose, xylose, arabinose, cellobiose, mannose, galactose, rhamnose, sucrose, lactose, and galacturonic acid (Sitepu et al. 2014). This yeast is distributed in many habitats, including natural environments where it employs a saprotrophic lifestyle, such as soil (Vieira et al. 2020a) and decaying plant material (Yalçın et al. 2021). Also, *P. laurentii* can be an opportunistic pathogen for humans with suppressed immune systems (Londero et al. 2019).

P. laurentii displays potential for application in different biotechnological fields. As mentioned previously, it assimilates sugar constituents of lignocellulosic biomasses; therefore, it can be used for lipid production from these biomasses in biorefineries. Compared to other oleaginous yeasts, such as *Yarrowia lipolytica* and *Rhodotorula toruloides*, there is little information regarding its genetic and physiological characteristics (Table 1 and Fig. 1). The

current knowledge on the regulation of the oleaginous phenotype in *P. laurentii* is described in the section *Lipid accumulation physiology*.

Fig. 1 Gene expression and lipid metabolism in *Papiliotrema laurentii* grown under nitrogen deprivation. The figure also depicts similarities and differences between *P. laurentii*, *R. toruloides* and *Y. lipolytica*. Under nitrogen-limiting conditions, the expression of malic enzyme (ME) is upregulated (green) and acetyl-CoA carboxylase (ACC) is downregulated (red). Otherwise, the cytosolic ME is not expressed in *Y. lipolytica* (##). The green arrows in the tricarboxylic acid (TCA) cycle indicate the accumulation of citrate and cis-aconitate in nitrogen-limiting conditions. Acetyl-CoA for lipid synthesis is mainly provided by the reaction catalyzed by ATP:citrate lyase (ACL) using the citrate accumulated in the mitochondria and transported to the cytosol; however, acetyl-CoA seems to be also formed from the leucine catabolism. So far there is no information regarding the post-traditional regulation of ACL and ACC in *P. laurentii*. In *Y. lipolytica*, these enzymes are phosphorylated when it is cultivated under nitrogen limitation (##). ACL, ACC, ME, and isocitrate dehydrogenase (IDH) are not transcriptionally regulated in *R. toruloides* (**). There is no evidence for the excretion of citrate in *P. laurentii* even at high carbon concentrations (##). In *Y. lipolytica*, citrate accumulates and is excreted under high C/N ratios. The red dashed box highlights the main source of NADPH in *P. laurentii*, *R. toruloides* and *Y. lipolytica* during the lipid accumulation phase. In *P. laurentii*, the reaction catalyzed by the ME plays a secondary role. Similar to *Y. lipolytica* (#), the cell requirement for ammonia under nitrogen limitation is likely provided by the breakdown of purines, such as AMP, which downregulates the activity of IDH and triggers the accumulation of citrate in *P. laurentii*. During nitrogen starvation, the expression of genes encoding transmembrane transporters, lipid homeostasis and β-oxidation genes, acetyl-CoA dehydrogenase, amino acid permeases, and chitin deacetylase is upregulated, while the expression of genes involved with the translation machinery is downregulated. The translation machinery is also downregulated in *Y. lipolytica* and *R. toruloides* under nitrogen deprivation (*, #). For more details regarding the similarities and differences between *P. laurentii*, *Y. lipolytica* and *R. toruloides* see Table 1.

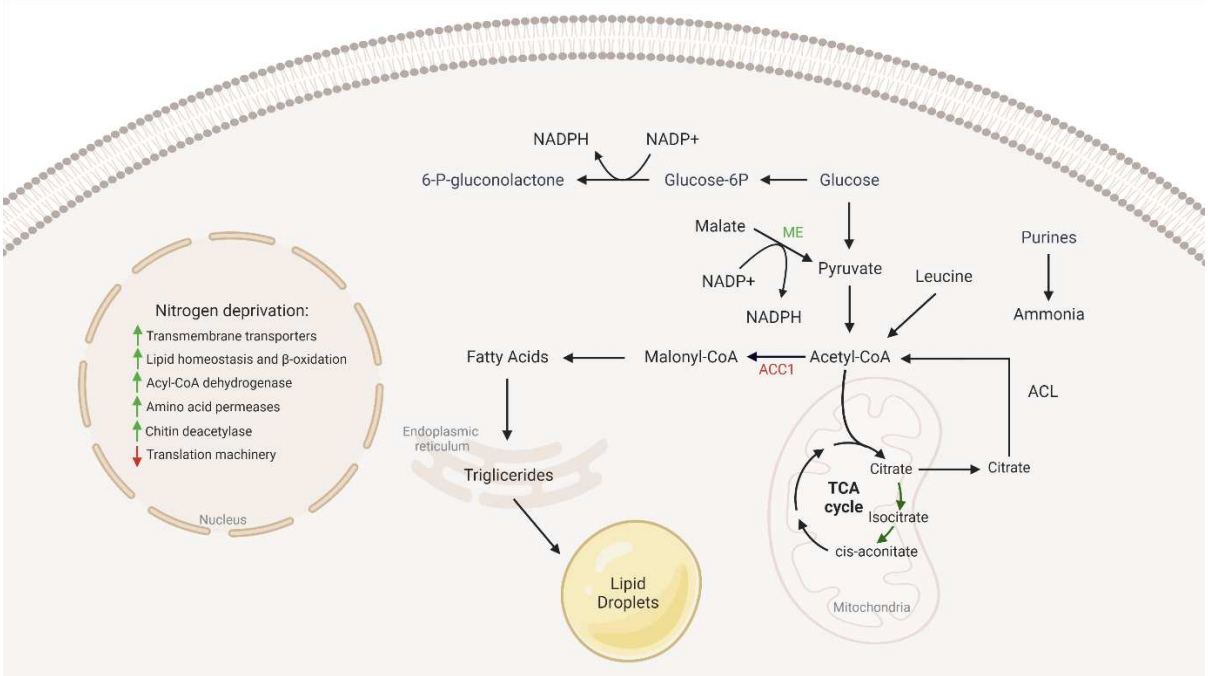


Table 1. Comparison of genetic and physiological features between *Papiliotrema laurentii*, *Rhodotorula toruloides*, and *Yarrowia lipolytica.*

Features	*Papiliotrema laurentii*	*Rhodotorula toruloides*	*Yarrowia lipolytica*
Genome	3 genomes are available in the NCBI database. Reference genome: 19.02 Mb, 56% GC (Bijlani et al. 2020)	20 genomes available. Reference genome: 20.2 Mb, 62% GC, 8.171 genes (Zhu et al., 2012)	27 genomes available. Reference genome: 20.5 Mb, 49% GC, 7,144 genes (Dujon et al. 2004)
Genome-scale metabolic model	papla-GEM (Ventorim et al. 2022)	rhto-GEM (Tiukova et al. 2019) iRhto1108 (Dinh et al. 2019) Rt_IFO0880 (Kim et al. 2021) ecRhtoGEM (Rekena et al. 2021) – rhto-GEM improved with enzymatic constraints	iNL895 (Loira et al. 2012) iYL_619 (Pan et al. 2012) iMK735 (Kavščček et al. 2015) iYali4 (Kerkhoven et al. 2016) iYL_2.0 (Wei et al. 2017) iYLI647 (Mishra et al. 2018) M21iYL (Czajka et al. 2021) – previous models improved with machine learning ec*i*Yali (Domenzain et al. 2022) – iYali4 improved with enzymatic constraints
Carbon source assimilation	Monosaccharides (hexoses and pentoses), oligosaccharides, glycerol, organic acids, oils, polymers derived from agricultural and/or industrial wastes (Carota et al. 2017; Castanha et al. 2014; Fonseca et al. 2011; Sitepu et al. 2014; Wang et al. 2018)	Monosaccharides (hexoses and pentoses), oligosaccharides, alcohols, organic acids, long-chain fatty acids, biomass hydrolysates, waste glycerol and gas fermentation products (Wen et al. 2020)	Monosaccharides (hexoses and pentoses), oligosaccharides, alcohols, organic acids. Although possessing all genes required for xylose assimilation, it is naturally unable to assimilate this sugar (Ledesma-Amaro et al. 2016). It is known for its capacity to metabolize hydrophobic compounds, including fatty acids, triacylglycerols, and alkanes (Liu et al. 2015a)
Highest lipid content	63.5% of CDW at 48h of cultivation with a culture medium containing xylose as the sole carbon source under optimized conditions of cultivation (Vieira et al. 2020a)	52.8% of CDW at 72 h with detoxified sugarcane bagasse hydrolysate as carbon source (Bonturi et al. 2017)	58.5% of CDW at 4 d with sugarcane bagasse hydrolysate as carbon source (Tsigie et al. 2011)
Tolerance to inhibitory compounds formed during the pretreatment of lignocellulosic biomasses	Can tolerate up to 1.0 g/L furfural and 2.5 g/L HMF (Sitepu et al. 2014) Unable to grow in concentrations of formic and acetic acid as low as 0.1 and 1.0 g/L, respectively (Almeida et al. 2022)	Able to grow in hydrolysate containing 4.0 g/L acetic acid, 2.0 g/L furfural, 2.5 g/L HMF, 8.7 g/L (Lopes et al. 2021); or minimal media with 10 g/L acetic acid or 2,0 g/L furfural (Díaz et al. 2018)	Tolerates up to 3.4 g/L acetic acid and 0.9 g/L formic acid; 0.15 g/L cinnamic acid and 0.18 g/L coniferyl aldehyde; 2.9 g/L furfural and 3.8 g/LHMF (Konzock et al. 2021)

Regulation of oleaginous phenotype	Upregulation of β-oxidation to provide acetyl-CoA and energy required to maintain cell growth during nitrogen deprivation (Sarkar et al. 2018) Accumulation of intermediates of the TCA cycle when cultivated under nitrogen-limiting conditions (Vieira et al. 2020b) G6PDH from PPP main source of NADPH, ME also contributes to the NADPH supply (Vieira et al. 2020b) ACL abundance appears not to be necessary for lipid accumulation (Vieira et al. 2020b) The expression of *ACC* gene encoding acetyl CoA carboxylase decreases during the lipid accumulation phase; it is still unknown mechanisms involved with the post-translational regulation of this enzyme (Vieira et al. 2020b) Citrate was not accumulated in the cytosol, indicating an efficient conversion into oxaloacetate and acetyl-CoA under high C:N ratio (Vieira et al. 2020b)	Inactivation of TORC1, which inhibits the transcription of genes related to translation initiation and ribosome biosynthesis, while activates autophagy-related ones (Zhu et al. 2012) *FAS* and *PYC* genes are upregulated, but *ACC*, *IDH*, *ME*, and *ACL* are not transcriptionally regulated (Zhu et al. 2012) The required NADPH is mainly supplied by the PPP and partially by the cytosolic ME (Bommareddy et al. 2015) Induction of proteins involved in β-oxidation (Tiukova et al. 2019)	Absence of expression of a cytosolic ME (Morin et al. 2011) PPP is the primary source of NADPH and limits lipid synthesis (Wasylenko et al. 2015) ACL, ACC, and lecithin cholesterol acyl transferase are phosphorylated under nitrogen limitation (Pomraning et al. 2016) Accumulation of intermediates of the TCA cycle when cultivated under nitrogen-limiting conditions (Beopoulos et al. 2009; Kerkhoven et al. 2017) Downregulation of genes related to protein synthesis and amino acid metabolism (Lazar et al. 2018) Upregulation of genes involved in protein turnover and autophagy (Lazar et al. 2018) Excretion of citrate in the presence of excessive carbon sources (Gálvez-López et al. 2019)

Abbreviations: CDW – cell dry weight; TCA – tricarboxylic acid; G6PDH – glucose-6-phosphate dehydrogenase; PPP – pentose phosphate pathway; ME – malic enzyme; IDH – isocitrate dehydrogenase; FAS – fatty acid synthase; PYC – pyruvate carboxylase; HMF – hydroxymethylfurfural

P. laurentii has also been used for biocontrol of phytopathogenic fungi (Wei et al. 2014), modulation of plant growth (Moller et al. 2016), degradation of industrial polymers and diesel (Barlow et al. 2020; Chandran and Das 2012; Hung et al. 2019), and bioremediation of heavy metals (Sarkar et al. 2019b) discussed in detail in the section *Other biotechnological applications*. In this context, this review focuses on the current knowledge about the physiology, genetics, ecology, and biotechnological applications of *P. laurentii*.

3.4. Natural history and ecology

The genus *Papiliotrema* was created in 2002 to house the newly described *P. bandonii* (Sampaio et al. 2002). This genus was first described morphologically and was named after the butterfly-like shape of the conidia and the mellaceous nature of the fungus. Its anamorphic stage was classified in the genus *Cryptococcus*. Molecular phylogenetics using rRNA markers performed by Sampaio et al. (2002) placed the genus *Papiliotrema* in the order Tremellales, itself belonging to the class Tremellomycetes and division Basidiomycota. The yeast *Papiliotrema laurentii* was first described as *Torula laurentii* by Kufferath (1920). Then, it was reclassified as *Torulopsis laurentii* by Diddens and Lodder (1934), and later as *Cryptococcus laurentii* by Skinner (1950). In 2015, an updated classification for the Tremellomycetes class was proposed by Liu et al. (2015b). They performed phylogenetic analyses using a dataset consisting of most species inside Tremellomycetes, with seven genes selected as molecular markers. They found that species from *Papiliotrema*, *Auriculibuller*, 16 *Cryptococcus* species, and 3 *Bullera* species formed a monophyletic clade, which were combined in a single genus that took the name *Papiliotrema*, as it preceded the other genus' names and contained the holotype for the *Papiliotrema* genus, *P. bandonii*, resulting in *Cryptococcus laurentii* being renamed to *Papiliotrema laurentii*.

P. laurentii has been found inhabiting a myriad of habitats. They include natural and agricultural environments such as rupestrian field soil (Vieira et al. 2020a), decaying mushrooms on trunks of living trees (Middelhoven 2004), tree barks and rotten fruits (Yalçın et al. 2021), the rhizosphere of blue lupin (Moller et al. 2016), the surface of wheat and corn kernels (Kurtzman 1973), vineyards (Wang et al. 2018), the surface of *Solanum torvum* leaves (Sitepu et al. 2014), *Populus tremuloides* exudates (Sitepu et al. 2013) and avian excrement (Brito et al. 2019). Other environments where it is found include hydrocarbon-contaminated soils (Chandran and Das 2012), kombucha tea (Chakraborty et al. 2016), sugarcane bagasse

(Gebbie et al. 2020), and palm oil (Polburee et al. 2015). *P. laurentii* can also be found in man-made environments, such as the internal surfaces of aircrafts (Hung et al. 2019), and even on surfaces inside the International Space Station (Satoh et al. 2021).

Although rare, *P. laurentii* can also be found in clinical cases as an opportunistic pathogen, causing cutaneous infections, fungemia, and meningitis (Intra et al. 2021). Its infections in humans are often associated with immunocompromised individuals, neoplasia, and multi-comorbidities (Londero et al. 2019; Al-Otaibi et al. 2021). The ecological niches adopted by *P. laurentii* are environments rich in decomposing matter, characteristic of saprophytic fungi. This is reflected in the capability of assimilating a wide range of carbon sources and its usage of many extracellular enzymes (Sitepu et al. 2014; Yalçın et al. 2021).

3.5. Genetic and Genome Characteristics, and Genome-Scale Metabolic Model

There currently are three genomes of *P. laurentii* available on the NCBI database (https://www.ncbi.nlm.nih.gov/datasets/genomes/?taxon=5418), belonging to the strains: IF7SW-F4, IF7SW-B5 and RY1. The reference genome, ASM1292261v1, belongs to the strain IF7SW-F4 isolated from the International Space Station (Bijlani et al. 2020). It comprehends 19.02 Mbp with a GC content of 56.22%. Furthermore, there are nine bioprojects available in the NCBI database (https://www.ncbi.nlm.nih.gov/bioproject/?term=Papiliotrema%20laurentii), including: genome sequencing of *P. laurentii* UFV-1, 5307AH and RY1 strains; and transcriptome or gene expression of 5307AH and RY1 strains. However, there is no information about the number of chromosomes of *P. laurentii*, and all its sequenced genomes are drafts assembled to contig or scaffold levels. Regarding the epigenetic characteristics, Tang et al. (2012) reported that *P. laurentii* presents 0.079% of cytosine methylation (5mC) on its genome under standard conditions.

Recently, the first genome-scale metabolic model (GEM) of *P. laurentii* was published (Ventorim et al. 2022). The model was based on homology reconstruction using *P. laurentii* UFV-1 genome (Vieira et al. 2020b) as the base, and using *Rhodotorula toruloides* (Tiukova et al. 2019) and *Yarrowia lipolytica* (Kerkhoven et al. 2016) GEMs as templates. The biomass equation was based on the quantification of biomass components of *P. laurentii* UFV-1 under different growth conditions. A total of 796 genes associated with 2465 reactions and 2127 metabolites are present in the model and represent the growth and lipid accumulation

physiology of this yeast. Therefore, this model is a valuable tool to guide the construction of engineered *P. laurentii* strains suitable to be applied as single cell oil (SCO) platforms. The model is available in a GitHub repository (https://github.com/SysBioChalmers/papla-GEM).

3.6. Lipid accumulation physiology

Similar to other oleaginous yeasts, nitrogen-limiting conditions are also pivotal to favor the lipid accumulation in *P. laurentii*. It has been reported in oleaginous yeasts that nitrogen depletion causes a decrease in intracellular adenosine monophosphate (AMP) concentration. As a result, isocitrate dehydrogenase (IDH) activity is decreased, which in turn leads to the accumulation of citrate in the mitochondria. Excess citrate is transported to the cytosol, where it is cleaved by ATP:citrate lyase (ACL) to acetyl-CoA and oxaloacetate. The acetyl-CoA produced is then directed towards *de novo* fatty acid synthesis by the fatty acid synthase (FAS) complex. The products released by the FAS complex, palmitoyl-CoA and stearoyl-CoA, are transported to the endoplasmic reticulum where they are used to produce TAGs, or undergo NADPH-dependent desaturation and/or a two-carbon elongation before being targeted to produce lysophosphatidic acid, phosphatidic acid, and diacylglycerol (DAG). Accumulated TAGs are stored in lipid droplets (Adrio 2017; Ledesma-Amaro and Nicaud 2016; Patel et al. 2016).

In *P. laurentii* UFV-1, phylogenetic analyses based on sequences of enzymes related to the oleaginous phenotype including acetyl-CoA carboxylase (ACC), ATP-citrate lyase (ACL), glucose-6P dehydrogenase (G6PDH), cytosolic NADP-dependent isocitrate dehydrogenase (NADP-ICDH), mitochondrial NAD-dependent isocitrate dehydrogenase (NAD-ICDH), and cytosolic malic enzyme (ME), surprisingly revealed an evolutionary proximity with *Cryptococcus neoformans* and *Kwoniella mangroviensis,* currently characterized as non-oleaginous. Therefore, the oleaginous phenotype might have evolved in *P. laurentii* after the divergence of the genera in the phylum Basidiomycota (Vieira et al. 2020b). Similar to other oleaginous yeasts, the lipid accumulation phase in *P. laurentii* begins in the late exponential phase (Vieira et al. 2020a; Vieira et al. 2020b).

In the lipid accumulation phase, the expression of ACC surprisingly decreases in *P. laurentii* (Fig. 1). This enzyme catalyzes the conversion of acetyl-CoA to malonyl-CoA, which is the first reaction that limits the carbon flux for fatty acid (FA) biosynthesis. So far, there is no information about the post-translational regulation of this enzyme in *P. laurentii*. Vieira et

al. (2020b) suggested that post-translational modifications of ACC could take place to maintain the required malonyl-CoA provision for fatty acid synthesis. Hence, it is important to investigate the regulation of ACC in *P. laurentii* in order to better understand the regulation of the first and rate-limiting step of the fatty acid biosynthesis in this yeast.

Moreover, the high expression of the gene encoding glucose-6-phosphate dehydrogenase during the lipid accumulation phase of *P. laurentii* indicates that the first reaction of the pentose phosphate pathway is the main NADPH source in this yeast. Importantly, the ME might also contribute to the NADPH supply as its expression increases during the lipid accumulation phase (Fig. 1). The expression of ACL does not change in the lipid accumulation phase of *P. laurentii* UFV-1, suggesting that an increase in ACL abundance is not required for lipid accumulation. Indeed, citrate was not accumulated in the cytosol of *P. laurentii* UFV-1, indicating that it was efficiently converted into oxaloacetate and acetyl-CoA in conditions with high sugar concentrations and C/N ratios (Vieira et al. 2020b).

Oleaginous yeasts, such as *P. laurentii* (Vieira et al. 2020b) and *Y. lipolytica* (Beopoulos et al. 2009; Kerkhoven et al. 2017), accumulate some intermediates of the tricarboxylic acid (TCA) cycle when cultivated under nitrogen-limiting conditions. In *P. laurentii* UFV-1, cis-aconitate and isocitrate accumulate in the lipid production phase (Fig. 1), indicating that the AMP-dependent isocitrate dehydrogenase is inhibited due to the AMP depletion in response to nitrogen limitation (Vieira et al. 2020b). Importantly, these authors reported that amino acids are also metabolized to provide acetyl-CoA for FA synthesis and ammonia to other metabolic pathways.

The strain *P. laurentii* RY1 isolated from Kombucha tea with low nitrogen availability can maintain cell viability without nitrogen and grow under nitrogen-limiting conditions (Sarkar et al. 2019a). RNA-seq analysis of this strain revealed that the nitrogen depletion induced the expression of genes encoding transmembrane transporters and proteins involved with lipid homeostasis and fatty-acid β-oxidation. Moreover, the translation machinery was repressed under this stress condition (Sarkar et al. 2018). The authors suggested that the upregulation of β-oxidation takes place to provide acetyl-CoA and generate the energy required to maintain cell growth during nitrogen deprivation. Indeed, the acyl-CoA dehydrogenase, an important enzyme in peroxisomal β-oxidation, was also upregulated in this condition. Furthermore, these authors pointed out that the overexpression of amino acid permeases seems to be a response of *P. laurentii* to maintain the nitrogen balance.

The lipid production by *P. laurentii* can reach levels ranging from 26.6 to 63.5% (g lipid/g dry weight) (Wang et al. 2018; Polburee et al. 2015; Castanha et al. 2014; Carota et al. 2017; Sitepu et al. 2013; Sitepu et al. 2014; Vieira et al. 2020a). The variety of habitats from which *P. laurentii* strains have been isolated reflects the diversity of carbon sources that it can metabolize, including monosaccharides, oligosaccharides, glycerol, organic acids, oils, and polymers derived from agricultural (e.g. lignocellulosic biomasses), and/or industrial wastes (e.g. polymers, oils) (Fonseca et al. 2011), which are promising feedstocks for oil production. *P. laurentii* UFV-1 achieved 40 and 30% lipid content in batch (shake flask) cultivations in culture media containing glucose and xylose, respectively (Vieira et al. 2020a). Remarkably, this strain reaches 63.5% lipid content when cultivated under optimized conditions using xylose as carbon source. *P. laurentii* AM113 cultivated in a culture medium supplemented with inulin hydrolysate in fed-batch modes reached 54.6% (g lipid/ g biomass) (Wang et al. 2018). Castanha et al. (2014) obtained 27.8% lipid content in *P. laurentii* UNESP 11 using cheese whey, an important effluent generated in cheese manufacture, as fermentation medium. When it was cultivated in diluted ricotta cheese whey supplemented with $(NH_4)_2SO_4$, *P. laurentii* UCD 68-201 accumulated 62.6% lipids (Carota et al. 2017). In batch cultivation using alkaline pretreated corn stover, *P. laurentii* UCDFST 12-803 accumulated 26.6% lipid content after 168 h of cultivation (Sitepu et al. 2014). Polburee et al. (2015) showed lipid accumulation of 28.4% (g lipid/g biomass) in *P. laurentii* DMKU AmC14 after 120 h of cultivation using a synthetic medium containing glycerol as carbon source.

So far, few studies have analyzed the responses of *P. laurentii* to stress conditions found in both environments and bioprocesses. Sitepu et al. (2014) qualitatively evaluated the growth of the strain UCDFST 12-803 in response to inhibitors derived from the treatment of lignocellulosic biomasses. They showed that it did not grow in the presence of 1.0 g/L of furfural and 2.5 g/L of acetic acid, but tolerated up to 2.0 g/L of hydroxymethylfurfural (HMF). However, the mechanisms involved with this tolerance remain unclear. Almeida et al. (2022) also reported a similar tolerance of *P. laurentii* UFV-1 to HMF and susceptibility to furfural. These authors showed that *P. laurentii* did not grow in concentrations as low as 1,0 g/L of acetic acid. By using adaptive laboratory evolution, they selected three evolved strains of *P. laurentii* more tolerant to acetic acid (up to 2,0 g/L). These strains were capable of detoxifying the medium during the transition from the lag phase to the exponential growth phase. One of them stood out due to its capacity of tolerating acetic acid and maintaining the oleaginous phenotype, opening perspectives of its application in lignocellulosic biorefineries; therefore, it had its

genome sequenced in order to identify genetic alterations related to acetic acid tolerance. Genomic analyses revealed that, compared to the parental strain (more susceptible to acetic acid), the promising evolved strain had mutations in genes encoding proteins related to multidrug resistance transporters, energy metabolism, detoxification, coenzyme recycling, and cell envelope.

It is noteworthy that the *P. laurentii* fatty acid profile is suitable for biodiesel production (Carota et al. 2017; Castanha et al. 2014; Vieira et al. 2020a; Wang et al. 2018). Additionally, the predicted properties of biodiesel produced from *P. laurentii* oil, which are based on its fatty acid profile, indicated that it was in agreement with the standards of quality (regarding European and Brazilian regulation agencies) (Wang et al. 2018; Vieira et al. 2020a).

3.7. Other biotechnological applications

3.7.1. Biocontrol and plant growth promotion

The ability of *P. laurentii* to inhibit other fungi underscores its potential in the biocontrol of phytopathogenic fungi (Table 2). This contributes to the maintenance of quality in post- and pre-harvest fruits, including pears, sweet cherry, table grapes, and strawberries (Benbow and Sugar 1999; Meng et al. 2010; Sharma et al. 2009; Tian et al. 2004; Wei et al. 2014). Besides being antagonistic, this yeast has a strong ability to survive on fruit surfaces under field conditions and easily adapts to postharvest storage conditions, such as low temperature, low O_2, and high CO_2 concentrations (Tian et al. 2004). The most accepted mechanism of action by which *P. laurentii* suppresses postharvest diseases is competition for nutrients (Castoria et al. 1997). Castoria et al. (1997) reported that the presence of exogenous glucose reduced the protective effect conferred by *P. laurentii* against the development of apple wounds by *Penicillium expansum* and *Botrytis cinerea*, as the competition for nutrients between species decreased. On the other hand, *P. laurentii* quickly colonized apple wounds caused by *B. cinerea* in the absence of exogenous nutrients, indicating that it is able to effectively exploit the availability of endogenous nutrients and protect apples against advancing wounds (Castoria et al. 1997). Additionally, it was reported an increase in the activity of antioxidant enzymes and enzymes related to defense against *B. cinerea* and *Alternaria alternata* in cherry tomatoes (Lai et al. 2018).

Table 2 – Overview of biotechnological applications of *Papiliotrema laurentii*.

Application	Strain	Description	Reference
Lipid production	UNESP 11	Obtained 27.8% of lipid content in shake flask cultivation for 240h using cheese whey as cultivation media	Castanha et al. 2014
	UCDFST12-803	Obtained 26.6 % of lipid content in shake flask cultivation with alkaline pretreated corn stover for 168h	Sitepu et al. 2014
	DMKU AmC14	Achieve 28.4 % (g lipid/ g biomass) of lipid content after 120h of cultivation using synthetic media with glycerol	Polburee et al. 2015
	UCD 68-201	Used diluted ricotta cheese whey plus $(NH_4)_2SO_4$ as cultivation media and obtained 62.6 % of lipid content with 96h of bioreactor batch cultivation	Carota et al. 2017
	AM113	Fed-batch bioreactor cultivation using supplemented inulin hydrolysate for 132h reaches 54.6 % of lipid content	Wang et al. 2018
	UFV -1	Shake flask cultivation using xylose, under optimized conditions for 48h, reaches 63.5 % of lipid content	Vieira et al. 2020a
Biocontrol	LS-28	Suppression of the development of postharvest apple wounds caused by *P. expansum* and *B. cinerea*.	Castoria et al. 1997
	zju 10	High activity of antioxidant enzymes and enzymes related to defense against *B. cinerea* and *A. alternata* in cherry tomato	Lai et al. 2018
Plant growth promotion	n.d.	Improved growth of soybean colonized by *Glomus mosseae*	Sampedro et al. 2004
	MTCC 3953	Increased growth and nutrition of cowpea plants by interaction with mycorrhizal *Glomus mosseae*	Boby et al. 2008
	CAB 578	Improved photosynthetic water use of *A. betulina* plants increasing the rate of CO_2 assimilation and the concentration of phosphorus, iron, and manganese in the root	Cloete et al. 2010a,b
	CAB 91	Increased respiration, biological nitrogen fixation efficiency, and improved mycorrhizal colonization in Blue lupin plants	Moller et al. 2016
	CBS 139	Displayed promising plant growth promoting activities when isolated from the soil of *Camellia sinensis* plantation	Kumla et al. 2020

Enzyme production	n.d.	Thirty strains exhibited extracellular activities of amylase, esterase, lipase, protease, and chitinase	Buzzini and Martini 2002
	CBS 2174, 8645, 8648	Extracellular laccase activity High phytase activity with properties suitable for use as animal feed additive	Ikeda et al. 2002
	ABO 510	Good producer of intracellular phytase with suitable properties for its direct inclusion in the feeding of monogastric animals	Staden et al. 2007 Pavlova et al. 2008
	AL_{27}	Extracellular xylanase and β-D-xylosidase activities	
	UFMG-HB-48	Extra and intracellular hyaluronidase activity, and thermostability useful for hyaluronan oligomers synthesis	Lara et al. 2014 Smirnou et al. 2015
	DSMZ 70766		
	RY1	Nitrogen limitation induced chitin deacetylase over-expression useful in the process of synthesis of chitosan	Chakraborty et al. 2016
	DMLU-Y34-A, Y38-C	Extracellular enzymatic activities of lipase, protease, amylase, and xylanase	
	n.d.	Several isolates exhibited extracellular laccase activity induced by copper	Jaiboon et al. 2016
	n.d.	Several *P. laurentii* isolates exhibited the extracellular activities of β-glucosidase, protease, lipase, amylase, xylanase, and phytase	Leguina et al. 2019 Yalçın et al. 2021
Bioremediation	RY1	Cells remained viable in a nitrogen-starved environment with high levels of lead, arsenic, and chromium, activating mechanisms of detoxification and antioxidant response	Sarkar et al. 2019b
	RY1	Immobilized and non-immobilized yeast were able to remove lead from water in neutral and alkaline environments	Mukherjee et al. 2020
Polymer degradation	n.d.	Degradation 97% of the diesel added to the culture medium after 10 days of cultivation	Chandran and Das 2012
	5307AH	Efficient degradation of polyester polyurethane material and polyethylene succinate (PES) coatings	Hung et al. 2019
Surfactant production	L62	Emulsifying and tensioactive activities observed in the fermentation supernatant using xylose	Chaves et al. 2021

n.d. not described

In the rhizosphere, *P. laurentii* can form single and dual biofilms with other microorganisms (e.g. *Pseudomonas capeferrum*), influencing its capacity of acting as a biocontrol agent and beneficial microorganisms for the host plant (Lacosegliaz et al. 2021). Another study reported the responses of *P. laurentii* to copper, commonly used in the form of copper sulfate to inhibit the growth of phytopathogenic fungi in horticulture. Leguina et al. (2019) isolated copper-resistant (up to 10 mM of copper sulfate) strains of *P. laurentii* from the rhizosphere of tomato plants. The presence of copper promoted the production of auxin and increased the activity of laccase and catechol oxidase, which are enzymes involved in the oxidation of phenolic compounds such as lignins, indicating that *P. laurentii* presents properties of ecological and agronomical interest.

P. laurentii is commonly found in a wide variety of agricultural and managed soils (Vadkertiová et al. 2017) and is described as a plant growth-promoting yeast (Boby et al. 2008; Cloete et al. 2009; Cloete et al. 2010a,b; Moller et al. 2016; Sampedro et al. 2004) evidencing its potential as a bio-fertilizer. For example, blue lupin (*Lupinus angustifolius* L.) plants treated with the rhizosphere yeast *P. laurentii* CAB 91 showed greater growth, which was related to increased respiration, biological nitrogen fixation efficiency, and improved mycorrhizal colonization (Moller et al. 2016). *P. laurentii* CAB 578 improved the photosynthetic water-use efficiency of *Agathosma betulina* Pillans plants growth under nutrient-poor conditions by increasing the rate of CO_2 assimilation and the concentration of phosphorus, iron, and manganese in the root portion (Cloete et al. 2010a,b). Moreover, root growth was also related to the ability of *P. laurentii* to produce polyamines (Cloete et al. 2009). *P. laurentii* strains can also enhance the growth of leguminous plants due to its interaction with mycorrhizal fungi (Sampedro et al. 2004; Boby et al. 2008). *P. laurentii* CBS 139 isolated from the soil of *Camellia sinensis* var. assamica plantations can also promote plant growth. This capacity seems to be associated with the production of indole-3-acetic acid (IAA), siderophores, ammonia, extracellular enzymes (amylase, lipase, pectinase, protease, and xylanase), as well as the solubilization of insoluble forms of phosphate [$Ca_3(PO_4)_2$], and zinc (ZnO and $ZnCO_3$) (Kumla et al. 2020).

3.7.2. Enzyme production

Extracellular enzymes of *P. laurentii* strains have a wide range of applications including cosmetics, textile, detergents, surfactants, pulp and paper industries, biodegradation, and waste

treatment (AMFEP 2015; Singh et al. 2016) (Table 2). *P. laurentii* strains isolated from peat in Thailand showed extracellular enzymatic activities of lipase, protease, amylase, and xylanase (Jaiboon et al. 2016). Thirty strains of *C. laurentii* (*P. laurentii*) isolated from different samples of water, soil, insect, and plant materials collected from natural Brazilian environments also displayed extracellular activities of amylase, esterase, lipase, protease, and chitinase (Buzzini and Martini 2002). Yalçın et al. (2021) evaluated the enzyme production by yeast strains isolated from tree barks and rotten fruit samples. The isolates identified as *P. laurentii* exhibited extracellular activities of β-glucosidase, protease, lipase, amylase, xylanase, and phytase.

C. laurentii (*P. laurentii*) ABO 510 isolated from soil in the Western Cape region showed a high phytase activity. The features of this phytase are suitable for application as an animal feed additive since it improves the bioavailability of phosphorus in plant feed (Staden et al. 2007). *C. laurentii* (*P. laurentii*) AL_{27} also produced phytase with suitable properties for its direct inclusion in the feeding of monogastric animals (Pavlova et al. 2008).

Among the xylanolytic yeasts isolated from decaying wood and sugarcane bagasse, *C. laurentii* (*P. laurentii*) UFMG-HB-48 showed the highest growth rate and cell density in xylan. This strain presents both extracellular xylanase and β-D-xylosidase activities (Lara et al. 2014). Furthermore, it presented, under xylose induction, the highest intracellular β-D-xylosidase activity among all yeasts tested.

C. laurentii (*P. laurentii*) DSMZ 70766 showed a high extra and intracellular hyaluronidase activity. These enzymes displayed the desirable feature of thermostability for the enzymatic synthesis of hyaluronan oligomers with application in pharmacy and cosmetics (Smirnou et al. 2015).

Laccase is another enzyme that can be identified in *P. laurentii* extracts (Ikeda et al. 2002; Leguina et al. 2019). Laccases can be applied in textile, food, pharmaceutical, cosmetic, and pulp and paper industries; bioremediation; biomedical and biofuel production; biosensors; and organic synthesis (Rodríguez-Couto 2019).

P. laurentii is also a good producer of chitin deacetylase (CDA) under nitrogen limitation (Chakraborty et al. 2016). This enzyme catalyzes the hydrolysis of N-acetylglucosamine (GlcNAc) to produce chitin, which is deacetylated to form chitosan. Chitosan has suitable properties that allow its application in water treatment, biomedicine and photography industries (Mukherjee et al. 2019; Crognale et al. 2022).

3.7.3. Bioremediation and polymer degradation

The ability of *P. laurentii* to produce extracellular enzymes enables the assimilation of a wide variety of carbon sources, including polymers such as polyesters, polyurethanes, and petroleum hydrocarbons (Barlow et al. 2020; Chandran and Das 2012; Hung et al. 2019). As such, it can be used in the remediation of industrial polymers. *P. laurentii* 5307AH, isolated from an aircraft (Hung et al. 2019) can actively degrade Irogran®, a thermoplastic polyester polyurethane that displays great resistance to biodegradation (Hung et al. 2019). Additionally, this yeast can also degrade polyethylene succinate (PES) releasing succinate. Importantly, the degradation of polyesters is dissociated from the *P. laurentii* growth. Therefore, even in environments where the yeast does not have sufficient or appropriate nutrients for its growth, it can still promote polymer degradation.

Besides plastic polymer degradation, *P. laurentii* can also be used to degrade diesel oil. Chandran and Das (2012) isolated a strain of *P. laurentii* able to degrade 97% of the diesel added to the culture medium, highlighting its potential in the bioremediation of petroleum oil-polluted environments.

Another important property of *P. laurentii* is its ability to bioremediate heavy metals remaining viable in nitrogen-starved environments with high levels of lead, arsenic, and chromium, showing significant biotolerant, biosorption, bioprecipitation properties, and bioaccumulation capabilities (Sarkar et al. 2019b). The mechanisms related to the maintenance of cell viability under these environmental conditions involve: i) the interaction of cell surface polysaccharides with part of the lead, generating pyromorphite and lead phosphates that are less soluble, bioavailable, and toxic, and with part of the chromium generating grimaldiite, reducing the Cr (VI) for the less toxic Cr (III); ii) part of the lead, chromium, and arsenic are internalized by yeast cells and activate an antioxidant defense mechanism, which is evidenced by the increase in the activities of antioxidant enzymes, such as catalase and superoxide dismutase. Additionally, Mukherjee et al. (2020) analyzed the capacity of *P. laurentii* RY1, in non-immobilized and immobilized agar-agar preparations, to remove lead from water. They verified lead removal in neutral and alkaline environments using batch and column modes. Also, the biobeads might be used for several cycles of biosorption, making it a cost-effective process.

3.7.4. Surfactant production

Papiliotrema laurentii strains are described as good surfactant producers (Table 2). Biosurfactants are eco-friendly products substitutes for synthetic surfactants due to their high biodegradability, reduced toxicity, and environmental compatibility (Chaves et al. 2021). When cultivated in semi-defined medium containing xylose as the sole carbon source, the Antarctic yeast *P. laurentii* L62 displayed tensioactive and emulsifying properties with a high xylose consumption (over 75%) (Chaves et al. 2021).

3.8. Conclusions and perspectives

P. laurentii displays potential for the production of large amounts of lipids, surfactants, and enzymes, as well as control of phytopathogenic fungi, contributing to quality maintenance in post- and pre-harvest fruits, bioremediation of heavy metals, and polymer degradation. Despite its remarkable biotechnological potential, little is known about its physiology and genetics. A better understanding of its molecular physiology will provide insights regarding the regulation of both carbon and nitrogen metabolism, as well as adaptive responses to stress conditions found in bioprocesses. Therefore, the reconstruction of its first genome-scale metabolic model (GEM) opens perspectives to better understand the regulation of its metabolism by using systems biology approaches. From a biotechnological point of view, this model will also be useful to guide future metabolic engineering studies, which is expected to improve the application of *P. laurentii* in bioprocesses. In this sense, it is important the development of molecular tools to boost the construction of engineered *P. laurentii* strains with improved capacity to produce bio-based products.

3.9. Authors Contribution

RZV and ELMA performed the literature search and wrote the review. MAMF contributed to the literature search and writing of the review. WBS designed the structure of the review, contributed to the writing and revised the review. All authors critically revised and approved the final manuscript.

3.10. Funding

This work was supported by the Brazilian National Council for Scientific and Technological Development (CNPq, Finance Code 140538/2021-6); the Foundation for Research Support of the State of Minas Gerais (FAPEMIG); and the Coordination for the Improvement of Higher Education Personnel (CAPES, Finance Code 001).

3.11. Availability of data and material

Not applicable.

3.12. Code availability

Not applicable.

3.13. Declarations

Ethics approval

Not applicable.

3.14. Consent to participate

Not applicable.

3.15. Consent for publication

All authors have read and approved the manuscript for publication.

3.16. Conflict of interest

The authors declare no competing interests.

3.17. References

Adrio JL (2017) Oleaginous yeasts: Promising platforms for the production of oleochemicals and biofuels. Biotechnol Bioeng 114:1915–1920. doi: 10.1002/bit.26337

Almeida E, Ventorim R, Ferreira M, Costa M, Mantovani H, da Silveira W (2022) New *Papiliotrema laurentii* UFV-1 strains with improved acetic acid tolerance selected by adaptive laboratory evolution. J Appl Microbiol. Submitted manuscript.

Al-Otaibi H, Asadzadeh M, Ahmad S, Al-Sweih N, Joseph L (2021) *Papiliotrema laurentii* fungemia in a premature, very low-birth-weight neonate in Kuwait successfully treated with liposomal amphotericin B. J Med Mycol 31:101123. doi: 10.1016/j.mycmed.2021.101123

AMFEP (2015) List of commercial enzymes. Association of Manufacturers and Formulators of Enzyme Products. https://amfep.org/_library/_files/Amfep_List_of_Enzymes_update_May_2015.pdf. Accesssed 29 March 2022

Barlow DE, Bi JC, Estrella L, Lu Q, Hung C, Nadeau LJ, Crouch AL, Russell JN, Crookes-goodson WJ (2020) Edge-localized biodeterioration and secondary microplastic formation by *Papiliotrema laurentii* unsaturated biofilm cells on polyurethane films. Langmuir 36:1596–1607. doi: 10.1021/acs.langmuir.9b03421%0A

Benbow JM, Sugar D (1999) Fruit surface colonization and biological control of postharvest diseases of pear by preharvest yeast applications. Plant Dis 83:839–844. doi: 10.1094/PDIS.1999.83.9.839

Beopoulos A, Cescut J, Haddouche R, Uribelarrea JL, Molina-Jouve C, Nicaud JM (2009) *Yarrowia lipolytica* as a model for bio-oil production. *Prog Lipid Res* 48:375–387. doi: 10.1016/j.plipres.2009.08.005

Bijlani S, Singh NK, Mason CE, Wang CCC, Venkateswaran K (2020) Draft genome sequences of Tremellomycetes strains isolated from the International Space Station. Microbiol Resour Announc 25:e00504-20. doi: 10.1128/MRA.00504-20

Boby VU, Balakrishna AN, Bagyaraj DJ (2008) Interaction between *Glomus mosseae* and soil yeasts on growth and nutrition of cowpea. Microbiol Res 163:693–700. doi: 10.1016/j.micres.2006.10.004

Bommareddy RR, Sabra W, Maheshwari G, Zeng AP (2015) Metabolic network analysis and experimental study of lipid production in *Rhodosporidium toruloides* grown on single and mixed substrates. Microb Cell Fact 14:1–13. doi: 10.1186/s12934-015-0217-5

Bonturi N, Crucello A, Viana AJC, Miranda EA (2017) Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a *Rhodosporidium toruloides* adapted strain. Process Biochem 57:16–25. doi: 10.1016/j.procbio.2017.03.007

Brito MO, Bessa MAS, Menezes RP, Röder DVDB, Penatti MPA, Pimenta JP, Dias PA, Aguiar F, Pedroso RS (2019) Isolation of *Cryptococcus* species from the external environments of hospital and academic areas. J Infect Dev Ctries 13:545–553. doi: 10.3855/jidc.10849

Buzzini P, Martini A (2002) Extracellular enzymatic activity profiles in yeast and yeast-like strains isolated from tropical environments. J Appl Microbiol 93:1020–1025. doi: 10.1046/j.1365-2672.2002.01783.x

Carota E, Crognale S, D'Annibale A, Gallo AM, Stazi SR, Petruccioli M (2017) A sustainable use of Ricotta Cheese Whey for microbial biodiesel production. Sci Total Environ 584–585:554–560. doi: 10.1016/j.scitotenv.2017.01.068

Castanha RF, Mariano AP, Morais LAS, Scramin S, Monteiro RTR (2014) Optimization of lipids production by *Cryptococcus laurentii* 11 using cheese whey with molasses. Brazilian J Microbiol 45:379–387

Castoria R, Curtis F, Lima G, Cicco V (1997) β-1,3-glucanase activity of two saprophytic yeasts and possible mode of action as biocontrol agents against postharvest diseases. Postharvest Biol Technol 12:293–300. doi: 10.1016/S0925-5214(97)00061-6

Chakraborty W, Sarkar S, Chakravorty S, Bhattacharya S, Bhattacharya D, Gachhui R (2016) Expression of a chitin deacetylase gene, up-regulated in *Cryptococcus laurentii* strain RY1, under nitrogen limitation. J Basic Microbiol 56:576–579. doi: 10.1002/jobm.201500596

Chandran P, Das N (2012) Role of plasmid in diesel oil degradation by yeast species isolated from petroleum hydrocarbon-contaminated soil. *Environ Techno*l 33:645-652. doi: 10.1080/09593330.2011.587024

Chaturvedi S, Bhattacharya A, Rout PK, Nain L, Khare SK (2022) An overview of enzymes and rate-limiting steps responsible for lipid production in oleaginous yeast. Ind Biotechnol 18:20-31. doi: 10.1089/ind.2021.0003

Chaves FS, Brumano LP, Marcelino PRF, da Silva SS, Sette LD, Felipe MGA (2021) Biosurfactant production by Antarctic-derived yeasts in sugarcane straw hemicellulosic hydrolysate. Biomass Conv Bioref. doi: 10.1007/s13399-021-01578-8

Cloete KJ, Valentine AJ, Botha A (2010a) Effect of the soil yeast *Cryptococcus laurentii* on the photosynthetic water and nutrient-use efficiency and respiratory carbon costs of a Mediterranean sclerophyll, *Agathosma betulina* (Berg.) Pillans. Symbiosis 51:245–248. doi: 10.1007/s13199-010-0084-0

Cloete KJ, Przybylowicz WJ, Mesjasz-Przybylowicz J, Barnabas AD, Valentine AJ, Botha A (2010b) Micro-particle-induced X-ray emission mapping of elemental distribution in roots of a Mediterranean-type sclerophyll, *Agathosma betulina* (Berg.) Pillans, colonized by *Cryptococcus laurentii*. Plant Cell Environ 33:1005–1015. doi: 10.1111/j.1365-3040.2010.02122.x

Cloete KJ, Valentine AJ, Botha A (2009) Evidence of symbiosis between the soil yeast *Cryptococcus laurentii* and a sclerophyllous medicinal shrub, *Agathosma betulina* (Berg.) Pillans. Microb Ecol 57:624–632. doi: 10.1007/s00248-008-9457-9

Crognale S, Russo C, Petruccioli M, D'Annibale A (2022) Chitosan production by fungi: Current state of knowledge, future opportunities and constraints. Fermentation 8: 76. doi: 10.3390/fermentation8020076

Czajka JJ, Oyetunde T, Tang YJ (2021) Integrated knowledge mining, genome-scale modeling, and machine learning for predicting *Yarrowia lipolytica* bioproduction. Metab Eng 67:227–236. doi: 10.1016/j.ymben.2021.07.003

Díaz T, Fillet S, Campoy S, Vázquez R, Viña J, Murillo J, Adrio JL (2018) Combining evolutionary and metabolic engineering in *Rhodosporidium toruloides* for lipid production with non-detoxified wheat straw hydrolysates. Appl Microbiol Biotechnol 102:3287–3300. doi: 10.1007/s00253-018-8810-2

Diddens, Lodder (1934) *Torulopsis laurentii* (Kuff.) Lodder, Verhandelingen Koninklijke Nederlandse Akademie van Wetenschappen Afdeling Natuurkunde 32: 160. [MB#269102].

Dinh H V., Suthers PF, Chan SHJ, Shen Y, Xiao T, Deewan A, Jagtap SS, Zhao H, Rao C V., Rabinowitz JD, Maranas CD (2019) A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. Metab Eng Commun 9:e00101. doi: 10.1016/j.mec.2019.e00101

Domenzain I, Sánchez B, Anton M, Kerkhoven EJ, Millán-Oropeza A, Henry C, Siewers V, Morrissey JP, Sonnenschein N, Nielsen J (2022) Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. Nat Commun 13:1–13. doi: 10.1038/s41467-022-31421-1

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, Montigny J De, Blanchin S, Beckerich J, Beyne E, Bleykasten C, Babour A, Boyer J, Cattolico L, Confanioleri F, Daruvar A De, Despons L, Fabre E (2004) Genome evolution in yeasts. Nature 430:35–44. doi: 10.1038/nature02579

Fonseca Á, Boekhout T, Fell JW (2011) *Cryptococcus* Vuillemin (1901). In: Kurtzman CP, Fell JW, Boekhout T (eds) Yeasts, 5th edn. Elsevier, London, pp. 1661–1737. doi: 10.1016/B978-0-444-52149-1.00138-5

Gálvez-López D, Chávez-Meléndez B, Vázquez-Ovando A, Rosas-Quijano R (2019) The metabolism and genetic regulation of lipids in the oleaginous yeast *Yarrowia lipolytica*. Brazilian J Microbiol 50:23–31. doi: 10.1007/s42770-018-0004-7

Gebbie L, Dam TT, Ainscough R, Palfreyman R, Cao L, Harrison M, O'Hara I, Speight R (2020) A snapshot of microbial diversity and function in an undisturbed sugarcane bagasse pile. BMC Biotechnol 20:12. doi: 10.1186/s12896-020-00609-y

Hung C-S, Barlow DE, Varaljay VA, Drake CA, Crouch AL, Russell JN, Nadeau LJ, Crookes-Goodson WJ, Biffinger JC (2019) The biodegradation of polyester and polyester polyurethane coatings using *Papiliotrema laurentii*. Int Biodeterior Biodegradation 139:34–43. doi: 10.1016/j.ibiod.2019.02.002

Ikeda R, Sugita T, Jacobson ES, Shinoda T (2002) Laccase and melanization in clinically important *Cryptococcus* species other than *Cryptococcus neoformans*. J Clin Microbiol 40:1214-1218. doi: 10.1128/JCM.40.4.1214-1218.2002

Intra J, Sarto C, Brambilla P (2021) A rare case of cutaneous *Papiliotrema* (*Cryptococcus*) *laurentii* infection in a 23-year-old Caucasian woman affected by an autoimmune thyroid disorder with hypothyroidism. Eur J Clin Microbiol Infect Dis 40:647–650. doi: 10.1007/s10096-020-04058-5

Jaiboon K, Lertwattanasakul N, Limtong P, Limtong S (2016) Yeasts from peat in a tropical peat swamp forest in Thailand and their ability to produce ethanol, indole-3-acetic acid and extracellular enzymes. Mycol Prog 15:755–770. doi: 10.1007/s11557-016-1205-9

Kamineni A, Shaw J (2020) Engineering triacylglycerol production from sugars in oleaginous yeasts. Curr Opin Biotechnol 62:239–247. doi: 10.1016/j.copbio.2019.12.022

Kavšček M, Bhutada G, Madl T, Natter K (2015) Optimization of lipid production with a genome-scale model of *Yarrowia lipolytica*. BMC Syst Biol 9:72. doi: 10.1186/s12918-015-0217-4

Kerkhoven EJ, Kim YM, Wei S, Nicora CD, Fillmore TL, Purvine SO, Webb-Robertson BJ, Smith RD, Baker SE, Metz TO, Nielsena J (2017) Leucine biosynthesis is involved in regulating high lipid accumulation in *Yarrowia lipolytica*. MBio 8:1–12. doi: 10.1128/ mBio.00857-17

Kerkhoven EJ, Pomraning KR, Baker SE, Nielsen J (2016) Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. Npj Syst Biol Appl 2:16005. doi: 10.1038/npjsba.2016.5

Kim M, Park BG, Kim E-J, Kim J, Kim BG (2019) In silico identification of metabolic engineering strategies for improved lipid production in *Yarrowia lipolytica* by genome-scale metabolic modeling. Biotechnol Biofuels 12:187. doi: 10.1186/s13068-019-1518-4

Kim J, Coradetti ST, Kim YM, Gao Y, Yaegashi J, Zucker JD, Munoz N, Zink EM, Burnum-Johnson KE, Baker SE, Simmons BA, Skerker JM, Gladden JM, Magnuson JK (2021) Multi-Omics driven metabolic network reconstruction and analysis of lignocellulosic carbon utilization in *Rhodosporidium toruloides*. Front Bioeng Biotechnol 8. doi: 10.3389/fbioe.2020.612832

Konzock O, Zaghen S, Norbeck J (2021) Tolerance of *Yarrowia lipolytica* to inhibitors commonly found in lignocellulosic hydrolysates. BMC Microbiol 21:1–10. doi: 10.1186/s12866-021-02126-0

Kufferath (1920) *Torula laurentii* Kuff., Bulletin de la Societé Royale des Sciences Medicales et Naturelles de Bruxelles 74: 38. [MB#252974].

Kumla J, Nundaeng S, Suwannarach N, Lumyong S (2020) Evaluation of multifarious plant growth promoting trials of yeast isolated from the soil of assam tea (*Camellia sinensis* var. assamica) plantations in Northern Thailand. Microorganisms 8:1168. doi: 10.3390/microorganisms8081168

Kurtzman CP (1973) Formation of hyphae and chlamydospores by *Cryptococcus laurentii*. Mycologia 65:388–395. doi: 10.1080/00275514.1973.12019447

Lacosegliaz MJ, Torres MA, Leguina ACdelV, Venturi IBV, de Figueroa LIC, Fernández PM, Nieto-Peñalver CG (2021) Copper sulfate inhibition of quorum sensing in *Pseudomonas capeferrum* is dependent on biotic interactions. Rhizosphere 20:100434. doi: 10.1016/j.rhisph.2021.100434

Lai J, Cao X, Yu T, Wang Q, Zhang Y, Zheng X, Lu H (2018) Effect of *Cryptococcus laurentii* on inducing disease resistance in cherry tomato fruit with focus on the expression of defense-related genes. Food Chem 254:208–216. doi: 10.1016/j.foodchem.2018.01.100

Lara CA, Santos RO, Cadete RM, Ferreira C, Marques S, Gírio F, Oliveira ES, Rosa CA, Fonseca C (2014) Identification and characterisation of xylanolytic yeasts isolated from decaying wood and sugarcane bagasse in Brazil. Antonie van Leeuwenhoek 105:1107–1119. doi: 10.1007/s10482-014-0172-x

Lazar Z, Liu N, Stephanopoulos G (2018) Holistic approaches in lipid production by *Yarrowia lipolytica*. Trends Biotechnol 36:1157–1170. doi: 10.1016/j.tibtech.2018.06.007

Ledesma-Amaro R, Nicaud J-M (2016) *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. Prog Lipid Res 61:40–50. doi: 10.1016/j.plipres.2015.12.001

Leguina AC del V, Barrios AC, Soro M del MR, Lacosegliaz MJ, Pajot HM, de Figueroa LIC, Nieto-Peñalver CG (2019) Copper alters the physiology of tomato rhizospheric isolates of *Papiliotrema laurentii*. Sci Hortic (Amsterdam). 243:376–384. doi: 10.1016/j.scienta.2018.08.057

Liu HH, Ji XJ, Huang H (2015a) Biotechnological applications of *Yarrowia lipolytica*: Past, present and future. Biotechnol Adv 33:1522–1546. doi: 10.1016/j.biotechadv.2015.07.010

Liu X-Z, Wang Q-M, Göker M, Groenewald M, Kachalkin A V, Lumbsch HT, Millanes AM, Wedin M, Yurkov AM, Boekhout T, Bai F-Y (2015b) Towards an integrated phylogenetic classification of the Tremellomycetes. Stud Mycol 81:85–147. doi: 10.1016/j.simyco.2015.12.001

Loira N, Dulermo T, Nicaud J-M, Sherman DJ (2012) A genome-scale metabolic model of the lipid-accumulating yeast *Yarrowia lipolytica*. BMC Syst Biol 6:35. doi: 10.1186/1752-0509-6-35

Londero MR, Zanrosso CD, Corso LL, Michelin L, Soldera J (2019) Catheter-related infection due to *Papiliotrema laurentii* in an oncologic patient: Case report and systematic review. Brazilian J Infect Dis 23:451–461. doi: 10.1016/j.bjid.2019.10.005

Lopes HJS, Bonturi N, Miranda EA (2021) Induction of resistance mechanisms in *Rhodotorula toruloides* for growth in sugarcane hydrolysate with high inhibitor content. Appl Microbiol Biotechnol 105:9261–9272. doi: 10.1007/s00253-021-11687-z

Meng X-H, Qin G-Z, Tian S-P (2010) Influences of preharvest spraying *Cryptococcus laurentii* combined with postharvest chitosan coating on postharvest diseases and quality of table grapes in storage. LWT - Food Sci Technol 43:596–601. doi: 10.1016/j.lwt.2009.10.007

Middelhoven WJ (2004) The yeast flora of some decaying mushrooms on trunks of living trees. Folia Microbiol (Praha) 49:569–573. doi: 10.1007/BF02931535

Moller L, Kessler KD, Steyn A, Valentine AJ, Botha A (2016) The role of *Cryptococcus laurentii* and mycorrhizal fungi in the nutritional physiology of *Lupinus angustifolius* L. hosting N2-fixing nodules. Plant Soil 409:345–360. doi: 10.1007/s11104-016-2973-3

Mishra P, Lee NR, Lakshmanan M, Kim M, Kim BG, Lee DY (2018) Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica.* BMC Syst Biol 12. doi: 10.1186/s12918-018-0542-5

Morin N, Cescut J, Beopoulos A, Lelandais G, Berre V Le, Uribelarrea J-L, Molina-Jouve C, Nicaud J-M (2011) Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast *Yarrowia lipolytica*. PLoS One 6:1–13. doi: 10.1371/journal.pone.0027966

Mukherjee A, Sarkar S, Gupta S, Banerjee S (2019) DMSO strengthens chitin deacetylase-chitin interaction: Physicochemical, kinetic, structural and catalytic insights. Carbohydr Polym 223:115032. doi: 10.1016/j.carbpol.2019.115032

Mukherjee A, Sarkar S, Parvin R, Bera D, Roy U, Gachhui R (2020) Remarkably high Pb^{2+} binding capacity of a novel, regenerable bioremediator *Papiliotrema laurentii* RY1: Functional in both alkaline and neutral environments. Ecotoxicol Environ Saf 195. doi: 10.1016/j.ecoenv.2020.110439

Pan P, Hua Q (2012) Reconstruction and In Silico analysis of metabolic network for an oleaginous yeast, *Yarrowia lipolytica*. PLoS One 7:1–11. doi: 10.1371/journal.pone.0051535

Patel A, Arora N, Mehtani J, Pruthi V, Pruthi PA (2017) Assessment of fuel properties on the basis of fatty acid profiles of oleaginous yeast for potential biodiesel production. Renew Sustain Energy Rev 77:604–616. doi: 10.1016/j.rser.2017.04.016

Patel A, Arora N, Sartaj K, Pruthi V, Pruthi PA (2016) Sustainable biodiesel production from oleaginous yeasts utilizing hydrolysates of various non-edible lignocellulosic biomasses. Renew Sustain Energy Rev 62:836–855. doi: 10.1016/j.rser.2016.05.014

Pavlova K, Gargova S, Hristozova T, Tankova Z (2008) Phytase from Antarctic yeast strain *Cryptococcus laurentii* AL27. Folia Microbiol (Praha) 53:29–34. doi: 10.1007/s12223-008-0004-3

Polburee P, Yongmanitchai W, Lertwattanasakul N, Ohashi T, Fujiyama K, Limtong S (2015) Characterization of oleaginous yeasts accumulating high levels of lipid when cultivated in glycerol and their potential for lipid production from biodiesel-derived crude glycerol. Fungal Biol 119:1194–1204. doi: 10.1016/j.funbio.2015.09.002

Pomraning KR, Kim Y-M, Nicora CD, Chu RK, Bredeweg EL, Purvine SO, Hu D, Metz TO, Baker SE (2016) Multi-omics analysis reveals regulators of the response to nitrogen limitation in *Yarrowia lipolytica*. BMC Genomics 17:138. doi: 10.1186/s12864-016-2471-2

Rekena A (2021) Enzyme-constrained genome-scale metabolic model of *Rhodotorula toruloides*. Dissertation. University of Tartu

Rodríguez-Couto S (2019) Fungal Laccase: A versatile enzyme for biotechnological applications. In: Yadav AN, Mishra S, Singh S, Gupta A (eds) Recent advancement in white biotechnology through Fungi, Fungal Biology, Springer, Cham, pp. 429–457. doi: 10.1007/978-3-030-10480-1_13

Sampaio JP, Weiß M, Gadanho M, Bauer R (2002) New taxa in the Tremellales: *Bulleribasidium oberjochense* gen. et sp. nov., *Papiliotrema bandonii* gen. et sp. nov. and *Fibulobasidium murrhardtense* sp. nov. Mycologia 94:873–887. doi: 10.1080/15572536.2003.11833182

Sampedro I, Aranda E, Scervino JM, Fracchia S, García-Romera I, Ocampo JA, Godeas A (2004) Improvement by soil yeasts of arbuscular mycorrhizal symbiosis of soybean (*Glycine max*) colonized by *Glomus mosseae*. Mycorrhiza 14:229–234. doi: 10.1007/s00572-003-0285-y

Sarkar S, Chakravorty S, Mukherjee A, Bhattacharya D, Bhattacharya S, Gachhui R (2018) De novo RNA-Seq based transcriptome analysis of *Papiliotrema laurentii* strain RY1 under nitrogen starvation. Gene 645:146–156. doi: 10.1016/j.gene.2017.12.014

Sarkar S, Mukherjee A, Das S, Ghosh B, Chaudhuri S, Bhattacharya D, Sarbajna A, Gachhui R (2019a) Nitrogen deprivation elicits dimorphism, capsule biosynthesis and autophagy in *Papiliotrema laurentii* strain RY1. Micron 124:102708. doi: 10.1016/j.micron.2019.102708

Sarkar S, Mukherjee A, Parvin R, Das S, Roy U, Ghosh S, Chaudhuri P, Roychowdhury T, Mukherjee J, Bhattacharya S, Gachhui R (2019b) Removal of Pb (II), As (III), and Cr (VI) by nitrogen-starved *Papiliotrema laurentii* strain RY1. J Basic Microbiol 59:1016–1030. doi: 10.1002/jobm.201900222

Sharma RR, Singh D, Singh R (2009) Biological control of postharvest diseases of fruits and vegetables by microbial antagonists: A review. Biol Control 50:205–221. doi: 10.1016/j.biocontrol.2009.05.001

Satoh K, Alshahni MM, Umeda Y, Komori A, Tamura T, Nishiyama Y, Yamazaki T, Makimura K (2021) Seven years of progress in determining fungal diversity and characterization of fungi isolated from the Japanese Experiment Module KIBO, International Space Station. Microbiol Immunol 65:463–471. doi: 10.1111/1348-0421.12931

Singh R, Kumar M, Mittal A, Mehta PK (2016) Microbial enzymes: Industrial progress in 21st century. 3 Biotech 6:174. doi: 10.1007/s13205-016-0485-8

Sitepu I, Selby T, Lin T, Zhu S, Boundy-Mills K (2014) Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. J Ind Microbiol Biotechnol 41:1061–1070. doi: 10.1007/s10295-014-1447-y

Sitepu IR, Sestric R, Ignatia L, Levin D, German JB, Gillies LA, Almada LAG, Boundy-Mills KL (2013) Manipulation of culture conditions alters lipid content and fatty acid profiles of a wide variety of known and new oleaginous yeast species. Bioresour Technol 144:360–369. doi: 10.1016/j.biortech.2013.06.047

Skinner CE (1950) Generic name for imperfect yeasts, *Cryptococcus* or *Torulopsis*. Am Midl Nat 43:242–250. doi: 10.2307/2421894

Smirnou D, Krčmář M, Kulhánek J, Hermannová M, Bobková L, Franke L, Pepeliaev S, Velebný V (2015) Characterization of hyaluronan-degrading enzymes from yeasts. Appl Biochem Biotechnol 177:700–712. doi: 10.1007/s12010-015-1774-0

Sreeharsha RV, Mohan SV (2020) Obscure yet promising oleaginous yeasts for fuel and chemical production. Trends Biotechnol 38:873–887. doi: 10.1016/j.tibtech.2020.02.004

Staden J Van, Haan R Den, Zyl WH Van, Botha A, Viljoen-bloom M (2007) Phytase activity in *Cryptococcus laurentii* ABO 510. FEMS Yeast Res 7:442–448. doi: 10.1111/j.1567-1364.2006.00196.x

Szczepańska P, Hapeta P, Lazar Z (2022) Advances in production of high-value lipids by oleaginous yeasts. Crit Rev Biotechnol 42:1–22. doi: 10.1080/07388551.2021.1922353

Tang Y, Gao X-D, Wang Y, Yuan B-F, Feng Y-Q (2012) Widespread existence of cytosine methylation in yeast DNA measured by gas chromatography/mass spectrometry. Anal Chem 84:7249–7255. doi: 10.1021/ac301727c

Tian S, Qin G, Xu Y (2004) Survival of antagonistic yeasts under field conditions and their biocontrol ability against postharvest diseases of sweet cherry. Postharvest Biol Technol 33:327–331. doi: 10.1016/j.postharvbio.2004.03.010

Tiukova IA, Prigent S, Nielsen J, Sandgren M, Kerkhoven EJ (2019) Genome-scale model of *Rhodotorula toruloides* metabolism. Biotechnol Bioeng 116:3396–3408. doi: 10.1101/528489

Tsigie YA, Wang CY, Truong CT, Ju YH (2011) Lipid production from *Yarrowia lipolytica* Po1g grown in sugarcane bagasse hydrolysate. Bioresour Technol 102:9216–9222. doi: 10.1016/j.biortech.2011.06.047

Vadkertiová R, Dudášŏvá H, Balaščáková M (2017) Yeasts in agricultural and managed soils. In: Buzzini P, Lachance MA, Yurkov A (eds) Yeasts in natural ecosystems: Diversity, Springer, Cham, pp. 117–144. doi: 10.1007/978-3-319-62683-3_4

Vasconcelos B, Teixeira JC, Dragone G, Teixeira JA (2019) Oleaginous yeasts for sustainable lipid production - from biodiesel to surf boards, a wide range of “green” applications. Appl Microbiol Biotechnol 103:3651–3667. doi: 10.1007/s00253-019-09742-x

Ventorim RZ, Ferreira MAM, de Almeida ELM, Kerkhoven EJ, da Silveira WB (2022) Genome-scale metabolic model of oleaginous yeast *Papiliotrema laurentii*. Biochem Eng J 180:108353. doi: 10.1016/j.bej.2022.108353

Vieira NM, dos Santos RCV, Germano VKC, Ventorim RZ, de Almeida ELM, da Silveira FA, Ribeiro Júnior JI, da Silveira WB (2020a). Isolation of a new *Papiliotrema laurentii* strain that

displays capacity to achieve high lipid content from xylose. 3 Biotech 10:1–14. doi: 10.1007/s13205-020-02373-4

Vieira NM, Ventorim RZ, Ferreira MAM, Souza GB, de Almeida ELM, Vidigal PMP, Nesi AN, Fietto LG, da Silveira WB (2020b) Insights into oleaginous phenotype of the yeast *Papiliotrema laurentii*. Fungal Genet Biol 144:103456. doi: 10.1016/j.fgb.2020.103456

Wang G, Liu L, Liang W (2018) Single cell oil production from hydrolysates of inulin by a newly isolated yeast *Papiliotrema laurentii* AM113 for biodiesel making. Appl Biochem Biotechnol 184:168–181. doi: 10.1007/s12010-017-2538-9

Wasylenko TM, Ahn WS, Stephanopoulos G (2015) The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in *Yarrowia lipolytica*. Metab Eng 30:27–39. doi: 10.1016/j.ymben.2015.02.007

Wei S, Jian X, Chen J, Zhang C, Hua Q (2017) Reconstruction of genome-scale metabolic model of *Yarrowia lipolytica* and its application in overproduction of triacylglycerol. Bioresour Bioprocess 4. doi: 10.1186/s40643-017-0180-6

Wei Y, Mao S, Tu K (2014) Effect of preharvest spraying *Cryptococcus laurentii* on postharvest decay and quality of strawberry. Biol Control 73:68–74. doi: 10.1016/j.biocontrol.2014.02.016

Wen Z, Zhang S, Odoh CK, Jin M, Zhao ZK (2020) *Rhodosporidium toruloides* - A potential red yeast chassis for lipids and beyond. FEMS Yeast Res 20:1–12. doi: 10.1093/femsyr/foaa038

Yalçın HT, Fındık B, Terzi Y, Uyar E, Shatila F (2021) Isolation and molecular identification of industrially important enzyme producer yeasts from tree barks and fruits. Arch Microbiol 203:1079–1088. doi: 10.1007/s00203-020-02104-6

Zhu Z, Zhang S, Liu H, Shen H, Lin X, Yang F (2012) A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. Nat Commun 3:1112. doi: 10.1038/ncomms2112

CHAPTER 4 - INSIGHTS INTO THE RESPONSE AND TOLERANCE MECHANISMS OF *Papiliotrema laurentii* TO ACETIC ACID STRESS BY RNA-seq AND GENOME-SCALE METABOLIC MODELING ANALYSIS

4.1. Abstract

The production of lipids and fatty acids by oleaginous yeasts from lignocellulosic biomasses is a sustainable alternative to produce oleochemicals. The pretreatment of these biomasses releases various toxic compounds, including acetic acid, that impair yeast growth and reduce productivity. Therefore, selecting and developing robust yeast strains tolerant to acetic acid is imperative. *Papiliotrema laurentii* UFV-1 strain, isolated by our research team, can assimilate sugars derived from agricultural wastes and convert them into high lipid contents; however, wild strains are highly sensitive to acetic acid. Previously, we have selected an acetic acid-tolerant strain of *P. laurentii*, ATS, by Adaptive Laboratory Evolution and identified mutations that might contribute to the tolerance phenotype. Here, we combined transcriptome and genome-scale metabolic modeling to deepen our understanding regarding the targets of acetic acid stress, as well as the adaptive responses in *P. laurentii*. Acetic acid stress promoted global expression changes in both strains and most repressed genes were related to transcriptional and translational processes and ribosome biogenesis. In the presence of acetic acid, the sensitive strain induced DNA mismatch repair mechanisms and meiosis, while the tolerant strain negatively regulated autophagy and the cell cycle. Moreover, the tolerant strain induced processes responsible for increasing the intracellular pH (e.g., arginase, ornithine metabolism, urea cycle), detoxification of toxic compounds (e.g., glutathione metabolism), and proton efflux. Notably, ATS also presented a remarkable NAD(P)H pool in the metabolic modeling analysis, which might support the reducing power required by these tolerance mechanisms. On the other hand, the Parental strain induced genes related to cell wall biogenesis, consistent with its morphological changes, and cobalamin synthesis. Overall, the genes and pathways described herein as tolerant-related might be used in future metabolic engineering strategies to improve the tolerance of *P. laurentii* to weak acids, boosting its application in lignocellulosic-based biorefineries.

Keywords: Oleaginous yeast; Transcriptomics; Stress response; Weak-acids.

4.2. Introduction

From a circular economy standpoint, efficient bioprocessing should combine three crucial aspects: the substrate, organism, and product (NURWONO et al., 2023). Since the current environmental crisis is caused at least in part by the use of oil derivatives as fuels and chemicals, the use of sustainable oil sources is pivotal to mitigate the impact provoked by petrochemicals. In this sense, lipids and fatty acids from oleaginous yeasts produced from abundant feedstocks such as lignocellulosic hydrolysates might be a sustainable alternative source for oleochemicals (GALLEGO-GARCÍA et al., 2023). During the pretreatment of lignocellulosic biomasses, which is required to separate hemicellulose and cellulose and facilitate their hydrolysis, various toxic compounds that can impair yeast growth and reduce productivity, are released, including furfural, hydroxymethylfurfural (HMF), phenolic compounds, formic acid, and acetic acid (JÖNSSON; MARTÍN, 2016). Therefore, selecting and developing robust yeast strains that can metabolize the main sugars found in lignocellulosic hydrolysates (i.e., xylose, arabinose, glucose, galactose), as well as maintain high product yields and tolerate these inhibitors, is required to broaden their applicability and viability.

Given the potential of non-*Saccharomyces* yeasts in lignocellulosic biorefineries, our research team has isolated the *Papiliotrema laurentii* UFV-1 strain, which is capable of assimilating sugars derived from agricultural wastes, such as glucose and xylose from lignocellulosic biomasses, and converting them into high lipid contents (VIEIRA et al., 2020). However, wild *P. laurentii* strains are highly sensitive to acetic acid, one of the primary inhibitors in lignocellulosic hydrolysates (ALMEIDA et al., 2023; SITEPU et al., 2014). Even after the detoxification of lignocellulosic hydrolysates, acetic acid remains in inhibitory concentrations (CHANDEL; DA SILVA; SINGH, 2013). In pH under 4.76, acetic acid is mainly in its undissociated form and freely diffuses through the plasma membrane or is transported by facilitated diffusion via aquaporins (MOLLAPOUR; PIPER, 2007). In the cytoplasm, it dissociates, releasing acetate and protons. The release of protons acidifies the cytoplasm, reducing metabolic rates, biosynthetic processes, and enzyme activity and promoting protein misfolding. Acetate accumulation promotes oxidative stress, protein aggregation, and lipid oxidation, affecting different cell structures and organelles. Hence, to thrive in the presence of acetic acid, yeast cells need to employ complex regulatory and response mechanisms comprising a multifactorial process (GUARAGNELLA; BETTIGA, 2021; LI et al., 2024).

Since the response and tolerance to acetic acid requires the coordinate regulation of multiple genes and functions and the molecular mechanisms underlying this stress in *P. laurentii* have not been elucidated so far, we have previously applied Adaptive Laboratory Evolution (ALE) - a systemic metabolic engineering approach with great potential to develop robust yeast strains (LONG; ANTONIEWICZ, 2018; SANDBERG et al., 2019) – to select acetic acid-tolerant strains of *P. laurentii* (ALMEIDA et al., 2023). One of them, here referred to as acetic acid tolerant strain (ATS), displayed tolerance (up to 2.0 g/L of acetic acid) and preserved the oleaginous phenotype in all conditions tested. Besides the physiological characterizations, we sequenced and compared its genome with the wild strain (the Parental strain of ATS) and identified mutations that might contribute to the tolerance phenotype.

Furthermore, the understanding of multifactorial stress responses can be facilitated by the application of holistic approaches from a systems biology perspective, including biological networks, mathematical models, and high-throughput (omics) data (NIELSEN, 2017). Hence, we combined transcriptome and genome-scale metabolic modeling in the current study to deepen our understanding of the acetic acid stress response and tolerance of *P. laurentii.* The genes and pathways described here as tolerant-related might be used in future genetic and metabolic engineering rounds to improve the tolerance of this yeast to weak acids and boost its application in lignocellulosic biorefineries.

4.3. Materials and Methods

4.3.1. Study design and sample collection

As mentioned in the introduction, we have previously used ALE to select acetic acid-tolerant strains of *P. laurentii* UFV-1 (ALMEIDA et al., 2023). One strain, ATS, presented tolerance and preserved the oleaginous phenotype in all conditions tested. In this previous work, we cultivated ATS and its Parental strain under chemostats with xylose (1.5 g/L) and YNB (Yeast Nitrogen Base – 6.7 g/L) medium without amino acids (Sigma Chemical CO., St. Louis, MO, USA) in the absence and presence (0.8 g/L) of acetic acid at a dilution rate of 0.055 h^{-1}. Cultivations were conducted in a 1.3-liter bioreactor (BioFlo/CelliGen 115, Eppendorf, Germany) with a working volume of 1.0 L, air flow of 1 vvm, dissolved oxygen above 30%, and pH below 4.0 at 30 °C. The steady state was considered as achieved after five residence times with constant dissolved oxygen, pH, and optical density at 600 nm. For more details on the cultivation strategy and parameters, see ALMEIDA et al. (2023). Three samples from each

chemostat culture were collected and immediately frozen with liquid nitrogen and stored at -70 °C. In the present study, we carried out RNA extraction for transcriptomic analysis. For genome-scale metabolic modeling tuning, we also retrieved the biomass composition and xylose and acetic acid uptake data for each condition and strain from the previous study (ALMEIDA et al., 2023).

4.3.2. RNA extraction and sequencing

For total RNA extraction, we ground three samples (n = 3) from each cultivation with liquid nitrogen using pistils for microtubes. Then, we extracted the total RNA using the TRI Reagent® (Sigma-Aldrich) following the manufacturer's instructions. We assessed RNA quality and quantity using NanoDrop (Thermo Fisher Scientific Inc.) and integrity by 1% agarose gel electrophoresis. Next, we sent total RNA samples to GenOne (Rio de Janeiro, Brazil) for library preparation and sequencing using the Illumina HiSeq 2500 platform with the 2 x 150 bp paired-end method. The sequencing data are available in Genbank (BioProject: PRJNA879120).

4.3.3. Data processing and differential expression analysis

First, we evaluated the quality of reads using FastQC v. 0.11.9 (ANDREWS, 2010). Then, we removed adaptor sequences and low-quality reads from the raw data using Trimmomatic (BOLGER; LOHSE; USADEL, 2014). We mapped the resulting clean reads to the reference genome of *P. laurentii* (JGI Project id: 1167266) using Bowtie 2 (LANGMEAD; SALZBERG, 2012), determined read counts using featureCounts software (LIAO; SMYTH; SHI, 2014), and normalized them using DEseq2 (LOVE; HUBER; ANDERS, 2014). We analyzed the distribution of samples by Principal Component Analysis (PCA) based on normalized reads. Further, we assessed the differential gene expression between conditions (absence and presence of acetic acid) for each strain (Parental and ATS) and between strains in the same condition using DEseq2. We visualized the overall results using volcano plots and the intersections between conditions and strains using Venn diagrams. We conducted the PCA and constructed the Venn diagrams in Orange Data Mining (v. 3.34.0). We built the volcano plots using the "EnhancedVolcano" package (BLIGHE; RANA; LEWIS, 2023) based on normalized DEseq2 reads in R 4.2.2.

In DEseq2, the *p*-values determined by the Wald test are corrected for multiple testing using the Benjamini and Hochberg method [false discovery rate (FDR) values]. Thus, we considered an FDR cutoff < 0.01 to determine differentially expressed genes (DEGs). We considered DEGs with $\log_2$ [fold change (FC)] > 1 as upregulated and $\log_2$(FC) < -1 as downregulated. We also analyzed DEGs regarding the enrichment of Gene Ontology (GO) terms and pathways [Kyoto Encyclopedia of Genes and Genomes (KEGG)] using the "ClusterProfiler" R package (WU et al., 2021). The cutoff for enriched terms was set as $p < 0.05$ and adjusted-$q < 0.2$ (Bonferroni-Holm correction).

4.3.4. Protein-protein interaction (PPI) networks

Furthermore, we constructed PPI networks based on DEGs for each condition and strain using STRING 12.0 (SZKLARCZYK et al., 2023). First, we uploaded the reference protein annotaiton of *P. laurentii* to the STRING website (the whole-proteome network can be found at https://version-12-0.string-db.org/organism/STRG0A79TGQ), then built networks for each group of DEGs. We considered only networks with significant PPI enrichments ($p < 0.1$). We prioritized networks with only high confidence interaction scores (> 0.7). However, for the upregulated DEGs of ATS vs. Parental strain in the presence and absence of acetic acid and the unique downregulated DEGs for ATS in the presence of acetic acid, we could not obtain a network with only high confidence scores; thus, in these cases, we also considered medium confidence scores (> 0.4). Even considering medium confidence scores, we could not build a significant network for the upregulated DEGs of ATS in the presence vs. absence of acetic acid, including unique ATS upregulated DEGs, as well as for the intersection between ATS and Parental upregulated DEGs in the presence and absence of the acid.

For each significant network, we also analyzed the enrichment of nodes (proteins) regarding GO terms, local STRING network clusters, KEGG pathways, Reactome pathways, and UniProt keywords using default STRING cutoffs (i.e. FDR < 0.05; Hypergeometric test; Benjamini and Hochberg correction). The nodes in the network represent proteins, and the edges represent the confidence score; the thickness of the edge indicates the strength of data support, with higher thickness indicating stronger support (interactions).

4.3.5. Reaction Activity Score (RAS)

Moreover, to determine how gene expression levels could affect metabolic activity, we applied the RAS. First, we retrieved the GEM of *P. laurentii* (*papla*-GEM; https://github.com/SysBioChalmers/papla-GEM) (VENTORIM et al., 2022) and mapped the model genes to the reference proteome. Next, we calculated the RAS for each reaction as described by Di Filippo et al. (2022). The RAS metric considers that the isoforms of an enzyme – represented in the model by Gene-Protein-Reaction (GPR) rules with the OR operator – additively contribute to the activity of a certain reaction (*r*), which is given by:

$$RAS_r = \sum Exp_i : i \in G_r$$

where Exp_i represents the expression of the i^{th} gene of a given reaction *r* in the model, and G_r is the set of genes in the GPR of reaction *r*.

In contrast, for enzymatic complexes – represented in the model by GPR rules with the AND operator – the subunit with the lowest expression determines the RAS. This can be represented as:

$$RAS_r = \min\{Exp_i : i \in G_r\}$$

After determining the RAS for each reaction, we determined their $\log_2$(FC) between conditions (absence and presence of acetic acid) and strains (Parental and ATS) in the same conditions. To determine the statistical significance, we conducted paired t-tests and calculated the adjusted *p-value* using the Benjamini and Hochberg correction; we considered a *p*-adjusted < 0.05 as statistically significant. We considered reactions with RAS with $\log_2$(FC) > 1.2 as induced and $\log_2$(FC) < -1.2 as repressed.

4.3.6. Reconstruction of the enzyme-constrained GEM (ecGEM) of P. laurentii

Subsequently, we reconstructed the ecGEM of *P. laurentii* (*ecpapla*-GEM) using the GECKO toolbox (v. 3.1.2) (https://github.com/SysBioChalmers/GECKO) (CHEN et al., 2024). In this approach, the enzyme turnover number (k_{cat}) of a given enzyme and its molecular weight (*MW*) are incorporated in the stoichiometric matrix of the model as the coefficient of the pseudometabolite that represents the protein abundance added to the reaction associated with the enzyme GPR, preserving the mass conservation. For example, consider the following reaction with reactants *A* and *B* and products *C* and *D* catalyzed by the hypothetical *Protein_1*. In the conventional model, the reaction can be represented as:

$$A + B \rightarrow C + D$$

Meanwhile, in the ecGEM, the reaction can now be represented as:

$$A + B + \frac{MW}{k_{cat}} Prot\_1_{abund} \rightarrow C + D$$

$$Prot\_1_{abund} \rightarrow P_{pool}$$

where $Prot\_1_{abund}$ represents the abundance of *Prot_1* and the P_{pool} is the total protein pool available for metabolic reactions in a determined condition.

For reactions with GPR rules with the OR operator, GECKO creates novel reactions for each isoform. For the AND operator, GECKO creates a series of dependent reactions for each subunit of the model, where the reactants of the first reaction represent the original reactants in the conventional model, the products of the last reaction represent the original products and pseudoreactants/products are created to fulfill the intermediate reactions, also preserving the mass conservation statement. In the case of reversible reactions, GECKO breaks them into two irreversible reactions since the k_{cat} is different for each direction. Pseudoreactions are also created to supply the required enzyme abundances (pseudometabolites), as well as a protein pool reaction that represents the enzymatic protein fraction of the cell. The upper bound of this pool (ub_{Ppool}) is represented by:

$$ub_{Ppool} = P_{total} * \sigma * f * 1000$$

where P_{total} represents the total protein content measured for each condition in the cell (g_{prot}/g DW); σ represents an average saturation factor (0.5); and f accounts for the estimated fraction of the total proteome of the cell that was not measured (0.5).

GECKO 3 has different strategies to incorporate k_{cat} values into the model. In the first strategy, the values are retrieved from the BRENDA database, considering data for the organism of interest; then, if not present, for the phylogenetically closest organism with available data; next, it tries to match the E.C. number for the organism; if still no match is found, it tries to match a k_{cat} for the E.C. number and any substrate for the phylogenetically closest organism; finally, it introduces one wild card to the E.C. number and repeats the steps above. In the second strategy, k_{cat} values are predicted by the deep-learn model DLKcat, which requires substrate structures and protein sequences. The third approach combines the values retrieved from BRENDA and those predicted by DLKcat to construct a "merged" model. Since GECKO also uses KEGG and UniProt databases data and *P. laurentii* is not present in KEGG and has

scarce data in UniProt, we matched the GPR rules in *papla*-GEM to its orthologs with *Cryptococcus gattii* (identified by bidirectional best hit BLASTp), another Basidiomycota yeast with enough data for ecGEM reconstruction in these databases.

After incorporating the k_{cat} values, GECKO performs a model tunning to allow growth in the desired conditions. Herein, the model that required the lowest number of adjustments was the one containing only k_{cat} values from BRENDA; thus, we used this model in the subsequent simulations.

4.3.7. Model adjustments and parsimonious flux balance analysis (pFBA)

We conducted model adjustments and simulations using the RAVEN toolbox 2.5.3 (WANG et al., 2018) and COBRA toolbox 3.0 (HEIRENDT et al., 2019) in the MATLAB environment with Gurobi® v. 9.5.2 as the solver. We adjusted the biomass equation of the model - total protein, total lipid, fatty acid profile, total carbohydrate, glycogen, glucan, mannan, and trehalose fractions - using data for each condition (absence and presence of acetic acid) and strain (Parental and ATS) provided by ALMEIDA et al. (2023). We also constrained the uptake rates of carbon sources (xylose and acetic acid) and biomass formation [dilution rate (D) = 0.055 h^{-1}] using data from the chemostat cultivations (ALMEIDA et al. 2023). Overall, we generated four condition/strain specific models. Then, we applied pFBA (LEWIS et al., 2010) to minimize the flux vectors for GPR-associated reactions and fine tune the exchange rates (i.e. CO_2 and O_2) that were not measured during the cultivations.

4.3.8. Random sampling (RS) and flux variability analysis (FVA)

To compare the different availability of cofactors in the absence and presence of acetic acid, as well as between strains, we conducted an RS analysis using the condition/strain-adjusted *ecpapla*-GEM and the constraints described above. We conducted 10,000 random samplings considering a variability of 5% in the growth rate and 10% in the protein pool, O_2, and CO_2 exchanges. Then, we mapped the solution to the conventional GEM and calculated the ATP, NADH, and NADPH yield as the sum of median fluxes in the reactions that consumed or produced these cofactors divided by the total carbon uptake rate.

Besides the RS, we applied FVA to evaluate how the flux space was affected by the presence of acetic acid and between different strains. In FVA, the maximum and minimum possible flux through a given reaction is determined, representing the range of metabolic capabilities. Here, we calculated the log_2FC of ranges between conditions for the same strain and between strains in the same condition. We considered a reaction with log_2FC of ranges > 1.2 as relaxed and < -1.2 as retrained. Similar to the RS analysis, we conducted the FVA using the condition/strain-adjusted *ecpapla*-GEM with similar constraints, then mapped the results to the conventional model. Since the CO_2, O_2, xylose, and acetic acid exchange ranges were pre-determined, we did not include them in the FC analysis. We also analyzed the reactions with differential ranges regarding the enrichment of metabolic subsystems using the "ClusterProfiler" R package. The cutoff for enriched subsystems was set as $p < 0.05$ and adjusted $q < 0.2$ (Bonferroni-Holm correction).

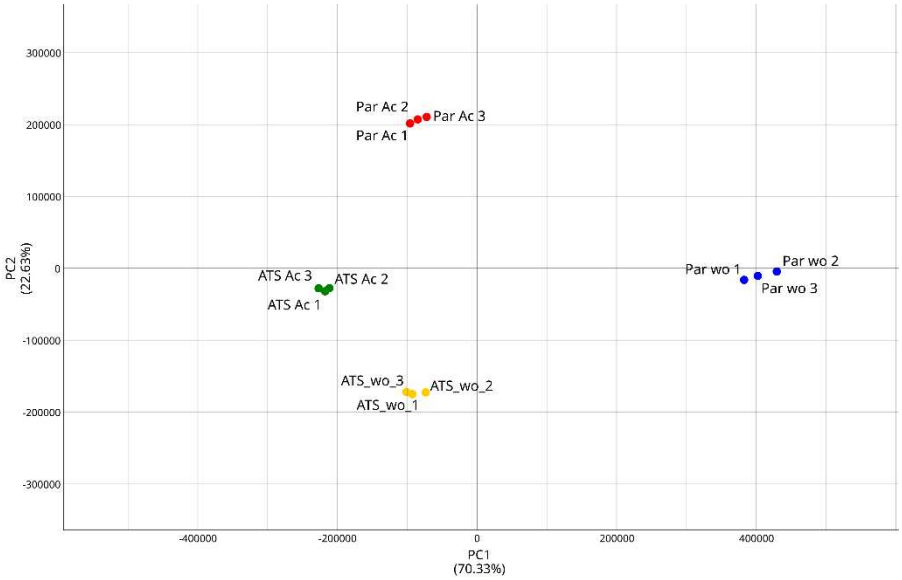
4.4. Results

4.4.1. Global gene expression changes

Previously, we selected a strain of *P. laurentii* UFV-1 tolerant to acetic acid, ATS, by ALE and compared its genomic mutations, as well as physiological behavior, to its Parental strain (ALMEIDA et al., 2023). Herein, to gain new insights and better understand the molecular mechanisms related to the responses of the tolerant (ATS) and sensitive (Parental) strains to acetic acid stress, we conducted transcriptomics and genome-scale metabolic modeling analyses using samples from xylose chemostats in the presence and absence of acetic acid.

After RNA extraction, sequencing, pre-processing, mapping, counting, and data normalization, we conducted PCA and differential gene expression analysis. Samples from the same treatment were grouped in the PCA, demonstrating the reliability and consistency of the sampling and data analysis (Figure 1). The PCA also indicated that, under stress, the global mRNA profile of the Parental strain changed to a higher degree compared to that of ATS. Notably, the evolved strain, ATS, also presented a great difference in the expression profile compared to the Parental strain in the absence of acetic acid. This result indicates that the ALE promoted constitutive alterations in the global mRNA pool of ATS that can be detected even in the absence of selective pressure.

Figure 1 – Principal component analysis (PCA) for normalized (DESeq2) RNA-seq counts. ATS: Acetic acid tolerant strain; Par: Parental strain; Par wo: *P. laurentii* Parental strain without acetic acid stress; Par Ac: *P. laurentii* Parental strain without acetic acid stress; ATS wo: *P. laurentii* ATS without acetic acid stress; ATS Ac: *P. laurentii* Parental strain without acetic acid stress.



The differential expression analysis further corroborated this global expression difference. In the presence of acetic acid, the Parental strain presented 3.12 times more DEGs than ATS, being most of them downregulated. Moreover, ATS upregulated only 56 genes in the presence of acetic acid, while the Parental strain upregulated 283, that is, 5.05 times more genes than the evolved strain (Figures 2; Table 1). This lower expression profile change of ATS when stressed is indicative of its more adapted cellular condition, requiring a less harsh global expression change than the Parental strain.

Figure 2 – **Volcano plots**: **(A)** *P. laurentii* ATS under acetic acid stress vs. without; **(B)** *P. laurentii* Parental strain under acetic acid stress vs. without; **(C)** *P. laurentii* ATS vs. Parental strain under acetic acid stress; **(D)** *P. laurentii* ATS vs. Parental strain without acetic acid stress; Grey dots represent non-significant genes; Blue dots represent genes with *p*-adjusted < 0.05 but $|Log_2$ [fold change (FC)]| < 1; Red dots represent genes with *p*-adjusted < 0.05 but $|Log_2$ (FC)| > 1. ATS: Acetic acid tolerant strain; Par: Parental strain; Par w/o: *P. laurentii* Parental strain without acetic acid stress; Par Ac: *P. laurentii* Parental strain without acetic acid stress; ATS w/o: *P. laurentii* ATS without acetic acid stress; ATS Ac: *P. laurentii* Parental strain without acetic acid stress.

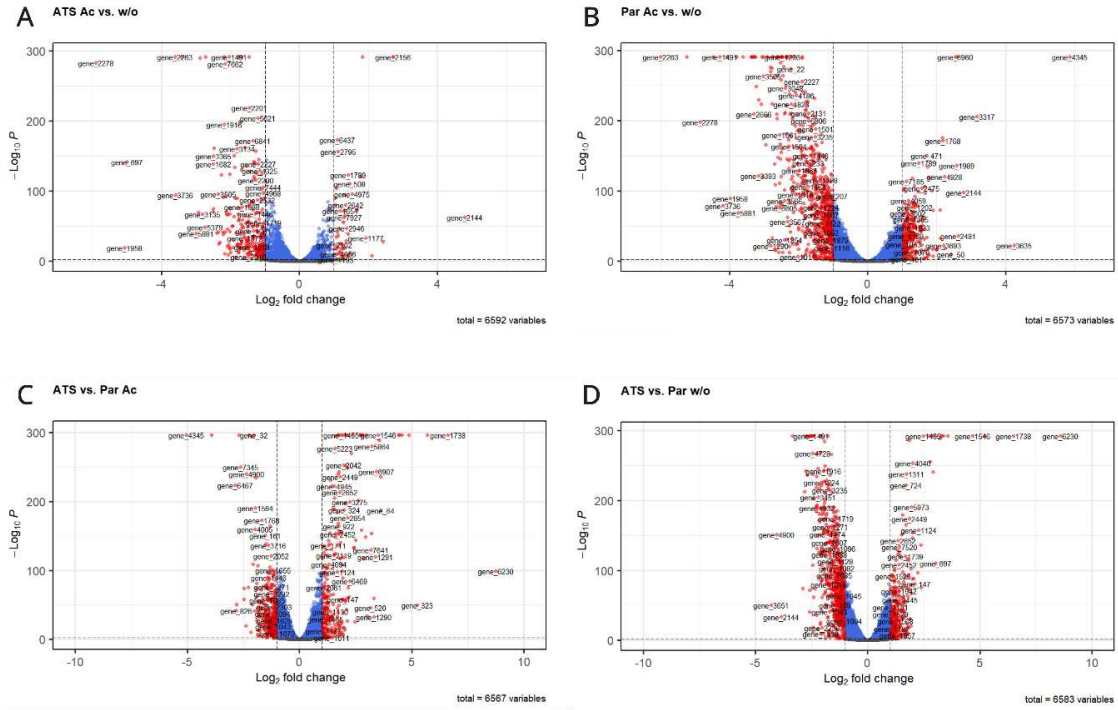


Table 1 – Summary of differentially expressed genes (DEGs) in *Papiliotrema laurentii* ATS and Parental strain in the presence and absence of acetic acid stress.

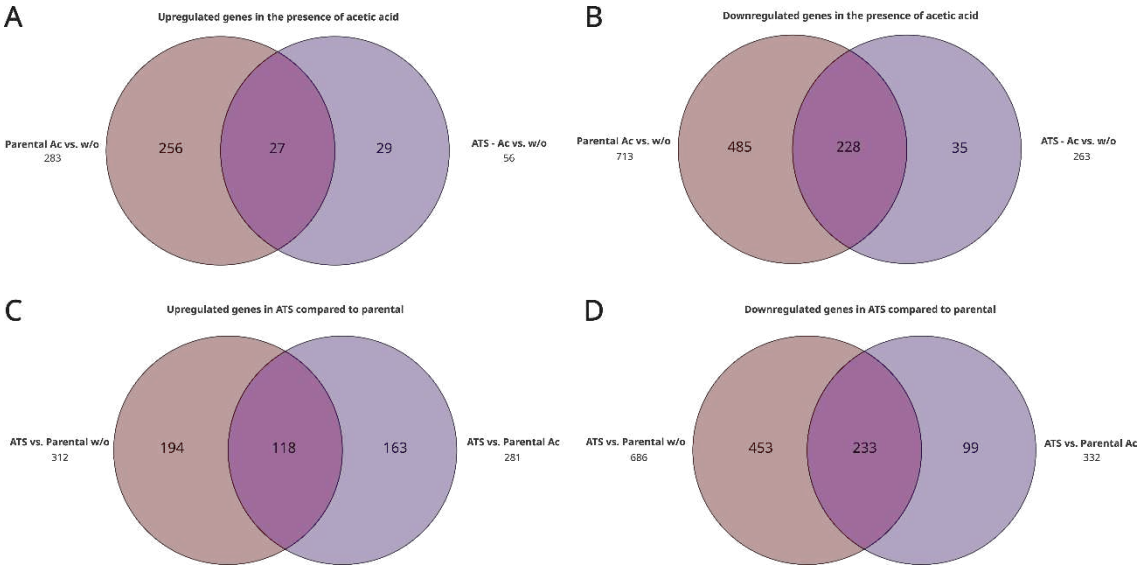
	Differentially expressed genes*		
Condition	**Total**	**Upregulated****	**Downregulated****
Parental (presence vs. absence of acetic acid)	996	283 (28.4%)	713 (71.6%)
ATS (presence vs. absence of acetic acid)	319	56 (17.5%)	263 (82.5%)
ATS vs. Parental in the absence of acetic acid	998	312 (31.3%)	686 (68.7%)
ATS vs. Parental in the presence of acetic acid	613	281 (45.8%)	332 (54.2%)

* *p*-adjusted < 0.05 and $|Log_2$ [fold change (FC)]| > 1. ** Upregulated: Log_2 (FC) > 1; Downregulated: Log_2 (FC) < -1

The expression profile of ATS presented significant differences compared to the Parental strain in both unstressed and stressed conditions (Figures 2C and D). For example, even in the absence of acetic acid, ATS differentially expressed 998 genes compared to the Parental strain (Table 1), indicating that the evolution process promoted major global expression profile changes, even when the selective pressure (acetic acid) is absent. Additionally, in the presence of acetic acid, ATS presented a more balanced up/downregulated profile, suggesting that both gene expression activation and repression might be related to the emergence of a more acetic acid-tolerant phenotype in *P. laurentii*.

Furthermore, ATS shared most of the downregulated genes in the presence of acetic acid with the Parental strain (228 of 263). Importantly, we did not observe this pattern for induced genes, as more than half of ATS-induced genes were not upregulated in the Parental strain (29 of 56) (Figures 3A and B). Although ATS shared many DEGs in the presence and absence of acetic acid versus the Parental strain, many DEGs (42.7 and 64.8% in the presence and absence of acetic acid, respectively) were still condition-specific, also highlighting ATS's great expression changes compared to the Parental strain (Figures 3C and D).

Figure 3 – **Venn diagrams**: **(A)** Upregulated genes in the presence of acetic acid; **(B)**; Downregulated genes in the presence of acetic acid; **(C)** Upregulated genes in ATS compared to the Parental strain; **(D)** Downregulated genes in ATS compared to the Parental strain. ATS: Acetic acid tolerant strain; Par: Parental strain; Par w/o: *P. laurentii* Parental strain without acetic acid stress; Par Ac: *P. laurentii* Parental strain without acetic acid stress; ATS w/o: *P. laurentii* ATS without acetic acid stress; ATS Ac: *P. laurentii* Parental strain without acetic acid stress.



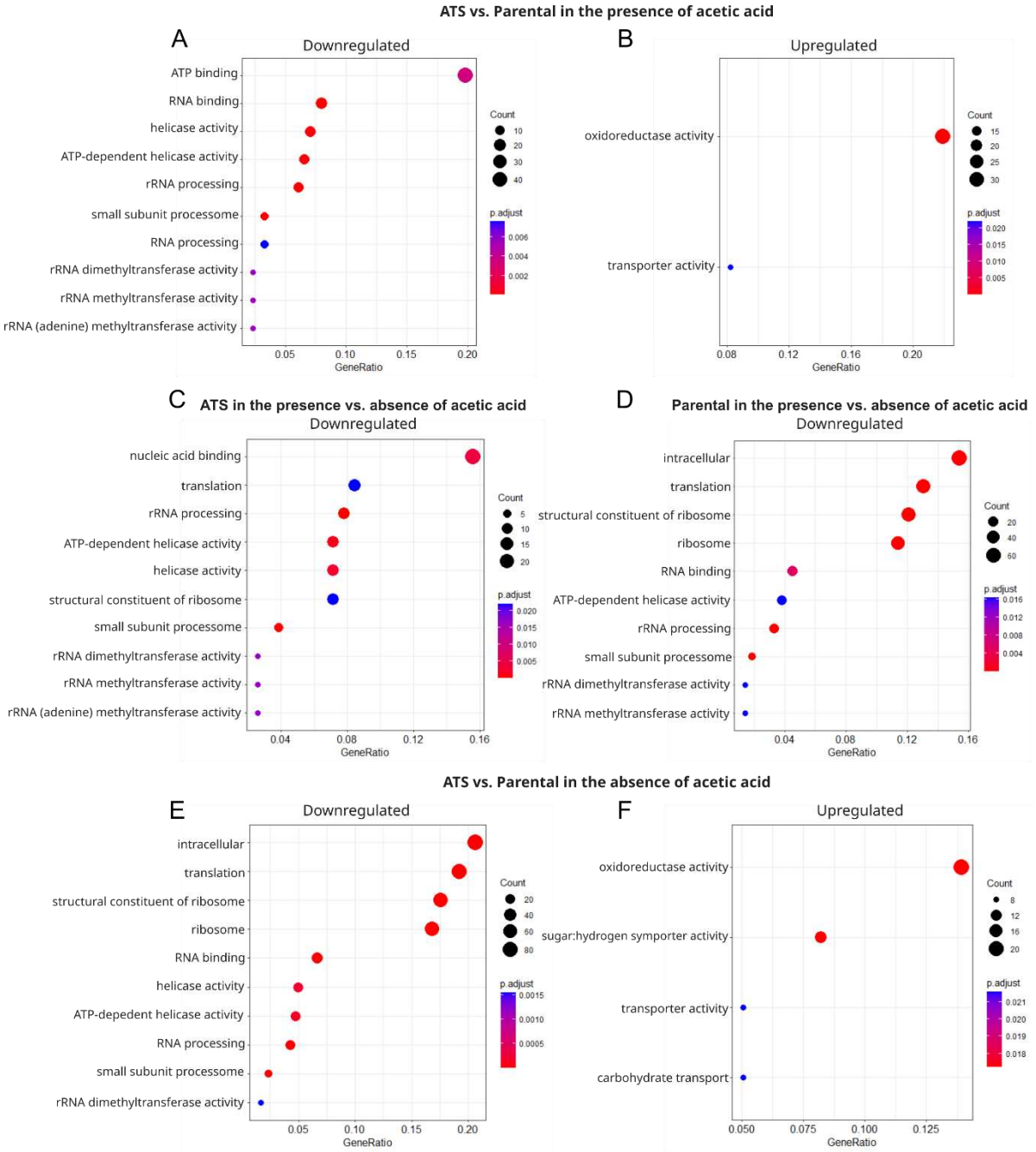
4.4.2. Expression profile changes induced by acetic acid stress

The GO functional enrichment analysis of DEGs indicated that, in the presence of acetic acid, the genes downregulated by ATS compared to the Parental strain were related to DNA processing and transcription (e.g., helicase activity), and translation and rRNA metabolism (e.g., rRNA processing, RNA binding, small unit processome, and di and methyltransferase activity) (Table 2 and Figure 4A). Besides, the top 10 downregulated genes included proteins related to cobalamin (vitamin B12) synthesis, virulence and capsule formation (PriA/CPL1 family), decarboxylation of aromatic acids (UbiD decarboxylase), ubiquitination (ASK1-D3 ubiquitin ligase), spermine, copper and phosphorus transport, and galactose metabolism (UDP-glucose 4-epimerase GalE). Consistent with the GO enrichment results, the Top 10 downregulated genes also included the fold of ribosomal proteins and an ATP-dependent RNA helicase (Table 2).

Table 2 – Top 10 down and upregulated genes for *P. laurentii* ATS vs. Parental strain in the presence of acetic acid.

Gene ID	Protein ID	Log_2(FC)	*p*-adjusted	Function
		Downregulated		
gene_4345	490704	-5.06	0E+00	Putative cobalamin synthesis protein
gene_5483	494674	-3.94	0E+00	PriA/CPL1 family protein
gene_6467	442885	-2.88	2E-224	UbiD decarboxylase
gene_826	478721	-2.83	8E-43	ASK1-D3 ubiquitin ligase
gene_3649	488204	-2.80	9E-51	Ribosomal protein S5 domain 2-type fold
gene_6842	84801	-2.69	0E+00	Spermine transporter
gene_7345	106947	-2.64	7E-251	Putative Pi-transporter A-1
gene_2770	451424	-2.48	8E-75	Putative copper-exporting ATPase
gene_4901	315141	-2.47	6E-59	ATP-dependent RNA helicase DBP3
gene_3647	254469	-2.40	4E-20	UDP-glucose 4-epimerase GalE
		Upregulated		
gene_6230	382679	8.75	5E-100	Alpha/Beta hydrolase (Epoxide hydrolase-like)
gene_1738	508509	6.62	0E+00	Putative CIP1
gene_6158	512494	5.70	0E+00	Sterol receptor domain-containing protein
gene_323	476944	5.25	5E-50	Nitrosoguanidine resistance protein SNG1
gene_433	22001	4.87	0E+00	Short-chain dehydrogenase/reductase SDR family
gene_3597	488059	4.56	0E+00	Wax synthase domain-containing protein
gene_3876	264447	4.46	0E+00	Heme-bound CFEM protein (Csa2)
gene_3016	534783	4.42	0E+00	NAD-dependent oxidoreductase
gene_64	476044	3.62	6E-188	Efflux protein (EncT)
gene_806	507690	3.62	2E-236	SUR7/PalI family-domain-containing protein

Figure 4 – **Dot plots for Gene Ontology (GO) enrichments**: **(A, B)** *P. laurentii* ATS vs. Parental strain in the presence of acetic acid: (A) downregulated and (B) upregulated genes. **(C)** *P. laurentii* ATS in the presence vs. absence of acetic acid (downregulated genes); no GO terms were enriched in the upregulated gene set for this condition. **(D)** *P. laurentii* Parental strain in the presence vs. absence of acetic acid (downregulated genes) no GO terms were enriched in the upregulated gene set for this condition. **(E, F)** *P. laurentii* ATS vs. Parental strain in the absence of acetic acid.



Meanwhile, the genes upregulated by ATS compared to the Parental strain in the presence of acetic acid were related to oxidoreductase and transporter activity (Figure 4B). Consistently, the top 10 upregulated genes in this scenario included proteins related to exposure to toxic compounds (epoxide hydrolase-like), an efflux protein (EncT) and nitrosoguanidine resistance protein (export permease SNG1), heme transport and iron uptake (Csa2), NAD(P)-dependent short-chain dehydrogenase/reductase (SDR family), and NAD-dependent oxidoreductase. Additionally, ATS upregulated genes related to lipid metabolism (Wax synthase domain-containing protein), eisosome formation (SUR7/PalI family-domain-containing protein), and response to environmental stress, G1 cell cycle phase arrest, and negative regulation of autophagy (Table 2).

Consistently, considering the downregulated genes of ATS in the presence vs. absence of acetic acid, GO terms related to translation, rRNA processing, helicase activity, small subunit processome, and methyltransferase activity were enriched (Figure 4C). The Parental strain presented a similar enrichment profile (Figure 4D), suggesting that these functions are affected by acetic acid and/or might be a strategy for saving energy and reducing power for stress response. Some of the main downregulated functions were also shared by the two strains, such as peptide transport and amino acid transport in the cell membrane and vacuole (peptide transporter PTR2, AAT family amino acid transporter), major facilitator superfamily (MFS) transporters, glycosidases that act in the components of the cell wall (glycan and mannan), nucleolin, and the ribosomal production factor 2 (Table 3).

Table 3 – Top 10 downregulated genes for *P. laurentii* ATS and Parental strain in the presence vs. absence of acetic acid.

Gene ID	Protein ID	Log_2(FC)	*p*-adjusted	Function
		ATS		
gene_2278	471404	-5.90	4E-284	Peptide transporter PTR2
gene_1958	174669	-5.07	5E-20	Major facilitator superfamily domain-containing protein
gene_697	537457	-5.01	3E-142	Alpha-mannoside beta-1,6-n-acetylglucosaminyltransferase
gene_2263	538760	-3.61	0E+00	AAT family amino acid transporter
gene_3736	488485	-3.60	2E-95	Six-hairpin glycosidase-like protein
gene_7019	499983	-3.42	0E+00	Nucleolin
gene_5881	496105	-3.00	1E-40	Ribosome production factor 2
gene_6098	542018	-2.92	6E-71	Chaperone (Nopp140-like)
gene_4071	438635	-2.88	3E-291	Transmembrane receptor (DUF1996)
gene_3135	451850	-2.82	5E-68	ATP-dependent RNA helicase HAS1
		Parental		
gene_2263	538760	-6.00	0E+00	AAT family amino acid transporter
gene_307	14588	-5.24	0E+00	Major facilitator superfamily domain-containing protein
gene_2278	471404	-4.85	3E-199	Peptide transporter PTR2
gene_6158	512494	-4.47	0E+00	Sterol receptor domain-containing protein
gene_1491	504211	-4.28	0E+00	3-oxoacid CoA-transferase
gene_3736	488485	-4.20	2E-79	Six-hairpin glycosidase-like protein
gene_7019	499983	-4.09	0E+00	Nucleolin
gene_1958	174669	-3.98	4E-90	Major facilitator superfamily domain-containing protein
gene_7024	93202	-3.80	0E+00	Hypothetical protein
gene_5881	496105	-3.75	1E-69	Ribosome production factor 2

Even though we could not detect enrichment terms for upregulated genes for both strains comparing stressed with unstressed conditions, their most differentially induced genes also presented some similarities, specifically, ricin B-like lectin and RTA-like protein (Table 4). However, the most upregulated genes mainly differed between strains. It is noteworthy that the top 10 upregulated genes of ATS included three transporters – MFS, ATP-binding cassette (ABC), and general amino acid permease – and the ferric reductase transmembrane component 4. Meanwhile, the top 10 of the Parental strain presented two proteins related to carbohydrate/cell wall metabolism (glycosyl hydrolase catalytic core-domain containing and putative cell wall organization and biogenesis-related proteins), Swi5-dependent recombination DNA repair protein 1, and a protein related to cobalamin synthesis (Table 4).

Table 4 – Top 10 upregulated genes for *P. laurentii* ATS and Parental strain in the presence vs. absence of acetic acid.

Gene ID	Protein ID	Log_2(FC)	*p*-adjusted	Function
		ATS		
gene_2144	522523	4.79	1E-63	Expressed protein
gene_6960	499819	2.72	0E+00	Ricin B-like lectin
gene_2156	471301	2.71	0E+00	Ferric reductase transmembrane component 4
gene_4442	491066	2.43	2E-29	Major facilitator superfamily general substrate transporter
gene_7325	500964	2.09	2E-08	Putative general amino acid permease
gene_1177	59445	1.92	6E-34	ABC transporter family protein
gene_7214	500585	1.83	0E+00	HotDog domain-containing protein
gene_3584	526018	1.79	5E-121	Expressed protein
gene_6961	90526	1.67	2E-74	Ricin B-like lectin
gene_76	2738	1.66	2E-75	RTA-like protein
		Parental		
gene_4345	490704	5.86	0E+00	Putative cobalamin synthesis
gene_3635	488166	4.25	3E-23	Hypothetical protein
gene_3317	452077	3.17	2E-206	Glycosyl hydrolase catalytic core-domain containing protein
gene_2144	522523	2.79	9E-99	Expressed protein
gene_2481	198321	2.63	1E-36	Expressed protein
gene_76	2738	2.60	0E+00	RTA-like protein
gene_1989	508744	2.59	8E-137	CRIB domain-containing protein
gene_6960	499819	2.54	0E+00	Ricin B-like lectin
gene_50	475966	2.41	2E-10	Putative cell wall organization and biogenesis-related protein (glycosylase)
gene_5716	360080	2.24	5E+19	Swi5-dependent recombination DNA repair protein 1

Moreover, even in the absence of acetic acid, the main enriched functions were maintained for down and upregulated genes in the ATS vs. the Parental strain scenario, including the downregulation of translation, helicase activity, and ribosome processing, and the induction of oxidoreductase activity and transport (Figures 4E and F). Notably, the main (top 10) downregulated genes/functions differed, including monosaccharide transport, carbamoyl-phosphate synthesis, sodium/potassium transport, and amino acid transport (Table 5).

Table 5 – Top 10 down and upregulated genes for *P. laurentii* ATS vs. Parental strain in the absence of acetic acid.

Gene ID	Protein ID	Log_2(FC)	*p*-adjusted	Function
Downregulated				
gene_3651	535023	-4.30	3E-51	Monosaccharide transporter
gene_4900	492621	-4.05	8E-152	Sodium/potassium-transporting ATPase subunit alpha
gene_2144	522523	-3.85	2E-34	Expressed protein
gene_6467	442885	-3.36	0E+00	UbiD decarboxylyase family
gene_144	459091	-2.89	9E-75	Carbamoyl-phosphate synthase (glutamine-hydrolyzing)
gene_5506	494794	-2.81	2E-213	Putative amino acid transporter
gene_306	476904	-2.78	0E+00	Glycosyl hydrolase catalytic core-domain containing protein
gene_2608	484575	-2.77	1E-242	RNA cytidine acetyltransferase NAT10
gene_4804	464948	-2.70	2E-237	rRNA biogenesis protein RRP5
gene_2263	538760	-2.70	3E-51	AAT family amino acid transporter
Upregulated				
gene_6230	382679	8.60	0E+00	Alpha/Beta hydrolase (Epoxide hydrolase-like)
gene_1738	508509	6.51	0E+00	Putative CIP1
gene_433	22001	5.20	0E+00	Short-chain dehydrogenase/reductase SDR family
gene_1546	521346	4.68	0E+00	Nitroreductase
gene_7306	500850	3.59	0E+00	Epoxide hydrolase
gene_3016	534783	3.38	0E+00	NAD-dependent oxidoreductase
gene_6158	512494	3.30	0E+00	Hypothetical protein
gene_64	476044	3.26	0E+00	Efflux protein EncT
gene_323	476944	3.23	1E-289	Nitrosoguanidine resistance protein SNG1
gene_697	537457	3.03	1E-110	Glycosyltransferase family 18 protein

To further explore the expression profile differences between ATS and the Parental strain, we analyzed in detail their main (top 20) unique DEGs in the presence vs. absence of acetic acid (Tables 6 and 7). ATS uniquely downregulated glycosyltransferase and glycoside hydrolase, Ahpc/TSA antioxidation, phosphate transport, amino acid transport, purine transport, vacuolar ATP synthase, cell cycle regulation, and lipid metabolism (Table 6). Regarding the upregulated functions, ATS presented a unique induction of transporters (aquaporin-like, ABC, and amino acid permease), ferric reductases, malate dehydrogenases, oxidoreductase, glycosyltransferase, and signal transduction (Table 6).

In contrast to ATS, the Parental strain displayed eight transporters in the top 20 uniquely downregulated genes, enzymes related to carbohydrate and secondary metabolism, detoxification of nitric oxide and organic acids, and lipid synthesis (Table 7). The profile of the top 20 uniquely induced genes of the Parental strain also differed from ATS and included proteins related to cobalamin synthesis, cell wall synthesis and organization, DNA repair, oxidases, transcriptional regulation, and carbohydrate and ammonium transport (Table 7).

Table 6 – Top 20 unique down and upregulated genes for *P. laurentii* ATS in the presence vs. absence of acetic acid.

Gene ID	Protein ID	Log_2(FC)	*p*-adjusted	Function
		Downregulated		
gene_697	537457	-5.01	3E-142	Glycosyltransferase family 18 protein
gene_5483	494674	-2.34	8E-26	Expressed protein
gene_6804	499269	-2.07	8E-10	Expressed protein
gene_2533	200583	-1.93	2E-20	Expressed protein
gene_2631	484648	-1.83	7E-07	Expressed protein
gene_4269	287012	-1.70	2E-25	Ahpc/TSA antioxidant-domain-containing protein
gene_3875	526601	-1.59	6E-05	CFEM domain-containing protein
gene_3716	535042	-1.57	1E-162	Vacuolar ATP synthase subunit E
gene_3289	239725	-1.56	7E-26	DUF4385 family protein
gene_7345	106947	-1.55	4E-55	Putative Pi-transporter A-1
gene_5084	493242	-1.55	7E-20	ATP-dependent RNA helicase PRH1
gene_5476	346524	-1.50	2E-50	SH3 domain-containing protein
gene_4863	492483	-1.48	1E-64	Putative long-chain-fatty-acid-CoA ligase
gene_7665	502290	-1.39	1E-10	Transmembrane amino acid transporter
gene_2961	436451	-1.39	3E-04	Purine nucleoside permease
gene_6355	497663	-1.35	2E-23	Phosphatidylethanolamine-binding protein PEBP
gene_3608	488095	-1.30	6E-21	Putative small nucleolar ribonucleoprotein
gene_2978	485861	-1.28	4E-09	Glycoside hydrolase family 5 protein
gene_2200	185921	-1.27	2E-116	Hypothetical protein
gene_7514	548724	-1.22	3E-15	Bud site selection-related protein
		Upregulated		
gene_2156	471301	2.71	0E+00	Ferric reductase transmembrane component 4
gene_7325	500964	2.09	2E-08	Putative general amino acid permease
gene_1177	59445	1.92	6E-34	ABC transporter family protein
gene_7214	500585	1.83	0E+00	Thioesterase/thiol ester dehydrase-isomerase
gene_3584	526018	1.79	5E-121	Expressed protein
gene_6961	90526	1.67	2E-74	Ricin B-like lectin
gene_5231	493734	1.47	5E-13	Aquaporin-like protein
gene_4063	426274	1.41	5E-09	Expressed protein
gene_7152	443644	1.36	1E-29	Ferric reductase NAD binding domain-containing protein
gene_2642	484708	1.34	3E-81	Malate dehydrogenase
gene_7324	421953	1.32	2E-65	Clavaminate synthase-like protein
gene_379	18512	1.31	8E-20	Expressed protein
gene_7927	503648	1.30	9E-64	Malate dehydrogenase
gene_7322	500950	1.30	6E-72	Hri1-like protein
gene_5033	493090	1.29	1E-03	Glycosyltransferase family 18 protein
gene_806	507690	1.27	2E-56	SUR7/PalI family-domain-containing protein
gene_5864	496059	1.23	3E-19	UDP-Glycosyltransferase/glycogen phosphorylase
gene_2644	484710	1.20	3E-25	Expressed protein
gene_1654	159427	1.19	2E-72	GAL4-like protein
gene_4482	298123	1.18	7E-13	S-adenosyl-L-methionine-dependent methyltransferase

Table 7 – Top 20 unique down and upregulated genes for *P. laurentii* Parental strain in the presence vs. absence of acetic acid.

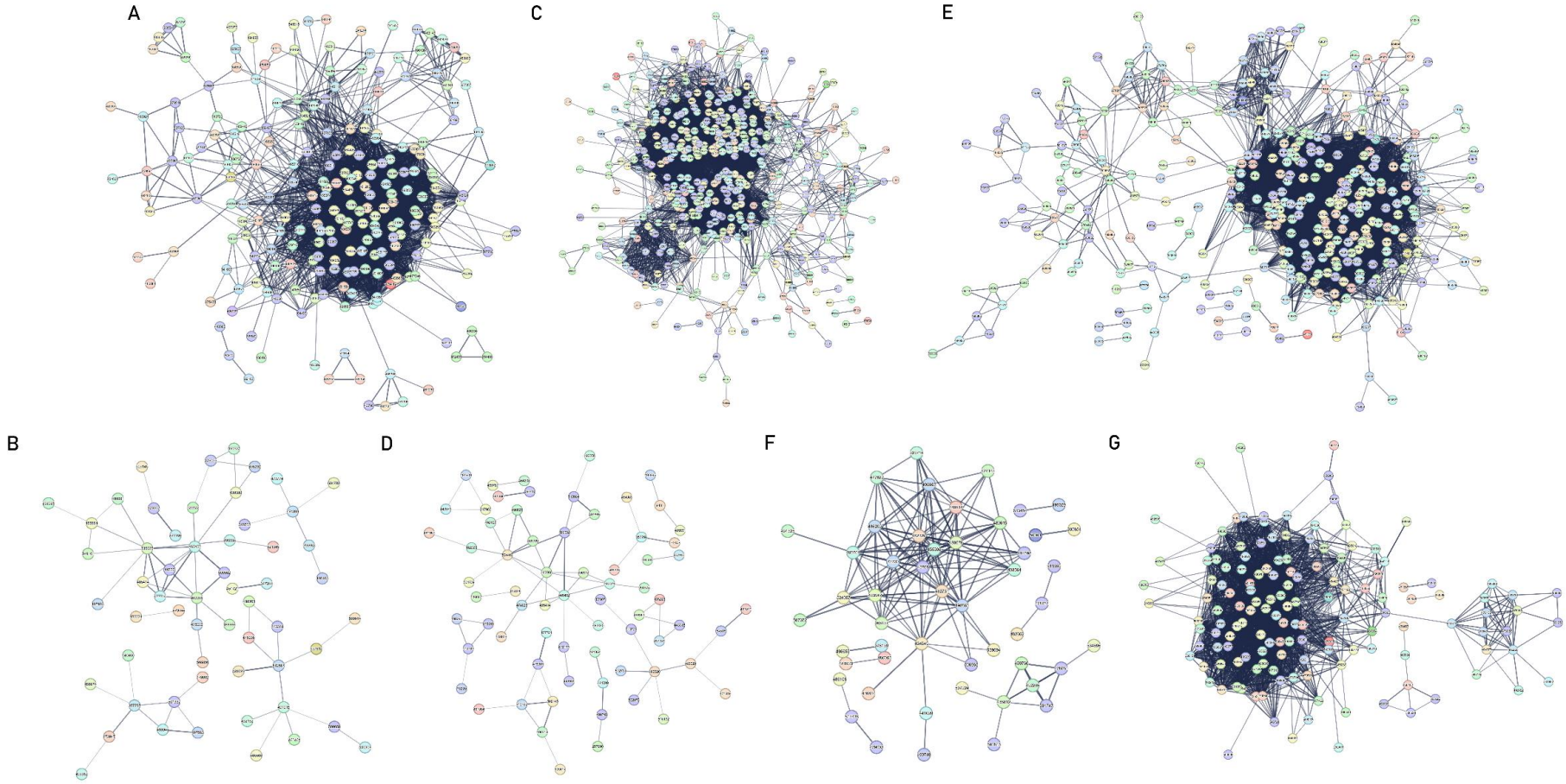
Gene ID	Protein ID	Log_2(FC)	*p*-adjusted	Function
Downregulated				
gene_3394	487408	-3.61	0E+00	Putative MFS monocarboxylate transporter
gene_5108	493334	-3.35	0E+00	Nitric oxide dioxygenase
gene_306	476904	-3.31	0E+00	Glycosyl hydrolase catalytic core-domain containing protein
gene_4185	490091	-3.25	0E+00	Formate dehydrogenase
gene_3393	525572	-3.17	3E-122	Alpha/beta hydrolase fold-1 protein
gene_5506	494794	-3.15	1E-230	Putative amino acid transporter
gene_4882	473435	-3.04	5E-121	Expressed protein
gene_4019	489409	-2.92	0E+00	Putative zinc-finger protein (Zpr1p)
gene_6657	498806	-2.91	0E+00	NADP-dependent alcohol dehydrogenase 6
gene_1290	480260	-2.81	2E-22	MFS multidrug transporter
gene_3008	451721	-2.77	5E-34	2-epi-5-epi-valiolone synthase
gene_1291	480268	-2.71	3E-17	PNRC100 replication protein H-like
gene_5557	494988	-2.63	6E-265	Amino acid permease-2F SLC12A domain-containing protein
gene_3568	487987	-2.62	8E-204	Major facilitator superfamily domain-containing protein
gene_7470	458405	-2.61	0E+00	Major facilitator superfamily domain-containing protein
gene_6605	498604	-2.55	8E-77	Solute carrier family 25 (mitochondrial dicarboxylate transporter)
gene_4952	318427	-2.50	0E+00	Beta-glucuronidase
gene_1236	449605	-2.47	0E+00	Putative glycerol-3-phosphate dehydrogenase
gene_6953	90295	-2.46	2E-13	Acyl-CoA N-acyltransferase
gene_4926	505587	-2.44	1E-277	Putative mitochondrial iron transporter
Upregulated				
gene_4345	490704	5.86	0E+00	Putative cobalamin synthesis protein
gene_2481	198321	2.63	1E-36	Expressed protein
gene_50	475966	2.41	2E-10	Cell wall organization and biogenesis protein (glycosylase)
gene_5716	360080	2.24	6E-19	Swi5-dependent recombination DNA repair protein 1
gene_4928	473472	2.22	5E-121	Methionyl-tRNA formyltransferase
gene_3693	488377	2.19	2E-23	General substrate transporter
gene_1768	470991	2.18	3E-172	Galactose oxidase
gene_1918	482371	2.17	1E-176	RNI-like protein
gene_4060	438704	2.12	6E-33	Amine oxidase
gene_6693	498934	2.00	8E-17	Ammonium transporter AmtB-like domain-containing protein
gene_1305	480348	1.90	2E-17	Fungal-specific transcription factor domain-containing protein
gene_4404	490904	1.89	6E-73	Fungal transcriptional regulatory protein
gene_852	537581	1.89	1E-07	Expressed protein
gene_4308	472974	1.86	1E-101	Expressed protein
gene_6001	506180	1.79	2E-09	Trehalose transport-related protein (maltose permease)
gene_5019	529511	1.73	4E-11	Myc-type, basic helix-loop-helix (bHLH) domain-containing protein
gene_573	537346	1.71	1E-34	Expressed protein
gene_471	24159	1.71	7E-151	Expressed protein
gene_6945	499748	1.71	4E-06	DNA mismatch repair protein MSH5
gene_3989	489321	1.70	5E-06	Expressed protein

4.4.3. Protein-protein interaction (PPI) network

Thereafter, we constructed PPI networks using DEGs in different conditions/strains to gain more insights into the response to acetic acid stress. We found that the networks for downregulated genes (Figures 5A, C, E, and G) were more complex and presented more interactions than those for upregulated ones (Figures 5B, D, and F). This result could be related to the number of proteins downregulated compared to upregulated ones, except for ATS vs. Parental strain in the presence of acetic acid, which showed a more balanced relationship (332 downregulated vs. 281 upregulated) (Table 1). Besides being more complex and involving more proteins, the networks for downregulated genes had more interactions and with higher confidence scores, suggesting that the regulation associated with gene repression is more coordinated than the gene induction (Figure 5).

Consistent with the expression profile, the PPI networks for downregulated genes for ATS vs. Parental strain in both presence and absence of acetic acid (Figures 5A and C) were enriched for transcriptional and translational processes/functions/pathways and regulation, such as ribosome biogenesis, rRNA processing, non-coding RNA (ncRNA) processing, and Cap-dependent translation initiation and elongation. Interestingly, in the absence of the selective pressure (acetic acid), the PPI network of ATS was also enriched for three different Nonsense-Mediated Decay (NMD) maps, which act on mRNAs with premature nonsense codons. The enrichments in the upregulated networks for these two scenarios (Figures 5B and D) were also consistent with the expression profile. For instance, in the presence of acetic acid, the induced network of ATS was enriched for oxidoreductase activity, transmembrane proteins, and the synthesis of secondary metabolites. Meanwhile, apart from the similar enrichment for secondary metabolite synthesis, the upregulated network of ATS in the absence of acetic acid was enriched for inositol transport, threonine metabolism, and alpha-linolenic acid metabolism. The complete list of enriched terms in the PPI networks can be found at: https://github.com/LabFisUFV/PlaurentiiAceticAcidStress/tree/master/STRING.

Figure 5 - Protein-protein interaction (PPI) networks. (A) Down and (B) upregulated genes for ATS vs. Parental strain in the presence of acetic acid; (C) Down and (D) upregulated genes for ATS vs. Parental strain in the absence of acetic acid; (E) Down and (F) upregulated genes for the Parental strain in the presence vs. absence of acetic acid; (E) Downregulated genes for ATS in the presence vs. absence of acetic acid. Nodes represent proteins, and edges represent interactions; the thickness of the edge represents the confidence score for the interaction, and the thicker the line, the higher the score.



The PPI network enrichments of downregulated genes for ATS and Parental strain in the presence vs. absence of acetic acid (Figures 5E and G) presented a similar profile to the gene expression profile, including transcriptional and translational functions and mRNA and ncRNA processing. Notably, the NMD was enriched only for the parental strain. Since this function, repressed in the Parental strain under stress, was enriched in the network of downregulated genes of ATS vs. Parental strain in the absence of acetic acid, it might comprehend a constitutive alteration developed during the ALE, which would explain why a similar behavior was not detected in the PPI network of ATS in the presence vs. absence of acetic acid.

Furthermore, the functions enriched in the PPI network for upregulated genes of the Parental strain in the presence of acetic acid (Figure 5F) corroborated the induction of DNA replication and repair. Notably, cell cycle-related functions/pathways, including yeast meiosis, were enriched. This result indicates that the acetic acid stress may induce the sexual reproduction of the sensitive (Parental) strain of *P. laurentii*. Although no significant network was found for the upregulated genes of ATS in the presence of acetic acid, only one protein (Protein ID: 506965) was related to the cell cycle (the mitotic spindle assembly checkpoint), indicating that the induction of meiosis and DNA replication/repair was specific for the Parental strain.

4.4.4. Reaction activity score (RAS)

Furthermore, we integrated the RNA-seq data with the genome-scale metabolic network of *P. laurentii* (papla-GEM) to determine which metabolic reactions with GPR associations were induced or repressed at the transcriptional level (Table 8). In the absence of acetic acid, ATS induced cysteine transport (vacuole to cytosol), nitrogen metabolism (L-asparaginase), and amino/nucleotide sugar metabolism (N-acetyl-D-glucosamine-6-phosphate amidohydrolase); meanwhile, ATS repressed (top 20) the transport of many amino acids, oxidoreductase activity, and galactose metabolism (Table 9).

Table 8 – Summary of reactions from *papla*-GEM with differential RAS in *Papiliotrema laurentii* ATS and Parental strain in the presence and absence of acetic acid stress.

	Differential RAS*		
Condition	**Total**	**Induced****	**Repressed****
Parental (presence vs. absence of acetic acid)	150	-	150 (100%)
ATS (presence vs. absence of acetic acid)	36	3 (8.33%)	33 (91.67%)
ATS vs. Parental in the absence of acetic acid	107	4 (3.73%)	103 (96.27%)
ATS vs. Parental in the presence of acetic acid	56	23 (41.1%)	34 (58.9%)

* *p*-adjusted < 0.05 but |Log_2 [fold change (FC)]| > 1.2. ** Induced: Log_2 (FC) > 1.2; Repressed: Log_2 (FC) < -1.2.

Table 9 – Reactions on *papla*-GEM with differential RAS for *P. laurentii* ATS vs. Parental strain in the absence of acetic acid.

Reaction ID	**Reaction Name**	**Log_2(FC)**	***p*-adjusted**
	Induced		
r_1193	L-cysteine transport (vacuole to cytosol)	1.81	0.0044
r_0679	L-asparaginase (cytosol)	1.26	0.0192
r_0680	L-asparaginase (extracellular)	1.26	0.0192
y300073	N-Acetyl-D-glucosamine-6-phosphate amidohydrolase	1.22	0.0093
	Top 20 repressed		
y300059	succinyl-CoA:acetoacetate CoA-transferase	-2.52	0.0062
r_0471	glutamate dehydrogenase (NADP)	-2.48	0.0006
r_4275	Fe(II):NADP+ oxidoreductase (cytosol – extracellular)	-2.44	0.0049
r_4276	Fe(II):NADP+ oxidoreductase (cytosol – vacuole)	-2.44	0.0049
r_1191	L-carnitine transport	-2.31	0.0012
r_0458	Galactokinase	-2.29	0.0012
r_4222	ATP:alpha-D-galactose 1-phosphotransferase	-2.29	0.0012
r_1255	S-adenosyl-L-methionine transport	-2.28	0.0006
r_1257	S-methylmethionine permease	-2.28	0.0006
r_1101	4-aminobutyrate transport	-2.15	0.0006
r_1173	glycine transport	-2.15	0.0006
r_1183	L-alanine transport	-2.15	0.0006
r_1184	L-arganine transport	-2.15	0.0006
r_1186	L-asparagine transport	-2.15	0.0006
r_1190	L-aspartate transport	-2.15	0.0006
r_1192	L-cysteine transport	-2.15	0.0006
r_1196	L-glutamate transport	-2.15	0.0006
r_1199	L-glutamine transport	-2.15	0.0006
r_1201	L-histidine transport	-2.15	0.0006
r_1205	L-isoleucine transport	-2.15	0.0006

In the presence of acetic acid, compared to the Parental strain, ATS repressed galactose metabolism and ammonia transport (Table 10), which was similar to the absence of the acid. Interestingly, ATS repressed carbohydrate metabolism, including erythrose and xylose reductase. The metabolism of lipids, purines, pyrimidines, and amino acids was repressed, including key enzymes, such as acetyl-CoA carboxylase, glutamate dehydrogenase, fatty-acyl-CoA synthase, CTP and GMP synthases, and pyruvate kinase. Additionally, the metabolism of one-carbon compounds presented many repressed reactions. These processes, especially biosynthetic ones, involve enzymes that require reduced cofactors [e.g., NAD(P)H], which are important in the cell response to toxic compounds.

Indeed, different reactions related to NAD(P)+ recycling in the cell (i.e., aldehyde and formaldehyde dehydrogenases) were induced by ATS compared to the Parental strain (Table 10). Moreover, reactions involved in stress response and metabolism of toxic compounds (i.e., lactoylglutathione lyase, water diffusion, ornithine transaminase, and dihydroxyacetone kinase) were induced by ATS, as well as the transport of various organic acids (Table 10). Besides, the transport of L-cystine from the vacuole to the cytosol was induced, similar to the absence of acetic acid (Table 9).

Table 10 – Reactions on *papla*-GEM with differential RAS for *P. laurentii* ATS vs. Parental strain in the presence of acetic acid.

Reaction ID	Reaction Name	Log_2(FC)	*p*-adjusted
	Induced		
r_0697	lactoylglutathione lyase (glyoxalase)	2.03	0.0046
r_1193	L-cystine transport (vacuole to cytosol)	1.94	0.0032
r_1277	water diffusion	1.84	0.0071
r_0146	adenosylmethionine-8-amino-7-oxononanoate transaminase	1.72	0.0012
r_0819	ornithine transaminase	1.72	0.0012
r_0171	aldehyde dehydrogenase (2-phenylethanol, NADP)	1.72	0.0017
r_0181	aldehyde dehydrogenase (isoamyl alcohol, NADP)	1.72	0.0017
r_0184	aldehyde dehydrogenase (isobutyl alcohol, NADP)	1.72	0.0017
r_0443	formaldehyde dehydrogenase	1.66	0.0027
r_0166	aldehyde dehydrogenase (2-methylbutanol, NAD)	1.60	0.0017
r_0169	aldehyde dehydrogenase (2-phenylethanol, NAD)	1.60	0.0017
r_0179	aldehyde dehydrogenase (isoamyl alcohol, NAD)	1.60	0.0017
r_0182	aldehyde dehydrogenase (isobutyl alcohol, NAD)	1.60	0.0017

r_0186	aldehyde dehydrogenase (tryptophol, NAD)	1.60	0.0017
r_1798	fumarate(2-) exchange	1.51	0.0012
r_1901	L-malate transport	1.51	0.0012
r_2057	succinate transport	1.51	0.0012
r_1136	D-lactate transport	1.51	0.0012
r_1206	D-lactate transport	1.51	0.0012
r_1207	L-lactate transport	1.51	0.0012
r_1254	pyruvate transport	1.51	0.0012
r_0354	dihydroxyacetone kinase	1.20	0.0016
r_0905	phosphopantothenate-cysteine ligase	1.23	0.0048
	Repressed		
r_0458	galactokinase	-2.83	0.0048
r_4222	ATP:alpha-D-galactose 1-phosphotransferase	-2.83	0.0048
r_0471	glutamate dehydrogenase (NADP)	-2.29	0.0050
r_1115	ammonia transport	-2.06	0.0012
y300040	erythrose reductase	-1.93	0.0086
r_0108	acetyl-CoA carboxylase	-1.86	0.0201
r_0459	galactose-1-phosphate uridylyltransferase	-1.80	0.0040
r_1071	UDPglucose--hexose-1-phosphate uridylyltransferase	-1.80	0.0040
r_2140	fatty-acyl-CoA synthase (n-C16:0CoA)	-1.77	0.0087
r_2141	fatty-acyl-CoA synthase (n-C18:0CoA)	-1.77	0.0087
r_0145	adenosylmethionine decarboxylase	-1.74	0.0067
r_1070	UDPglucose 4-epimerase	-1.59	0.0020
r_0548	homoserine kinase	-1.57	0.0050
r_0139	adenine phosphoribosyltransferase	-1.49	0.0030
r_0773	NADH:ubiquinone oxidoreductase	-1.47	0.0014
r_0075	5'-methylthioadenosine phosphorylase	-1.47	0.0029
r_0943	purine-nucleoside phosphorylase	-1.47	0.0029
r_0306	CTP synthase (glutamine)	-1.46	0.0054
r_0307	CTP synthase (NH3)	-1.46	0.0054
r_1093	xylose reductase	-1.39	0.0086
r_0214	aspartate carbamoyltransferase	-1.38	0.0229
r_0250	carbamoyl-phosphate synthase (glutamine-hydrolysing)	-1.38	0.0229
r_0446	formate-tetrahydrofolate ligase (cytosolic)	-1.36	0.0236
r_0447	formate-tetrahydrofolate ligase (mithocondrial)	-1.36	0.0236
r_0724	methenyltetrahydrifikate cyclohydrolase (mithocondrial)	-1.36	0.0236
r_0725	methenyltetrahydrifikate cyclohydrolase (cytosolic)	-1.36	0.0236
r_0732	methylenetetrahydrofolate dehydrogenase (NADP) (cytosolic)	-1.36	0.0236
r_0733	methylenetetrahydrofolate dehydrogenase (NADP) (mitochondrial)	-1.36	0.0236
r_0962	pyruvate kinase	-1.33	0.0041
r_0726	methionine adenosyltransferase	-1.30	0.0115
r_0514	GMP synthase	-1.24	0.0162
r_0349	dihydroorotase	-1.24	0.0046
r_0144	adenosylhomocysteinase	-1.21	0.0247
r_0555	hydroxybenzoate octaprenyltransferase	-1.20	0.0078

Concerning the presence vs. absence of acetic acid, ATS repressed various reactions involved in carnitine metabolism and transport, as well as the transport and metabolism of purines and pyrimidines (Table 11). Additionally, the conversion of citrate to isocitrate and CoA synthesis was repressed, which might be related to the increased supply and gluconeogenic flux due to acetic uptake and catabolism. ATS induced only three reactions: arginase and two Fe(II):NADP+ oxidoreductases (Table 11). This induction of oxidoreductase is consistent with the gene expression and PPI network analysis and underscores the importance of redox balance. The induction of arginase, which catalyzes the conversion of arginine to ornithine and urea, likely contributes to the pH homeostasis in the cytosol and counteract the pH decrease promoted by acetic acid dissociation.

Table 11 – Reactions on *papla*-GEM with differential RAS for *P. laurentii* ATS in the presence vs. absence of acetic acid.

Reaction ID	Reaction Name	Log_2(FC)	*p*-adjusted
	Induced		
r_4275	Fe(II):NADP+ oxidoreductase (cytosol – extracellular)	2.60	0.0016
r_4276	Fe(II):NADP+ oxidoreductase (cytosol – vacuole)	2.60	0.0016
r_0206	arginase	1.43	0.0008
	Repressed		
r_1120	carnitine-acetylcarnitine carrier	-2.19	0.0099
t_0015	fatty acylcarnitine transport (C16:0)	-2.19	0.0099
t_0016	fatty acylcarnitine transport (C18:0)	-2.19	0.0099
t_0017	fatty acylcarnitine transport (C20:0)	-2.19	0.0099
t_0018	fatty acylcarnitine transport (C22:0)	-2.19	0.0099
t_0019	fatty acylcarnitine transport (C24:0)	-2.19	0.0099
t_0020	fatty acylcarnitine transport (C26:0)	-2.19	0.0099
y300059	succinyl-CoA:acetoacetate CoA-transferase	-2.05	0.0012
r_1667	bicarbonate formation	-1.85	0.0008
r_0252	carnitine O-acetyltransferase	-1.69	0.0008
t_0009	cytoplasmatic carnitine acyltransferase (C16:0)	-1.69	0.0008
t_0010	cytoplasmatic carnitine acyltransferase (C18:0)	-1.69	0.0008
t_0011	cytoplasmatic carnitine acyltransferase (C20:0)	-1.69	0.0008
t_0012	cytoplasmatic carnitine acyltransferase (C22:0)	-1.69	0.0008
t_0013	cytoplasmatic carnitine acyltransferase (C24:0)	-1.69	0.0008
t_0014	cytoplasmatic carnitine acyltransferase (C26:0)	-1.69	0.0008
r_1108	adenine transport	-1.66	0.0107
r_1133	cytosine transport	-1.66	0.0107

r_1176	guanine transport	-1.66	0.0107
r_1255	S-adenosyl-L-methionine transport	-1.41	0.0009
r_1257	S-methylmethionine permease	-1.41	0.0009
r_0253	carnitine O-acetyltransferase (peroxisomal)	-1.34	0.0046
r_0254	carnitine O-acetyltransferase (mitochondrial)	-1.34	0.0046
r_1191	L-carnitine transport	-1.31	0.0008
r_1070	UDPglucose 4-epimerase	-1.26	0.0042
r_0565	IMP dehydrogenase	-1.25	0.0031
r_0280	cis-aconitate(3-) to isocitrate	-1.24	0.0269
r_0302	citrate to cis-aconitate(3-), mitochondrial	-1.24	0.0269
r_0303	citrate to cis-aconitate(3-), cytoplasmic	-1.24	0.0269
r_2305	cis-aconitate(3-) to isocitrate	-1.24	0.0269
r_0842	pantothenate kinase	-1.21	0.0099
y200015	ATP:pantothenate 4'-phosphotransferase	-1.21	0.0099

Although no reactions were predicted to be induced by the Parental strain in the presence vs. absence of acetic acid, two main functions were repressed (top 40): transport of amino acids and amino-derived compounds and the metabolism of polyamines (Table 12). The high regulation of ammonium/amino acid/peptide transport in the absence and presence of acetic acid by ATS and the Parental strain was in agreement with the differential expression and PPI network analyses, and appears to be related to the maintenance of intracellular pH homeostasis, as well as the abundance of ammonium in the growth media and the reduced requirement for protein synthesis based on the repression of many genes related to translation.

Table 12 – Top 40 repressed reactions on *papla*-GEM for *P. laurentii* Parental strain in the presence vs. absence of acetic acid.

Reaction ID	**Reaction Name**	**Log_2(FC)**	***p*-adjusted**
y300059	succinyl-CoA:acetoacetate CoA-transferase	-4.30	0.0074
r_1191	L-carnitine transport	-3.65	0.0012
r_1255	S-adenosyl-L-methionine transport	-3.60	0.0008
r_1257	S-methylmethionine permease	-3.60	0.0008
r_1101	4-aminobutyrate transport	-3.27	0.0008
r_1173	glycine transport	-3.27	0.0008
r_1183	L-alanine transport	-3.27	0.0008
r_1184	L-arginine transport	-3.27	0.0008
r_1186	L-asparagine transport	-3.27	0.0008
r_1190	L-aspartate transport	-3.27	0.0008

r_1192	L-cystine transport	-3.27	0.0008
r_1196	L-glutamate transport	-3.27	0.0008
r_1199	L-glutamine transport	-3.27	0.0008
r_1201	L-histidine transport	-3.27	0.0008
r_1205	L-isoleucine transport	-3.27	0.0008
r_1211	L-leucine transport	-3.27	0.0008
r_1213	L-lysine transport	-3.27	0.0008
r_1214	L-methionine transport	-3.27	0.0008
r_1215	L-phenylalanine transport	-3.27	0.0008
r_1216	L-proline transport	-3.27	0.0008
r_1218	L-threonine transport	-3.27	0.0008
r_1219	L-tryptophan transport	-3.27	0.0008
r_1223	L-tyrosine transport	-3.27	0.0008
r_1224	L-valine transport	-3.27	0.0008
r_1238	ornithine transport	-3.27	0.0008
r_1260	spermidine transport (cytosol – extracellular)	-2.64	0.0075
r_1251	putrescine transport (cytosol – extracellular)	-2.55	0.0098
r_1252	putrescine transport (cytosol – vacuole)	-2.55	0.0098
r_1261	spermidine transport (cytosol – vacuole)	-2.55	0.0098
r_1262	spermine transport (cytosol – extracellular)	-2.55	0.0098
r_1263	spermine transport (cytosol – vacuole)	-2.55	0.0098
r_0252	carnitine O-acetyltransferase	-2.54	0.0132
t_0009	cytoplasmatic carnitine acyltransferase (C16:0)	-2.54	0.0132
t_0010	cytoplasmatic carnitine acyltransferase (C18:0)	-2.54	0.0132
t_0011	cytoplasmatic carnitine acyltransferase (C20:0)	-2.54	0.0132
t_0012	cytoplasmatic carnitine acyltransferase (C22:0)	-2.54	0.0132
t_0013	cytoplasmatic carnitine acyltransferase (C24:0)	-2.54	0.0132
t_0014	cytoplasmatic carnitine acyltransferase (C26:0)	-2.54	0.0132
r_1217	L-serine transport	-2.52	0.0012

4.4.5. FVA and RS analysis

To better understand the flux distribution, flexibility, and cofactor pool/yield of *P. laurentii* in the presence and absence of acetic acid, we conducted FVA and RS analysis using the GEM with incorporated k_{cat} values (ecGEM). After reconstructing the enzyme-constrained *papla*-GEM (ec*papla*-GEM), we retrieved the xylose and acetic acid uptake rates, dilution rate, and biomass composition from ALMEIDA et al. (2023) and used pFBA to adjust the exchange rates that were not measured experimentally. Next, we conducted FVA to evaluate which reactions would be more prone to regulation, that is, would present more flexible fluxes, and those that would be restrained or less variable. The minimum and maximum flux range of a given reaction can reflect this variability/flexibility. We considered the range FC between two conditions for the same strain or strains in the same condition.

The cumulative distribution of variability ranges showed that the increase in carbon availability derived from acetic acid uptake increased the global metabolic flexibility for both strains. Notably, ATS presented a higher flexibility than the Parental strain in the presence of the acid, which might be related not only to its higher acetic acid uptake but also to the differences in biomass composition. In the absence of acetic acid, both strains presented a similar cumulative distribution, implying that their metabolic state was closer than compared to in the presence of the acid (Figure 6). The number of relaxed reactions also demonstrated the higher flexibility of ATS metabolism in the presence of the acid compared to the parental strain (628) and in the presence vs. absence of acetic acid (741). In contrast, the Parental strain restrained more reactions than relaxed when exposed to the acid, 24 and 15, respectively (Table 13).

Figure 6 – Cumulative distribution of the variability range for *P. laurentii* ATS and Parental strain in the presence and absence of acetic using ec*papla*-GEM constrained for each strain and condition (biomass composition, growth, and carbon uptake).

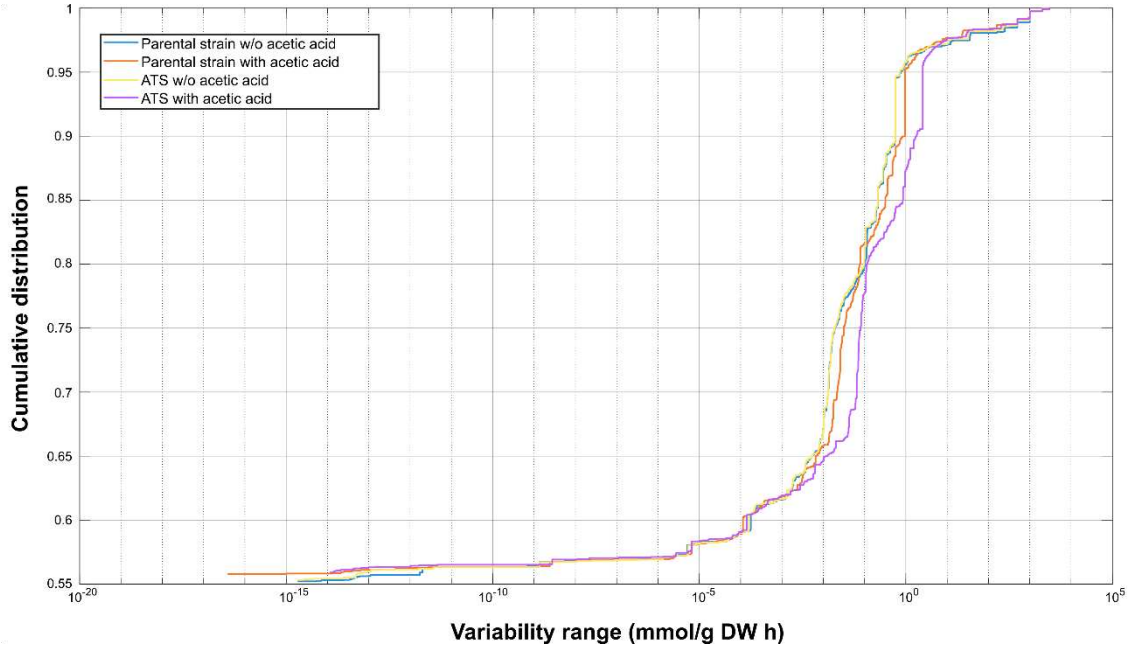


Table 13 - Summary of reactions with differential ranges from *ecpapla*-GEM for *P. laurentii* ATS and Parental strain in the presence and absence of acetic acid stress.

	Differential ranges*		
Condition	**Total**	**Relaxed****	**Restrained****
Parental (presence vs. absence of acetic acid)	39	15 (38.5%)	24 (61.5%)
ATS (presence vs. absence of acetic acid)	754	741 (98.3%)	13 (1.7%)
ATS vs. Parental in the absence of acetic acid	22	7 (31.8%)	15 (68.2%)
ATS vs. Parental in the presence of acetic acid	633	628 (99.2%)	5 (0.8%)

* *p*-adjusted < 0.05 but |Log_2 [fold change (FC)]| > 1.2. ** Relaxed: Log_2 (FC) > 1.2; Restrained: Log_2 (FC) < -1.2.

In the absence of acetic acid, although their cumulative variability was similar (Figure 6), ATS presented a slightly more restrained (15 reactions) than relaxed (7 reactions) flux distribution (Table 13). Most of these restrained reactions were related to lipid synthesis and oxidations, which might be related to the lower lipid content and different fatty acid profile of ATS compared to the Parental strain (Table 14). Interestingly, one reaction involved in the synthesis of phosphatidylcholine was strongly relaxed, as well as one acyl-CoA involved in the β-oxidation of fatty acids, which might also be related to the ATS lipid profile mentioned above.

Table 14 – Reactions with differential ranges on *ecpapla*-GEM for *P. laurentii* ATS vs. Parental strain in the absence of acetic acid.

Reaction ID	**Reaction Name**	**Log_2(FC)**
	Relaxed	
t_1083	PC (1-18:1, 2-18:1) diacylglycerol (1-16:0, 2-18:2) acyltransferase, endoplasmic reticulum membrane	38.87
t_0064	acyl-CoA oxidase (butanoyl-CoA)	9.12
r_0664	isoleucine transaminase	5.88
r_0771	NADH kinase	5.03
r_0503	glycine hydroxymethyltransferase	4.83
r_3533	NAD transport, cytoplasm-ER membrane	4.72
r_3573	octadecanoate (n-C18:0) transport, cytoplasm-lipid particle	2.81
	Restrained	
r_3751	1-acylglycerophosphoethanolamine (16:0) transport, ER membrane-lipid particle	-14.25
t_0041	3-hydroxyacyl-CoA dehydrogenase (3-oxodecanoyl-CoA)	-13.96
t_0049	3-hydroxyacyl-CoA dehydrogenase (3-oxooctanoyl-CoA)	-13.96
r_3652	ADP transport, cytoplasm-vacuolar membrane	-12.59
r_0018	2-aminoadipate transaminase	-5.63
t_0874	linoleate transport, cytoplasm-endoplasmic reticulum membrane	-5.57
t_1114	PE (1-18:1, 2-18:2) diacylglycerol (1-16:0, 2-18:1) acyltransferase, endoplasmic reticulum membrane	-4.92

y300065	L-Serine:pyruvate aminotransferase	-3.12
t_0351	PE (1-16:0, 2-18:2) diacylglycerol (1-16:0, 2-18:2) acyltransferase, endoplasmic reticulum membrane	-2.41
r_3574	octadecenoate (n-C18:1) transport, cytoplasm-lipid particle	-2.36
r_2125	coenzyme A: cytoplasm to LP	-2.23
t_1118	PE (1-18:1, 2-18:2) diacylglycerol (1-16:0, 2-18:2) acyltransferase, endoplasmic reticulum membrane	-2.13
t_1088	PC (1-18:1, 2-18:2) diacylglycerol (1-18:1, 2-18:1) acyltransferase, endoplasmic reticulum membrane	-1.48
t_0233	fatty-acid--CoA ligase (linoleate), lipid particle	-1.44
t_0052	acetyl-CoA C-acyltransferase (acetyl-CoA)	-1.41

In the presence of acetic acid, ATS presented a more flexible metabolic flux distribution than the Parental strain, with 628 relaxed reactions (Table 13 and Figure 6). This set of reactions was enriched in subsystems related to different biosynthetic pathways, including secondary metabolites, antibiotics, and amino acids (Figure 7A). This result, together with the enrichment of purine and pyrimidine metabolism, is consistent with our previous results, demonstrating a high regulation of processes that involve the use of cofactors, especially NAD(P)H, and the metabolism of nitrogenous compounds. The 2-aminoadipate transaminase reaction, involved in the transfer of nitrogenous groups, presented a relaxation of almost 50 times, also highlighting the importance of nitrogen metabolism (Table 15). Additionally, crucial pathways for the assimilation of the carbon sources available (xylose and acetic acid) were enriched, such as gluconeogenesis, glycolysis, glyoxylate and dicarboxylate metabolism, and oxidative phosphorylation (Figure 7A), which might be related to its higher uptake of both carbon sources in this condition, as well as its biomass composition. Similar to the scenario in the absence of acetic acid, the Top 50 relaxed reactions included many processes related to lipid synthesis and oxidation (Table 15). In contrast, the few restrained reactions were related to important transport processes in the presence of organic acids (Table 15), which is also consistent with our previous analysis.

Figure 7 – **Dot plots for enriched metabolic subsystems**: **(A)** Relaxed reactions of *P. laurentii* ATS vs. Parental strain in the presence of acetic acid. **(B)** Relaxed reactions of *P. laurentii* ATS strain in the presence vs. absence of acetic acid. No subsystems were enriched for the other scenarios.

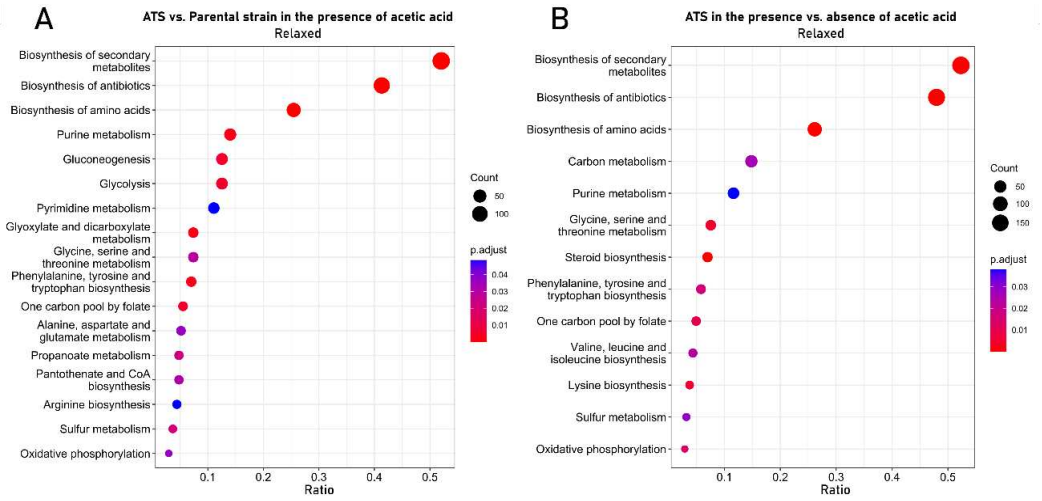


Table 15 – Reactions on *papla*-GEM with differential ranges on *ecpapla*-GEM for *P. laurentii* ATS vs. Parental strain in the presence of acetic acid.

Reaction ID	Reaction Name	Log_2(FC)
	Relaxed (Top 50)	
r_0018	2-aminoadipate transaminase	49.86
t_0874	linoleate transport, cytoplasm-endoplasmic reticulum membrane	4.68
r_3745	phosphatidylethanolamine (1-18:0, 2-18:1) transport, ER membrane-lipid particle	4.08
t_0752	phosphatidylethanolamine (1-18:0, 2-18:2) transport, ER membrane-lipid particle	4.08
t_0227	fatty-acid--CoA ligase (oleate), endoplasmic reticulum membrane	3.83
r_0771	NADH kinase	3.35
r_3513	oleate transport, cytoplasm-ER membrane	3.10
r_1241	pantothenate transport	2.81
r_2125	coenzyme A: cytoplasm to LP	2.57
r_0166	aldehyde dehydrogenase (2-methylbutanol, NAD)	2.52
r_2285	acetyl-CoA C-acyltransferase (hexanoyl-CoA)	2.43
t_0059	acetyl-CoA C-acyltransferase (lauroyl-CoA)	2.40
t_0231	fatty-acid--CoA ligase (stearate), lipid particle	2.21
y300065	L-Serine:pyruvate aminotransferase	2.02
t_0072	acyl-CoA oxidase (stearoyl-CoA)	1.98
t_1118	PE (1-18:1, 2-18:2) diacylglycerol (1-16:0, 2-18:2) acyltransferase, endoplasmic reticulum membrane	1.95
r_0255	catalase	1.94
r_1075	urea carboxylase	1.92
r_0191	allophanate hydrolase	1.92
r_0218	aspartate transaminase	1.92
r_0804	nucleoside triphosphatase	1.90
t_1088	PC (1-18:1, 2-18:2) diacylglycerol (1-18:1, 2-18:1) acyltransferase, endoplasmic reticulum membrane	1.90
t_0538	PS decarboxylase (1-18:0, 2-18:2), vacuolar membrane	1.88

r_1683	choline exchange	1.88
r_0142	adenosine kinase	1.88
r_0144	adenosylhomocysteinase	1.88
r_0726	methionine adenosyltransferase	1.88
r_0727	methionine synthase	1.88
r_3851	phosphatidyl-L-serine (1-18:0, 2-18:1) transport, ER membrane-vacuolar membrane	1.88
t_0763	phosphatidylethanolamine (1-18:0, 2-18:1) transport, ER membrane-vacuolar membrane	1.88
t_0764	phosphatidylethanolamine (1-18:0, 2-18:2) transport, ER membrane-vacuolar membrane	1.88
t_0782	phosphatidyl-L-serine (1-18:0, 2-18:2) transport, ER membrane-vacuolar membrane	1.88
r_3849	phosphatidyl-L-serine (1-16:0, 2-18:1) transport, ER membrane-vacuolar membrane	1.88
t_0760	phosphatidylethanolamine (1-16:0, 2-18:1) transport, ER membrane-vacuolar membrane	1.88
t_0761	phosphatidylethanolamine (1-16:0, 2-18:2) transport, ER membrane-vacuolar membrane	1.88
t_0779	phosphatidyl-L-serine (1-16:0, 2-18:2) transport, ER membrane-vacuolar membrane	1.88
t_0369	PE methyltransferase (1-18:0, 2-18:1), endoplasmic reticulum membrane	1.87
t_0435	phosphatidyl-N,N-dimethylethanolamine methyltransferase (1-18:0, 2-18:1), endoplasmic reticulum membrane	1.87
t_0441	phosphatidyl-N-methylethanolamine methyltransferase (1-18:0, 2-18:1), endoplasmic reticulum membrane	1.87
t_0370	PE methyltransferase (1-18:0, 2-18:2), endoplasmic reticulum membrane	1.87
t_0436	phosphatidyl-N,N-dimethylethanolamine methyltransferase (1-18:0, 2-18:2), endoplasmic reticulum membrane	1.87
t_0442	phosphatidyl-N-methylethanolamine methyltransferase (1-18:0, 2-18:2), endoplasmic reticulum membrane	1.87
t_0366	PE methyltransferase (1-16:0, 2-18:1), endoplasmic reticulum membrane	1.87
t_0432	phosphatidyl-N,N-dimethylethanolamine methyltransferase (1-16:0, 2-18:1), endoplasmic reticulum membrane	1.87
t_0438	phosphatidyl-N-methylethanolamine methyltransferase (1-16:0, 2-18:1), endoplasmic reticulum membrane	1.87
t_0367	PE methyltransferase (1-16:0, 2-18:2), endoplasmic reticulum membrane	1.87
t_0433	phosphatidyl-N,N-dimethylethanolamine methyltransferase (1-16:0, 2-18:2), endoplasmic reticulum membrane	1.87
t_0439	phosphatidyl-N-methylethanolamine methyltransferase (1-16:0, 2-18:2), endoplasmic reticulum membrane	1.87
r_2484	PS decarboxylase (1-16:0, 2-18:1), vacuolar membrane	1.87
r_2470	PS decarboxylase (1-18:0, 2-18:1), mitochondrial membrane	1.87
	Restrained	
r_1811	glycine transport	-13.27
r_1827	H+ diffusion	-4.98
r_3533	NAD transport, cytoplasm-ER membrane	-2.18
r_1184	L-arganine transport	-2.14
r_3538	CTP transport, cytoplasm-ER membrane	-1.66

Taking into account the presence vs. absence of acetic acid, the main metabolic subsystems enriched for ATS were similar to those versus the Parental strain in the presence of the acid (Figures 7A and B). The main difference was the presence of steroid biosynthesis, which, along with the presence of various reactions related to lipid

biosynthesis and oxidation in the Top 50, also highlighted the importance of these processes for ATS in the presence of acetic acid, possibly to control the cofactor pool and membrane composition (reflected by the fatty acid composition differences in the model) (Table 16). Consistently, two reactions related to the metabolism of nitrogenous compounds were highly relaxed - 2-isopropylmalate hydratase and 2-aminoadipate transaminase - as well as others with fewer effects, including aspartate transaminase, glycine hydroxymethyltransferase, and the urea cycle (e.g., argininosuccinate lyase and allophanate hydrolase). The metabolism of ROS was also relaxed in this scenario, represented, for example, by catalase (Table 16). Similar to the scenario compared to the Parental strain in the presence of acetic acid, most restrained reactions were related to important transports in the presence of organic acids (Table 16).

Table 16 – Reactions on *papla*-GEM with differential ranges for *P. laurentii* ATS in the presence vs. absence of acetic acid.

Reaction ID	Reaction Name	Log_2(FC)
	Relaxed (Top 50)	
r_0023	2-isopropylmalate hydratase	44.55
r_0018	2-aminoadipate transaminase	44.30
r_0326	dCMP deaminase	9.67
r_4039	succinyl-CoA:acetate CoA transferase	6.30
r_3513	oleate transport, cytoplasm-ER membrane	4.39
r_0218	aspartate transaminase	4.21
r_0502	glycine hydroxymethyltransferase	4.15
r_3745	phosphatidylethanolamine (1-18:0, 2-18:1) transport, ER membrane-lipid particle	3.91
t_0752	phosphatidylethanolamine (1-18:0, 2-18:2) transport, ER membrane-lipid particle	3.91
t_0231	fatty-acid--CoA ligase (stearate), lipid particle	3.77
r_0207	argininosuccinate lyase	3.68
t_0049	3-hydroxyacyl-CoA dehydrogenase (3-oxooctanoyl-CoA)	3.41
t_0041	3-hydroxyacyl-CoA dehydrogenase (3-oxodecanoyl-CoA)	3.41
t_0059	acetyl-CoA C-acyltransferase (lauroyl-CoA)	3.41
r_2285	acetyl-CoA C-acyltransferase (hexanoyl-CoA)	3.40
t_0055	acetyl-CoA C-acyltransferase (decanoyl-CoA)	3.32
t_0065	acyl-CoA oxidase (decanoyl-CoA)	3.17
t_0227	fatty-acid--CoA ligase (oleate), endoplasmic reticulum membrane	3.03
r_1660	ATP diffusion	2.95
t_0057	acetyl-CoA C-acyltransferase (hexanoyl-CoA)	2.93
t_0054	acetyl-CoA C-acyltransferase (butanoyl-CoA)	2.93
t_0349	PE (1-16:0, 2-18:2) diacylglycerol (1-16:0, 2-18:1) acyltransferase, endoplasmic reticulum membrane	2.91
t_1088	PC (1-18:1, 2-18:2) diacylglycerol (1-18:1, 2-18:1) acyltransferase, endoplasmic reticulum membrane	2.80
t_0067	acyl-CoA oxidase (lauroyl-CoA)	2.73

r_0255	catalase	2.72
r_0191	allophanate hydrolase	2.70
r_0804	nucleoside triphosphatase	2.68
r_1683	choline exchange	2.66
r_0142	adenosine kinase	2.66
r_0144	adenosylhomocysteinase	2.66
r_0726	methionine adenosyltransferase	2.66
r_0727	methionine synthase	2.66
r_3851	phosphatidyl-L-serine (1-18:0, 2-18:1) transport, ER membrane-vacuolar membrane	2.65
t_0763	phosphatidylethanolamine (1-18:0, 2-18:1) transport, ER membrane-vacuolar membrane	2.65
t_0764	phosphatidylethanolamine (1-18:0, 2-18:2) transport, ER membrane-vacuolar membrane	2.65
t_0782	phosphatidyl-L-serine (1-18:0, 2-18:2) transport, ER membrane-vacuolar membrane	2.65
r_3849	phosphatidyl-L-serine (1-16:0, 2-18:1) transport, ER membrane-vacuolar membrane	2.65
t_0760	phosphatidylethanolamine (1-16:0, 2-18:1) transport, ER membrane-vacuolar membrane	2.65
t_0761	phosphatidylethanolamine (1-16:0, 2-18:2) transport, ER membrane-vacuolar membrane	2.65
t_0779	phosphatidyl-L-serine (1-16:0, 2-18:2) transport, ER membrane-vacuolar membrane	2.65
t_0369	PE methyltransferase (1-18:0, 2-18:1), endoplasmic reticulum membrane	2.65
t_0435	phosphatidyl-N,N-dimethylethanolamine methyltransferase (1-18:0, 2-18:1), endoplasmic reticulum membrane	2.65
t_0441	phosphatidyl-N-methylethanolamine methyltransferase (1-18:0, 2-18:1), endoplasmic reticulum membrane	2.65
t_0370	PE methyltransferase (1-18:0, 2-18:2), endoplasmic reticulum membrane	2.65
t_0436	phosphatidyl-N,N-dimethylethanolamine methyltransferase (1-18:0, 2-18:2), endoplasmic reticulum membrane	2.65
t_0442	phosphatidyl-N-methylethanolamine methyltransferase (1-18:0, 2-18:2), endoplasmic reticulum membrane	2.65
t_0367	PE methyltransferase (1-16:0, 2-18:2), endoplasmic reticulum membrane	2.65
t_0433	phosphatidyl-N,N-dimethylethanolamine methyltransferase (1-16:0, 2-18:2), endoplasmic reticulum membrane	2.65
t_0439	phosphatidyl-N-methylethanolamine methyltransferase (1-16:0, 2-18:2), endoplasmic reticulum membrane	2.65
t_0366	PE methyltransferase (1-16:0, 2-18:1), endoplasmic reticulum membrane	2.65
	Restrained	
r_3751	1-acylglycerophosphoethanolamine (16:0) transport, ER membrane-lipid particle	-13.94
r_1811	glycine transport	-13.09
r_3652	ADP transport, cytoplasm-vacuolar membrane	-11.94
r_0664	isoleucine transaminase	-6.72
r_1827	H+ diffusion	-4.63
t_1114	PE (1-18:1, 2-18:2) diacylglycerol (1-16:0, 2-18:1) acyltransferase, endoplasmic reticulum membrane	-4.43
r_3533	NAD transport, cytoplasm-ER membrane	-4.02
t_0351	PE (1-16:0, 2-18:2) diacylglycerol (1-16:0, 2-18:2) acyltransferase, endoplasmic reticulum membrane	-1.49
r_3538	CTP transport, cytoplasm-ER membrane	-1.40
r_1184	L-arganine transport	-1.36
y200033	dCTP:uridine 5'-phosphotransferase	-1.35
t_0749	phosphatidylethanolamine (1-16:0, 2-18:2) transport, ER membrane-lipid particle	-1.21

Different from ATS, the Parental strain presented a small relaxation of its metabolism in the presence of acetic acid (Table 13). Nevertheless, many reactions related to the metabolism of nitrogenous compounds, including the urea cycle, similar to ATS, were relaxed (Table 17). Another similarity was the presence of lipid synthesis and oxidation in both relaxed and restrained reactions (Table 17), corroborating the importance of modulating these pathways to cope with the presence of acetic acid.

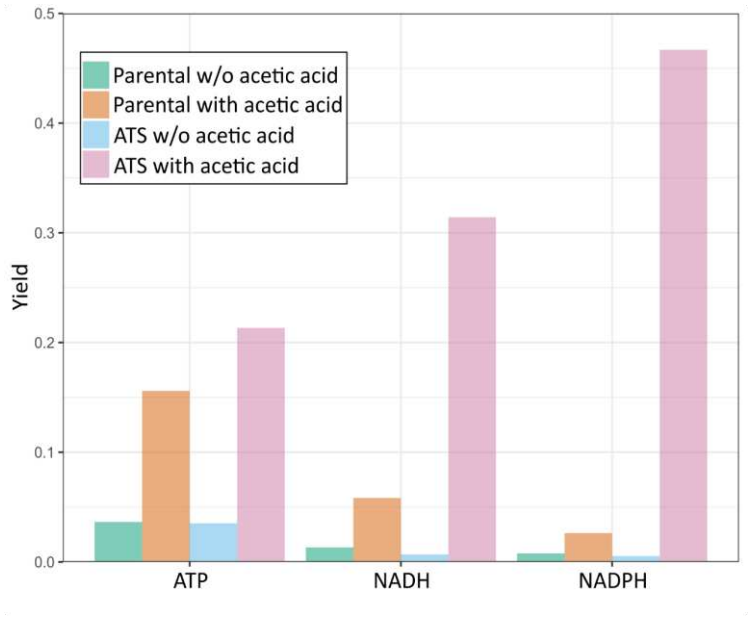
Table 17 – Reactions on *papla*-GEM with differential ranges on *ecpapla*-GEM for *P. laurentii* Parental strain in the presence vs. absence of acetic acid.

Reaction ID	Reaction Name	Log_2(FC)
	Relaxed	
r_0023	2-isopropylmalate hydratase	40.47
t_1083	PC (1-18:1, 2-18:1) diacylglycerol (1-16:0, 2-18:2) acyltransferase, endoplasmic reticulum membrane	38.52
t_0064	acyl-CoA oxidase (butanoyl-CoA)	9.04
r_0503	glycine hydroxymethyltransferase	5.80
r_4039	succinyl-CoA:acetate CoA transferase	4.40
r_3573	octadecanoate (n-C18:0) transport, cytoplasm-lipid particle	3.11
r_3533	NAD transport, cytoplasm-ER membrane	2.88
r_3525	H+ transport, cytoplasm-ER membrane	2.55
r_0502	glycine hydroxymethyltransferase	2.48
r_0218	aspartate transaminase	2.27
r_0105	acetyl-CoA C-acyltransferase (lauroyl-CoA)	1.95
r_0090	6-phosphofructo-2-kinase	1.76
r_0150	adenylate kinase (GTP)	1.76
r_0207	argininosuccinate lyase	1.60
t_0057	acetyl-CoA C-acyltransferase (hexanoyl-CoA)	1.52
r_1660	ATP diffusion	1.50
r_0361	dolichyl-phosphate D-mannosyltransferase	1.43
r_0722	mannose-1-phosphate guanylyltransferase	1.43
r_0902	phosphomannomutase	1.43
r_1063	tyrosine transaminase	1.40
t_0231	fatty-acid--CoA ligase (stearate), lipid particle	1.40
r_3513	oleate transport, cytoplasm-ER membrane	1.36
t_0225	fatty-acid--CoA ligase (palmitate), endoplasmic reticulum membrane	1.22
t_0070	acyl-CoA oxidase (hexanoyl-CoA)	1.21
	Restrained	
r_3751	1-acylglycerophosphoethanolamine (16:0) transport, ER membrane-lipid particle	-14.25
t_0041	3-hydroxyacyl-CoA dehydrogenase (3-oxodecanoyl-CoA)	-13.96
t_0049	3-hydroxyacyl-CoA dehydrogenase (3-oxooctanoyl-CoA)	-13.96
r_3652	ADP transport, cytoplasm-vacuolar membrane	-12.59
r_0018	2-aminoadipate transaminase	-5.63
t_0874	linoleate transport, cytoplasm-endoplasmic reticulum membrane	-5.57
t_1114	PE (1-18:1, 2-18:2) diacylglycerol (1-16:0, 2-18:1) acyltransferase, endoplasmic reticulum membrane	-4.92
y300065	L-Serine:pyruvate aminotransferase	-3.12
t_0351	PE (1-16:0, 2-18:2) diacylglycerol (1-16:0, 2-18:2) acyltransferase, endoplasmic reticulum membrane	-2.41
r_3574	octadecenoate (n-C18:1) transport, cytoplasm-lipid particle	-2.36
r_2125	coenzyme A: cytoplasm to LP	-2.23

t_1118	PE (1-18:1, 2-18:2) diacylglycerol (1-16:0, 2-18:2) acyltransferase, endoplasmic reticulum membrane	-2.13
t_1088	PC (1-18:1, 2-18:2) diacylglycerol (1-18:1, 2-18:1) acyltransferase, endoplasmic reticulum membrane	-1.48
t_0233	fatty-acid--CoA ligase (linoleate), lipid particle	-1.44
t_0052	acetyl-CoA C-acyltransferase (acetyl-CoA)	-1.41

In order to compare the different availability of cofactors in the absence and presence of acetic acid, as well as between strains, we conducted an RS analysis using the condition/strain-adjusted *ecpapla*-GEM. We conducted 10,000 random samples and calculated the ATP, NADH, and NADPH yield as the sum of median fluxes in the reactions that consumed or produced these cofactors divided by the total carbon uptake rate (Figure 8). The ATP, NADH, and NADPH yields were almost the same for both strains in the absence of acetic acid; however, ATS presented slightly lower NADH and NADPH yields, possibly due to its lower xylose uptake and biomass composition. In the presence of acetic acid, both strains presented a higher predicter yield of cofactors. Nevertheless, ATS presented superior yields, especially for NADH and NADPH. This remarkable increase in the pool of reduced cofactors might be related to the repression of reactions with high demand, such as biosynthetic pathways observed in the gene expression analysis, PPI networks, and FVA, mainly reflected in the model by the biomass composition and uptake differences. These cofactors are of great importance to cope with stressful conditions in the cell, such as the detoxification of ROS, and might be one of the major mechanisms underlying the tolerance of ATS compared to the Parental strain in the presence of acetic acid.

Figure 8 – ATP, NADH, and NADPH yield for *P. laurentii* ATS and Parental strain in the presence and absence of acetic acid based on the median fluxes from 10,000 random samplings using ec*papla*-GEM constrained for each strain and condition (biomass composition, growth, and carbon uptake).



4.5. Discussion

Lignocellulosic biomasses are abundant feedstocks that can be used as carbon sources for oil production by oleaginous yeasts in a circular economy context. However, the pretreatment and hydrolysis of these biomasses generate growth inhibitors, such as acetic acid, furfural, HMF, formic acid, and phenolic compounds (JIN et al., 2015; KUMAR; SINGH; KORSTAD, 2017). It is important to point out that even after the detoxification step, acetic acid remains in concentrations that impair yeast growth and production of biobased-products (CHANDEL; DA SILVA; SINGH, 2013; JÖNSSON; MARTÍN, 2016). Thus, the development of yeast strains tolerant to acetic acid and the comprehension of their adaptive responses to acetic acid stress are crucial for boosting the yeast oil production from lignocellulosic hydrolysate-based biorefineries. Previously, we used ALE to select an acetic-acid tolerant strain of *P. laurentii,* ATS. Besides the physiological characterization, we sequenced and compared its genome with the Parental strain and identified mutations that might contribute to the tolerance phenotype. In the present study, to deepen our understanding regarding the tolerance mechanisms acquired by ALE, we conducted transcriptome and genome-scale metabolic modeling analyses for

both strains using samples and data collected from xylose chemostat cultures conducted in the presence and absence of acetic acid.

We found that the presence of acetic acid led to global changes in the expression profile for both strains. Most genes had their expression downregulated under acetic acid stress. Similar to our results, major downregulation of gene expression was also observed for different yeasts, including *S. cerevisiae* (DONG et al., 2017; GENG; ZHANG; SHI, 2017; LEE et al., 2015; LI et al., 2020a), *Zygosaccharomyces bailli* (ANTUNES; PALMA; SÁ-CORREIA, 2018), *Issatchenkia orientalis* (LI et al., 2020b, 2022), and *Kluyveromyces marxianus* (LI et al., 2023). It is important to point out that these studies describe that the stress response is multifactorial and has complex mechanisms, differentially regulating various pathways and cell functions, which is also in line with our current results.

When *P. laurentii* ATS and Parental strain were stressed with acetic acid, transcriptional and translational processes were repressed and formed dense and integrated PPI networks. Interestingly, these processes were also repressed in ATS in the absence of acetic acid (Tables 2-5; Figures 4 and 5). Consistent with our findings, *I. orientalis* (Li et al., 2020) and *K. marxianus* (LI et al., 2023) exposed to acetic acid showed a downregulation of genes and their PPI network comprising mainly ribosome biogenesis and translational processes. In *Z. bailli*, the activity and binding of transcriptional factors are the main functions repressed under acetic acid stress (ANTUNES; PALMA; SÁ-CORREIA, 2018). Meanwhile, in *S. cerevisiae*, both transcriptional regulation and ribosomal-associated genes are repressed (DONG et al., 2017). The repression of these processes seems to be associated with the stress generated by the accumulation of acetate and reduction of the intracellular pH, which represses metabolic activity and affects protein folding (GUAN; LIU, 2020; PALMA; GUERREIRO; SÁ-CORREIA, 2018).

Additionally, transcription and translation, including the biosynthesis of purines/pyrimidines and amino acids, require high amounts of ATP and reducing power [NAD(P)H]. Thus, the repression of these processes under acetic acid exposure might also be related to saving energy and cofactors for coping with the stress imposed by acetic acid. Interestingly, the aforementioned repression was more evident in the tolerant strain of *P. laurentii* than in the Parental strain, both in the presence and absence of acetic acid.

This difference between the tolerant and Parental strains might be related to the improved tolerance to acetic acid displayed by ATS (Tables 2 and 5; Figures 4 and 5).

The cell cycle can also be affected by acetic acid, leading to autophagy and programmed cell death in more severe scenarios. In more stressed cells (150 mM of acetic acid; batch cultures) of *S. cerevisiae*, DNA repair mechanisms, sexual reproduction, and meiosis are induced (DONG et al., 2017), while in less stressed cells, the meiosis is repressed (LI et al., 2020a). Herein, the Parental (sensitive) strain presented a behavior closer to cells under severe stress since it uniquely induced proteins related to DNA mismatch repair (Table 7) and the PPI network of upregulated genes in the presence of acetic acid was enriched for yeast meiosis (Figure 5F). In contrast, the ATS induced genes related to cell cycle arrest and the negative regulation of autophagy (Table 2) and did not induce DNA repair and meiosis, suggesting a behavior more similar to less stressed cells. Importantly, we previously detected genomic variants in the coding and regulatory regions of genes related to cell cycle, autophagy, and sexual differentiation that emerged during the ALE (ALMEIDA et al., 2023). Consistent with these results, we demonstrated here that these functions were also affected in the transcriptional level in ATS.

In pH below 4.76, acetic acid is mainly in its undissociated form and can enter the cell via simple diffusion or through channels such as aquaglyceroporin (MOLLAPOUR; PIPER, 2007). Upon entering the cell, it dissociates, releasing acetate and protons (H^+), and two main challenges are imposed: the reduction of the cytoplasmic pH promoted by H^+ and the accumulation of acetate ions. The pH reduction leads to protein misfolding and decreased synthesis of nucleotides, enzyme activity, and metabolic rates. Acetate accumulation triggers oxidative stress, lipid oxidation, and protein aggregation and modifies the activity of different organelles, including the peroxisome, vacuole, and mitochondria (GUARAGNELLA; BETTIGA, 2021; LI et al., 2024). To cope with this reduction and maintain pH homeostasis, yeast cells can pump protons out of the cell or to the vacuole via ATPase activity (GUARAGNELLA; BETTIGA, 2021; LI et al., 2024; PALMA; GUERREIRO; SÁ-CORREIA, 2018). In the presence of acetic acid, ATS uniquely repressed the synthesis of a vacuolar ATPase (Table 6), restraining H+ diffusion to the extracellular environment (Table 16), despite presenting more metabolic plasticity in this condition (Figure 6), which highlights the importance of this process. This result

indicates that the proton transport to the vacuole might have a minor role compared to plasma membrane ATPases in *P. laurentii*.

Another strategy to maintain pH homeostasis is the regulation of the metabolism of nitrogenous compounds, such as amino acids and purines/pyrimidines, to accumulate ammonia and polyamines and consume H^+ via deamination and decarboxylation reactions, raising the intracellular pH and neutralizing cytoplasmic protons (GUAN; LIU, 2020). Herein, these cell processes were highly regulated in both strains in the presence of acetic acid. The top-downregulated genes of both strains included peptide and amino acid transporters (Tables 3, 6, and 7) in the cell membrane and vacuole. The tolerant strain (ATS) also repressed these processes under unstressed conditions (Tables 2 and 5). The RAS analysis corroborated the repression of transport reactions for many amino acids and ammonia compared to the Parental strain (Tables 9 and 10). Regarding the amino acid metabolism, ATS induced arginase (Table 11), which generates urea and ornithine from arginine and might contribute to increases in intracellular pH. ATS also presented more flexibility in the metabolism of amino acids and other nitrogenous compounds, including the urea cycle (Table 16). This regulation favors the ammonia accumulation and the activity of proton-consuming reactions in ATS, which would increase intracellular pH, and might be one of the main mechanisms for acetic acid tolerance in *P. laurentii*.

Furthermore, yeast cells can export acetate to the external environment or catabolize it. The first strategy is well described for *S. cerevisiae* and involves multidrug MFS transporters (ZHANG; NIJLAND; DRIESSEN, 2022), while the highly acid-tolerant yeast *Z. baillii* mainly employs the second strategy (PALMA; GUERREIRO; SÁ-CORREIA, 2018). Since acetic acid was not found in the culture media during chemostat cultivations (ALMEIDA et al., 2023), acetate catabolism might be important to deal with this stress. Thus, a higher capacity to metabolize acetate might be directly related to acetic acid tolerance in *P. laurentii.* Consistently, we found that the tolerant strain (ATS) induced an aquaporin-like protein (uptake of acetate by facilitated diffusion) and two malate dehydrogenases (Tables 2 and 6); ATS also presented more metabolic flexibility in glycolysis, gluconeogenesis, oxidative phosphorylation, and glyoxylate metabolism (Figure 7).

The metabolism of acetic acid and its accumulation also promote the generation of toxic compounds, including ROS. Notably, in the presence of acetic acid, the tolerant

strain (ATS) of *P. laurentii* presented higher levels of ROS during the chemostat cultivations (ALMEIDA et al., 2023). In accordance with this result, ATS induced different genes, reactions, and functions related to cell detoxification, including efflux proteins, uptake of copper and iron, heme transport, NAD(P)H dehydrogenases, eisosome formation, oxidoreductases, epoxide hydrolases, ornithine metabolism, and glutathione metabolism (Tables 2, 6, and 10), and relaxed catalase (Table 16). Importantly, the genome of ATS presented mutations in many genes and their regulatory regions related to these functions (ALMEIDA et al., 2023), underscoring their importance for acetic acid tolerance in *P. laurentii.*

Many genes and reactions related to unknown transports, including MFS transporters, were induced in the presence of acetic acid. Different ABC and MFS transporters were also affected at the genomic level during the ALE (ALMEIDA et al., 2023). Although the specific function of these transporters has not been elucidated in *P. laurentii*, they might be related to the export of byproducts and toxic compounds related to acetic acid stress, as well as the uptake of important nutrients and transport of organic acids in these conditions.

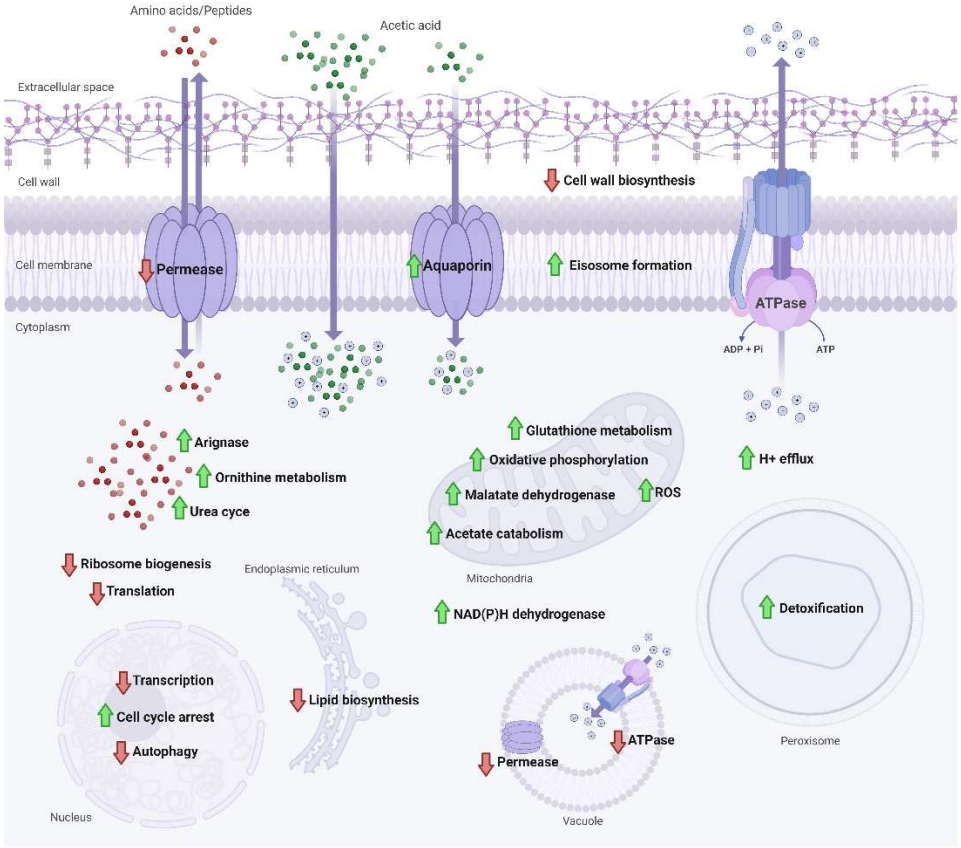
The cell envelope is another important target during the stress response to weak acids (GUAN; LIU, 2020; LI et al., 2024). In our previous study, we found that ATS presented a fatty acid profile that would allow more acetic acid diffusion, mutated genes related to cell-wall biogenesis, and preserved its cell morphology (ALMEIDA et al., 2023). Herein, we showed that ATS repressed genes and reactions related to the biosynthesis of cell wall components (Tables 3, 6, and 10). Besides, many reactions related to lipid metabolism, including sterol biosynthesis, were relaxed (Figure 7; Tables 14 and 16). In contrast to ATS, the Parental strain presented major alterations in cell morphology under acetic acid stress (ALMEIDA et al., 2023), which was consistent with the induction of many genes involved in the cell wall biogenesis and modification (Tables 3 and 7). Altogether, these results indicated that targeting the cell envelope might be more important for more sensitive *P. laurentii* cells under severe stress (Parental strain) than for those that are more tolerant and less stressed (ATS).

The response and tolerance mechanisms described above depend on the cell's pool of ATP and reducing power. Hence, an increased availability of these molecules indicates a higher capacity to cope with stressful conditions and thrive. To increase the availability

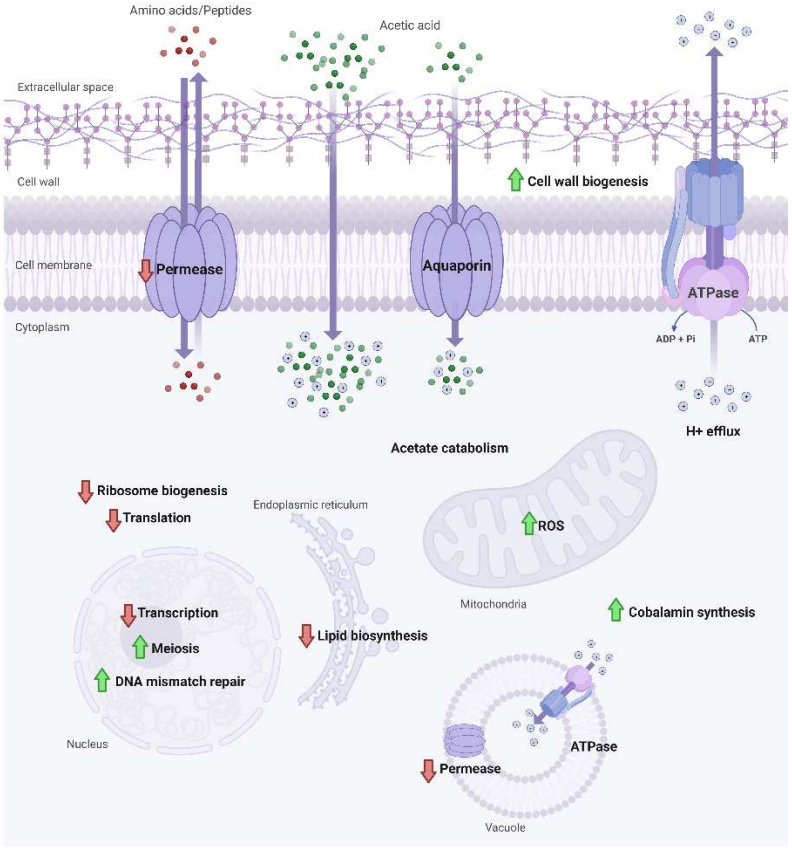
of these molecules for stress response, cells can save them by repressing consuming processes (i.e., transcription, translation, and biosynthesis of complex molecules) and increase their carbon uptake and catabolism. Our results indicated that the tolerant strain of *P. laurentii* used these two strategies, which is in line with its far superior theoretical yield of ATP, NADH, and NADPH in the presence of acetic acid compared to the Parental (sensitive) strain (Figure 8).

The main response mechanisms of *P. laurentii* to acetic acid stress described here, combined with our previous data (ALMEIDA et al., 2023), are summarized in Figure 9. To counteract the decreases in intracellular pH, ATS induced the efflux of protons and the metabolism of nitrogenous compounds, including arginase, the urea cycle and ornithine. The tolerant strain also induced the catabolism and uptake of acetate and various reactions related to detoxification, such as glutathione metabolism and NAD(P)H dehydrogenase. Since the stress response requires energy and reducing power, ATS repressed various processes that are expensive for the cell in terms of cofactors, including the biosynthesis of lipids and cell wall components, ribosome biogenesis, transcription, and translation, which was supported by its remarkably higher cofactor pool in the metabolic modeling analysis.

Figure 9 - Overview of the response mechanisms of *P. laurentii* to acetic acid stress. The red arrows turned down indicate repressed/decreased cellular functions, pathways, metabolites or genes; the green arrows turned up indicate induced/increased cellular functions, pathways, metabolites or genes. ROS: reactive oxygen species.



Tolerant strain



Sensitive strain

4.6. Conclusion

In the present study, we combined transcriptomics and genome-scale metabolic modeling to improve our understanding about the adaptive responses of *P. laurentii* to acetic acid stress. Our findings, based on transcriptomic and fluxomic analysis, allowed us to expand the current understanding about the acetic acid targets, as well as response mechanisms of *P. laurentii.* Acetic acid stress promoted global changes in both strains and most repressed genes were related to transcriptional and translational processes. The sensitive strain induced DNA mismatch repair mechanisms and meiosis, while the tolerant strain negatively regulated autophagy and the cell cycle. Moreover, the tolerant strain induced processes responsible for increasing the intracellular pH, including the metabolism of nitrogenous compounds and proton efflux, and the detoxification of toxic compounds. Notably, ATS presented a remarkable NAD(P)H pool in the metabolic modeling analysis, which might support the reducing power required by these tolerance mechanisms. Other omics such as proteomics and metabolomics can be further applied and combined with our current data to deepen this knowledge and uncover novel mechanisms. Importantly, we found genes and pathways involved in the adaptive responses to acetic acid stress that can be future metabolic engineering targets to improve the tolerance of *P. laurentii* to weak acids, boosting its application in lignocellulosic-based biorefineries.

4.7. Data availability

Sequencing data is available in Genbank (BioProject: PRJNA879120). The scripts for reconstruction and simulations, the models, and the raw data generated here are available in the GitHub repository at https://github.com/LabFisUFV/PlaurentiiAceticAcidStress.

4.8. Acknowledgments

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) (Finance Code 001), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) (Finance Code APQ-00326-21), and Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq) [Finance Code 140538/2021-6].

4.9. Conflict of interest

The authors declare no competing interests.

4.10. References

ALMEIDA, E. L. M. et al. New *Papiliotrema laurentii* UFV-1 strains with improved acetic acid tolerance selected by adaptive laboratory evolution. **Fungal Genetics and Biology**, v. 164, p. 103765, 2023.

ANDREWS, S. **FastQC: A Quality Control Tool for High Throughput Sequence Data.** Babraham Institute, 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

ANTUNES, M.; PALMA, M.; SÁ-CORREIA, I. Transcriptional profiling of *Zygosaccharomyces bailii* early response to acetic acid or copper stress mediated by ZbHaa1. **Scientific Reports**, v. 8, n. 1, p. 14122, 2018.

BLIGHE, K.; RANA, S.; LEWIS, M. **EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling**. 2023. Disponível em: <https://github.com/kevinblighe/EnhancedVolcano>

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014.

CHANDEL, A. K.; DA SILVA, S. S.; SINGH, O. V. Detoxification of Lignocellulose Hydrolysates: Biochemical and Metabolic Engineering Toward White Biotechnology. **BioEnergy Research**, v. 6, n. 1, p. 388–401, 2013.

CHEN, Y. et al. Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO Toolbox 3.0. **Nature Protocols**, 2024.

DI FILIPPO, M. et al. INTEGRATE: Model-based multi-omics data integration to characterize multi-level metabolic regulation. **PLOS Computational Biology**, v. 18, n. 2, p. e1009337, 2022.

DONG, Y. et al. RNA-Seq-based transcriptomic and metabolomic analysis reveal stress responses and programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. **Scientific Reports**, v. 7, n. 1, p. 42659, 2017.

GALLEGO-GARCÍA, M. et al. Challenges and prospects of yeast-based microbial oil production within a biorefinery concept. **Microbial Cell Factories**, v. 22, n. 1, p. 246, 2023.

GENG, P.; ZHANG, L.; SHI, G. Y. Omics analysis of acetic acid tolerance in *Saccharomyces cerevisiae*. **World Journal of Microbiology and Biotechnology**, v. 33, n. 5, p. 94, 2017.

GUAN, N.; LIU, L. Microbial response to acid stress: mechanisms and applications. **Applied Microbiology and Biotechnology**, v. 104, n. 1, p. 51–65, 2020.

GUARAGNELLA, N.; BETTIGA, M. Acetic acid stress in budding yeast: From molecular mechanisms to applications. **Yeast**, v. 38, n. 7, p. 391–400, 2021.

HEIRENDT, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. **Nature Protocols**, v. 14, n. 3, p. 639–702, 2019.

JIN, M. et al. Microbial lipid-based lignocellulosic biorefinery: feasibility and challenges. **Trends in Biotechnology**, v. 33, n. 1, p. 43–54, 2015.

JÖNSSON, L. J.; MARTÍN, C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. **Bioresource Technology**, v. 199, p. 103–112, 2016.

KUMAR, D.; SINGH, B.; KORSTAD, J. Utilization of lignocellulosic biomass by oleaginous yeast and bacteria for production of biodiesel and renewable diesel. **Renewable and Sustainable Energy Reviews**, v. 73, p. 654–671, 2017.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, 2012.

LEE, Y. et al. Transcriptome analysis of acetic-acid-treated yeast cells identifies a large set of genes whose overexpression or deletion enhances acetic acid tolerance. **Applied Microbiology and Biotechnology**, v. 99, n. 15, p. 6391–6403, 2015.

LEWIS, N. E. et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. **Molecular Systems Biology**, v. 6, n. 1, p. 390, 2010.

LI, B. et al. The response mechanisms of industrial *Saccharomyces cerevisiae* to acetic acid and formic acid during mixed glucose and xylose fermentation. **Process Biochemistry**, v. 91, p. 319–329, 2020a.

LI, M. et al. General mechanisms of weak acid-tolerance and current strategies for the development of tolerant yeasts. **World Journal of Microbiology and Biotechnology**, v. 40, n. 2, p. 49, 2024.

LI, Y. et al. Integrated transcriptomic and proteomic analysis of the acetic acid stress in *Issatchenkia orientalis*. **Journal of Food Biochemistry**, v. 44, n. 6, 2020b.

LI, Y. et al. Metabolic changes of *Issatchenkia orientalis* under acetic acid stress by transcriptome profile using RNA-sequencing. **International Microbiology**, v. 25, n. 3, p. 417–426, 2022.

LI, Y. et al. Transcriptomic analysis reveals hub genes and pathways in response to acetic acid stress in *Kluyveromyces marxianus* during high-temperature ethanol fermentation. **Stress Biology**, v. 3, n. 1, p. 26, 26 2023.

LIAO, Y.; SMYTH, G. K.; SHI, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923–930, 2014.

LONG, C. P.; ANTONIEWICZ, M. R. How adaptive evolution reshapes metabolism to improve fitness: recent advances and future outlook. **Current Opinion in Chemical Engineering**, v. 22, p. 209–215, 2018.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 1–21, 2014.

MOLLAPOUR, M.; PIPER, P. W. Hog1 Mitogen-Activated Protein Kinase Phosphorylation Targets the Yeast Fps1 Aquaglyceroporin for Endocytosis, Thereby Rendering Cells Resistant to Acetic Acid. **Molecular and Cellular Biology**, v. 27, n. 18, p. 6446–6456, 2007.

NIELSEN, J. Systems Biology of Metabolism. **Annual Review of Biochemistry**, v. 86, n. 1, p. 245–275, 2017.

NURWONO, G. et al. Sustainable metabolic engineering requires a perfect trifecta. **Current Opinion in Biotechnology**, v. 83, p. 102983, 2023.

PALMA, M.; GUERREIRO, J. F.; SÁ-CORREIA, I. Adaptive Response and Tolerance to Acetic Acid in Saccharomyces cerevisiae and *Zygosaccharomyces bailii*: A Physiological Genomics Perspective. **Frontiers in Microbiology**, v. 9, p. 274, 2018.

SANDBERG, T. E. et al. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. **Metabolic Engineering**, v. 56, p. 1–16, 2019.

SITEPU, I. et al. Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. **Journal of Industrial Microbiology and Biotechnology**, v. 41, n. 7, p. 1061–1070, 2014.

SZKLARCZYK, D. et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. **Nucleic Acids Research**, v. 51, n. D1, p. D638–D646, 2023.

VENTORIM, R. Z. et al. Genome-scale metabolic model of oleaginous yeast *Papiliotrema laurentii*. **Biochemical Engineering Journal**, v. 180, p. 108353, 2022.

VIEIRA, N. M. et al. Isolation of a new *Papiliotrema laurentii* strain that displays the capacity to achieve high lipid content from xylose. **3 Biotech**, v. 10, n. 9, p. 382, 2020.

WANG, H. et al. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. **PLOS Computational Biology**, v. 14, n. 10, p. e1006541, 2018.

WU, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. **The Innovation**, v. 2, n. 3, p. 100141, 2021.

ZHANG, X.; NIJLAND, J. G.; DRIESSEN, A. J. M. Combined roles of exporters in acetic acid tolerance in *Saccharomyces cerevisiae*. **Biotechnology for Biofuels and Bioproducts**, v. 15, n. 1, p. 67, 2022.

GENERAL CONCLUSIONS

Understanding the stress response and physiology of oleaginous yeasts is crucial for developing robust strains and making feasible the lipid production by oleaginous yeasts in lignocellulosic biorefineries. Systems biology approaches are suitable for holistically studying metabolism and the underlying mechanisms and complex interactions in stressed conditions. Herein, we first presented a literature review on the reconstruction and validations of GEMs of non-*Saccharomyces* yeasts, underscoring the physiological insights, as well as the identification of both metabolic engineering and drug targets. In the second chapter, we proposed *lista*-GEM, the first genome-scale metabolic model of *L. starkeyi*. We showed that *lista*-GEM successfully captured the growth and lipid-producing phenotype of *L. starkeyi* and is a useful platform for *in silico* metabolic engineering of this yeast. Furthermore, we revised the current knowledge about the physiology, genetics, ecology, and biotechnological applications of *P. laurentii*. In the final chapter, we combined transcriptome and genome-scale metabolic modeling to deepen our understanding regarding the targets of acetic acid stress, as well as the adaptive responses in *P. laurentii*. We described genes and pathways that can be used in future genetic and metabolic engineering rounds to improve the tolerance of this yeast to weak acids and boost its application in lignocellulosic-based biorefineries. Overall, we demonstrated that the integration of systems biology approaches, including GEMs and transcriptomics, can help us understand the metabolism and stress responses of oleaginous yeasts and determine suitable targets for metabolic engineering.