Enzyme-constrained genome-scale model of Yarrowia lipolytica predicts growth-phase specific metabolic engineering targets

Applied Microbiology and Biotechnology (2026) 110:123
https://doi.org/10.1007/s00253-026-13791-4

RESEARCH

Enzyme-constrained genome-scale model of *Yarrowia lipolytica* predicts growth-phase specific metabolic engineering targets

Juliano Sabedotti De Biaggi[1] · Young-Kyoung Park[2,3] · Eduard J. Kerkhoven[4,5] · Rodrigo Ledesma-Amaro[3,6,7] · Petri-Jaan Lahtvee[1]

Received: 10 November 2025 / Revised: 3 March 2026 / Accepted: 8 March 2026

Abstract

The oleaginous yeast *Yarrowia lipolytica* has been gaining increasing importance as an industrial biotech platform, supported by several available metabolic engineering tools. Genome-scale models (GEMs) are relevant to the iterative improvement of this yeast, and their predictive ability is enhanced by enzymatic activity constraints (ecGEMs). Although the newest tools for ecGEM reconstruction use deep learning to expand the coverage of these constraints, this approach has not yet been applied to *Y. lipolytica* models. This paper describes the reconstruction of an ecGEM of *Y. lipolytica* (*eciYali5-GEM*) and its application in predicting metabolic engineering targets for enhanced lipid and carotenoid production, respectively, in this yeast. To achieve this, we constrained a manually curated *Y. lipolytica* model with physiological flux and mass-spectrometry proteomics data collected from distinct growth phases of two engineered strains (producing lipids and carotenoids, respectively) and their parental strain. We found that the enzymatic constraints enable the prediction of growth-phase-specific metabolic engineering targets, a feature not displayed by regular GEMs. Combining these targets with other ecGEM-based insights, we propose two strategies for further metabolic engineering, including the use of inducible promoters for precise, growth-phase-specific expression of targets such as phytoene dehydrogenase for carotenoid production. These targets included genes previously validated elsewhere, as well as novel genes awaiting experimental validation. This model, which is publicly available, can be similarly adapted and used by different metabolic engineering efforts, making it a versatile tool for the development of *Y. lipolytica* as a microbial cell factory.

Key points

- *eciYali5-GEM reconstruction used in silico and in vitro proteomics data.*
- *Enzyme constraints enriched Y. lipolytica ecGEM predictions.*
- *eciYali5-GEM improves metabolic engineering rational design.*

Keywords Systems biology · *Yarrowia lipolytica* · Enzyme-constrained models · GECKO · FSEOF · Flux control coefficients

✉ Petri-Jaan Lahtvee
lahtvee@taltech.ee

[1] Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia

[2] Micalis Institute, Université Paris-Saclay, INRAE, AgroParisTech, Jouy-en-Josas, France

[3] Department of Bioengineering and Imperial College Centre for Synthetic Biology, Imperial College London, London, UK

[4] Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

[5] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

[6] Bezos Centre for Sustainable Protein, Imperial College London, London, UK

[7] UKRI Mission Hub On Microbial Food, Imperial College London, London, UK

Published online: 24 March 2026

Introduction

Yarrowia lipolytica is arguably the most consolidated oleaginous yeast for bioprocessing. This yeast naturally produces lipids, organic acids, and polyols. For instance, wild-type *Y. lipolytica* strains have been shown to accumulate lipid contents of 78% (Tomás-Pejó et al. 2023), produce 107 $g{\cdot}L^{-1}$ of citrate (Rymowicz et al. 2010), and yield 199 $g{\cdot}L^{-1}$ of erythritol (Rakicka et al. 2017). Notably, this reported lipid accumulation is compatible with economically viable key performance indicators (Kumar et al. 2023). *Y. lipolytica* also holds Generally Recognized as Safe (GRAS) status (Groenewald et al. 2014), which expands its applicability to a broader range of products, including those intended for human consumption. A recent study identified companies that have established or are scaling up commercial *Y. lipolytica*-derived products (Park and Ledesma-Amaro 2023), indicating that the industrial use of this yeast is already a reality. A common feature among most of these products (i.e., polyunsaturated fatty acids, recombinant enzymes, and carotenoids) is their derivation from genetically modified *Y. lipolytica*. This is facilitated by the ready availability of metabolic engineering tools for this yeast (Cao et al. 2023), potentially enabling multiple cycles of targeted genomic engineering.

The availability of metabolic engineering tools makes *Y. lipolytica* a suitable platform for strain development in Design-Build-Test-Learn (DBTL) cycles —an engineering approach to achieve optimal solutions. Implementing this approach requires standardized and modular DNA parts and robust genome-editing capabilities (Chao et al. 2017), both of which are established for *Y. lipolytica* (Larroude et al. 2019; Holkenbrink et al. 2018). Genome-scale models (GEMs), in turn, can also be employed within DBTL cycles for metabolic network optimization (Design) (Chao et al. 2017) and for the integration of cultivation and omics data (Learn) (Czajka et al. 2021; Vijayakumar et al. 2020). The integration of enzyme kinetics (derived from k_{cat} values) and abundance constraints into GEMs, yielding enzyme-constrained GEMs (ecGEMs), improves their predictive abilities. This integration enables the description of phenotypes previously considered unfeasible (Sánchez et al. 2017). These enhanced predictive abilities also make ecGEMs powerful tools for guiding metabolic engineering efforts. For example, an ecGEM of *Saccharomyces cerevisiae* was employed to achieve a 70-fold increase in heme production (Ishchuk et al. 2022).

To date, six original *Y. lipolytica* GEM (Loira et al. 2012; Kavščček et al. 2015; Kerkhoven et al. 2016; Wei et al. 2017; Mishra et al. 2018; Y. Guo et al. 2022) and one ecGEM (Domenzain et al. 2022) reconstructions have been published (excluding updates). However, the latter does not incorporate the latest features of ecGEM reconstruction pipelines, such as GECKO 3 (Chen et al. 2024), which allows for the prediction of k_{cat} values through deep learning and further improves ecGEM predictions (Li et al. 2022). Moreover, to our knowledge, the existing *Y. lipolytica* ecGEM has not yet been used for metabolic engineering efforts.

The goal of this work was to reconstruct an updated ecGEM of *Y. lipolytica* with expanded deep-learning k_{cat} constraints. This refined model was then applied to analyze strains engineered to produce industrially relevant molecules, namely lipids and carotenoids. The newly generated ecGEM, when constrained with bioreactor flux and proteomics data, provided additional insights into the metabolism of these strains, complementing a more traditional proteomics approach. Furthermore, these models were used to predict additional targets for metabolic engineering. This predictive application represented the start of a second round of a DBTL cycle. The outputs of these predictions, which appeared to be growth-phase-specific, enabled the rational conceptualization of subsequent steps for improving the genotype of these strains, thereby endorsing the value of ecGEMs as a tool for DBTL cycles.

Materials and methods

Y. lipolytica strain construction

Detailed strain genotypes and plasmid components are listed in Table 1. *Y. lipolytica* PAR (R. Dulermo et al. 2014), derived from strain W29 Po1d (Barth and Gaillardin 1996) was obtained from RLA Lab (Imperial College of London) strain collection and was used as the parental and control strain. OBE (Lazar et al. 2014), a PAR-derived strain with enhanced lipid accumulation, was obtained from the same collection.

The linearized carotenoid gene expression cassettes by NotI (HMG1-Nat and car-cassette) were sequentially transformed to PAR competent cells by the lithium acetate/DTT method to generate a carotenoid-producing strain (CAR_{LEU-}). The gene expression cassettes were randomly integrated into the genome of *Y. lipolytica*. Transformants were selected on YNBD media containing the appropriate amino acids for their specific genotype. Integration of gene expression cassettes was verified by colony PCR with Phire Plant Direct PCR master mix (Thermo Fisher, Waltham, USA). A prototroph strain (CAR) was constructed by integrating URA3- or LEU2-expressing fragments.

Table 1 Plasmid and strain table

Plasmid		Description			Reference
HMG1-Nat		JMP62-Nat ex-pTEF-tHMG1-TLip2			Larroude et al. 2018
Car-cassette		GGE-URA3 ex-pTEF-GSS1-TLip2-pTEF-CarB-TLip2-pTEF-CarRP-TLip2			Larroude et al. 2018
Strain	**Product**	**Genotype**	**Auxotrophy**	**Reference ID**	**Reference**
Po1d	WT	Po1d (MatA, leu2-270, ura3-302, xpr2-322, scSUC2, Δaep)	URA-LEU2-	W29 Po1d	Barth and Gaillardin 1996
PAR	WT	Po1d URA3 LEU2	URA + LEU +	JMY2900	Dulermo et al. 2014
CAR_{LEU-}	Carotene	Po1d URA3 ex -pTEF-GSS1-TLip2-pTEF-CarB-TLip2-pTEF-CarRP-TLip2 Nat ex-pTEF-tHMG1-TLip2	URA + LEU-		This study
CAR	Carotene	Po1d URA3 ex -pTEF-GSS1-TLip2-pTEF-CarB-TLip2-pTEF-CarRP-TLip2 Nat ex-pTEF-tHMG1-TLip2 LEU2	URA + LEU +		This study
OBE	Obese (lipid)	Po1d Δpox1-6 Δtgl4 LEU2 ex-pTEF-DGA2-TLip2 URA3 ex-pTEF-GPD1-TLip2	URA + LEU +	JMY3501	Lazar et al. 2014

Media composition

Selection YNBD medium contained 0.17% yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium chloride, 50 mM phosphate buffer (KH_2PO_4–Na_2HPO_4, pH 6.8), and 2% glucose. Complex medium (YPD) contained 2% peptone, 1% yeast extract, and 2% glucose. Cultivation minimal medium contained 2% glycerol as a sole carbon source, $(NH_4)_2SO_4$ 0.5 g/L, KH_2PO_4 3 g/L, and $MgSO_4$ 0.5 g/L, supplemented with trace minerals and vitamins (Verduyn et al. 1992). The resulting C/N ratio of the medium was therefore 86, resulting in nitrogen limiting conditions at the end of the cultivation. Antifoaming agent (Antifoam 204) in an amount of 0.01% (v/v) was added to the cultivation medium prior to wet heat sterilization of the medium components (121 °C, 5 min).

Unless differently stated, all other reagents were purchased from Sigma-Aldrich Co., St. Louis, MO, USA.

Bioreactor cultivation

Inoculums were prepared by transferring a single colony of each strain used herein to 30 mL of sterile YPD medium in a 250-mL conical flask, which were incubated overnight at 30 °C, 200 rpm on an orbital shaker. After the incubation, enough broth was centrifuged (5000xg, 5 min, room temperature) to achieve an initial OD_{600} of 0.5 in the bioreactors after washing and resuspending in 0.9% NaCl aiming for a 0.5% (v/v) inoculation ratio. Batch cultivations were carried out in 1 L glass bioreactors (miniBio, Applikkon, Delft, The Netherlands) with 900 mL of initial working volume. Temperature was maintained at 30 °C with a temperature probe and a heating jacket. The culture pH was maintained at 5.5 with a pH probe and the automatic addition of sterile 2 mol/L KOH. Stirring (600 rpm) was achieved with two Rushton type impellers, one located at the bottom of the stirrer shaft and the other placed at approximately the 450 mL level height. Airflow was set to 1 vvm and the oxygen saturation was monitored with a Lumisens DO probe (Applikkon, Delft, Netherlands). Off-gas (O_2 and CO_2) composition (%) was monitored with a BlueInOne gas analyzer (BlueSens, Herten, Germany). Cultivations were performed for 32 h for the initial strain (PAR), as the endpoint was being established. For all subsequent experiments and strains, cultivations were terminated at 2 h, as this duration was considered to be sufficient to achieve the stationary phase. Data was collected and processed with Lucullus PIMS Lite v3.7.4 software (Getinge, Gothenburg, Sweden). Cultivations were sampled at least four times in the exponential phase and three times during the stationary phase, after the CO_2 composition in the off-gas had peaked. The samples were meant for OD_{600} measurement, extracellular metabolite analysis, fatty acid quantification, protein quantification, proteomics and carotenoid analysis (when applicable). All strains were cultivated in biological triplicates.

Quantification of cell concentration and growth

Samples were appropriately diluted to fit the linearity range of the spectrophotometer used for OD_{600} analysis. Those measurements were converted to $g_{DCW} \cdot L^{-1}$ using a calibration curve with a conversion factor of 0.2. The maximum specific growth rate for the exponential phase was calculated by fitting a regression line from the DCW natural logarithm against time.

Lipid analysis

Fatty acid composition and lipid content were determined via GC–MS with an internal standard method (Tammekivi

et al. 2019, 2021). Briefly, lyophilized cells were sonicated in the presence of methanol (2 mL for 10–12 mg of dry cells) for 15 min. Next, 98% H_2SO_4 was added (0.4 mL) to the mixture, which was derivatized for 3 h at 80 °C. After that, the mixture was extracted three times with 2 mL of hexane, followed by pipetting the extracts through a layer of K_2CO_3 on top of a glass wool. The hexane in the combined extracts was evaporated and the remaining fraction was redissolved in 2 mL of toluene with the addition of toluene and hexadecane (internal standard) to ensure each sample fit within the bounds of the calibration curve.

A GC (Agilent 7890 A) connected to a triple-axis MS detector (Agilent 5975 C) and an autosampler (Agilent G4513A) were used for the analysis (Agilent Technologies, Santa Clara, USA). The equipment employed a capillary column (30 m × 0.25 mm diameter, 0.25-µm film thickness, Agilent DB-225MS) with a (50%-cyanopropylphenyl)-methylpolysiloxane stationary phase. The MS transfer line operated at 280 °C, the ion source at 280 °C, and the injection inlet was maintained at 300 °C. Sample injection volume was 0.5 µL, where the split was opened after 2 min (splitless mode). Oven temperature started at 80 °C (2 min), increased to 200 °C (20 °C/min), maintained at 200 °C (4 min), increased to 220 °C (5 °C/min), maintained at 220 °C (5 min), increased to 230 °C (10 °C/min), and maintained at 230 °C (12 min), a total run time of 34 min. Electron ionization was carried with 70 eV. Solvent delay was 5.6 min. Carrier gas was Helium 6.0 at 1.5 $mL \cdot min^{-1}$. Scan mode (mass range from 27 to 400 m/z) was used for qualitative analysis as selected ion monitoring was used for quantitative analysis in the same run. Agilent MSD ChemStation and NIST Mass Spectral Library Search 2.0. were used for data analysis.

Fatty acid identities were identified by comparing retention times and mass spectra with a commercial standard mixture of fatty acid methyl esters (C8–C24). The standard mixture was also used to quantify five fatty acids (C16:0, C18:0, C18:1, C18:2, C18:3) by preparing calibration curves (more information on Tammekivi et al. 2019). Fatty acid concentrations (g/g_{DCW}) were calculated from their derivatized equivalents concentration, the output from the quantification. Total lipid content was approximated as the sum of the fatty acid concentrations.

Extracellular metabolite analysis

Extracellular metabolites (glycerol, citrate, and ethanol) in the cultivation sample's supernatant (21,000 g, 5 min) were analyzed by HPLC (Shimadzu LC-2050C, Kyoto, Japan). Samples were filtered (0.22 µm) prior to analysis. Injection volume was 10 µL. Detection occurred at 210 nm using a UV coupled with an RI detector. Aminex HPX-87H 300 × 7.8 mm (Bio-Rad, Hercules, USA) column was employed, temperature maintained at 45 °C. Mobile phase (H_2SO_4 5.0 mM) was pumped at 0.6 $mL \cdot min^{-1}$ (isocratic elution). Metabolites were identified and quantified by comparison with commercially available standards, with calibration curves linear over the relevant concentration ranges (0.625 to 40 $g \cdot L^{-1}$).

Protein content quantification

Total protein content was quantified as described previously (Reķēna et al. 2023), with the following minor modifications. Briefly, cell pellets were incubated in a commercial protein extraction solution (Y-PER, Thermo Fisher, Waltham, USA) prior to lysis. Lysis was performed via four cycles of bead (425–600 µm diameter) milling using a MonoLyzer™ (Rotaprep, Tustin, USA), 4000 cycles per min, for 20 s with 5 min of ice incubation between cycles. Protein concentration from the combined extracts was quantified using a Micro BCA Protein Assay Kit (Thermo Fisher, Waltham, USA) employing a bovine serum albumin (BSA) standard curve ranging from 0.5 to 40 $\mu g \cdot mL^{-1}$. The total protein content for each sample was calculated from the ratio of the total protein mass after extraction and the known mass of biomass (600 µg) extracted.

Carotenoid analysis

Samples for carotenoid analysis were washed with 0.9% NaCl and lyophilized. Subsequently, carotenoid extraction was performed similarly to Pinheiro et al. (2020). Briefly, the lyophilized biomass was resuspended in acetone and lysed using a MonoLyzer™ bead mill (425–600 µm diameter, acid-washed beads) for three cycles. Each cycle consisted of 30 s of milling at 4,000 cycles per min, followed by 1 min of ice incubation. After centrifugation (15,000 g, 5 min), the solvent layer was collected. The remaining biomass was subjected to repeated extraction until the cell debris became colorless. Subsequently, the combined extracts were evaporated under a N_2 stream, and the dried extract was resolubilized in a known volume of acetone. Carotenoid concentration was quantified spectrophotometrically at 450 nm using a β-carotene standard curve prepared in acetone, ranging from 2.5 to 30 $mg \cdot mL^{-1}$.

Proteomics

Sample preparation

All samples were processed in a randomized order. Sample preparation followed the single-pot, solid-phase-enhanced sample-preparation (SP3) protocol (Hughes et al. 2019), with minor modifications.

Briefly, cell pellets were resuspended in 10 volumes of lysis buffer (50 mmol·L^{-1} HEPES pH 8, 1% SDS, 1% Triton X-100, 1% NP-40, 1% Tween 20, 1% Na-deoxycholate, 50 mmol. L^{-1} NaCl, 50 mmol. L^{-1} DTT, 1 × Roche cOmplete protease inhibitor mix, 5 mmol·L^{-1} EDTA, and 1% glycerol). Samples were heated at 95 °C for 10 min, and sonicated with Bioruptor sonicator (Diagenode, USA) for 15 min at "High" setting. Further homogenization was achieved with a FastPrep24 bead beater (MP Biomedicals, USA), applying two 30-s pulses at 4 m/s with cooling between cycles. Lysates were clarified by centrifugation at 17,000 g for 10 min at 4 °C.

Aliquots were taken for protein concentration measurement using a Micro BCA™ Protein Assay Kit (Thermo Fisher) after precipitation with 10% trichloroacetic acid (TCA). For the SP3 protocols, 20 µg of protein from each sample was alkylated with 100 mmol·L^{-1} chloroacetamide for 1 h at room temperature in the dark. The sample volume was adjusted to 96 µL with lysis buffer before adding 4 µL of SP3 bead mix (1:1 suspension of Sera-Mag SpeedBead carboxylate-modified E3 and E7 magnetic particles, Cytiva, Marlborough, USA). One volume of ethanol was added to the binding mixture, which was then incubated at 24 °C for 5 min at 1,000 rpm. Beads were washed three times with 80% ethanol, with magnetic collection after each wash and then resuspended in 100 µL of trypsin/LysC digestion solution (1:100 enzyme:substrate ratio) in 100 mM ammonium bicarbonate. The suspension was sonicated for 30 s to disaggregate the beads before overnight incubation at 37 °C at 1000 rpm. Tryptic peptides were used in-house made C18 (3 M Empore, Oxford, USA) tips.

Mass spectrometry

LC–MS/MS analysis began by loading the prepared tryptic peptides onto a 0.3 × 5 mm trap-column (5 µm C18 particles, Dionex) using an Ultimate 3000 RSLCnano system (Dionex, Thermo Fisher). Peptides were then eluted to an in-house packed 50 cm × 75 µm emitter-column (3 µm C18 particles, Dr Maisch, Germany; MS Wil, CoAnn Technologies, the Netherlands). Elution from the analytical column used a 60-min, 8–40% gradient of buffer B (80% acetonitrile + 0.1% formic acid) in buffer A (0.1% formic acid). Peptides were online electrosprayed into a Q Exactive HF mass spectrometer (Thermo Fisher Scientific) via a nano-electrospray source (positive mode, spray voltage of 2.6 kV).

A data independent analysis (DIA) method was employed. Each MS1 (60,000 resolution, 60 ms max ion injection time, 3e6 ion target) was followed by 24 variable-size MS2 DIA isolation window scans (5 × 20, 7 × 15, 5 × 20, 3 × 30, 1 × 50, 2 × 84, and 1 × 189, with 1 m/z overlap), covering a 400 to 1200 m/z range. DIA scans were collected at 30,000 resolution, 41 ms max ion injection time, 3e6 max ions, a default charge state + 3, and a normalized collision energy 27.

Data analysis

Raw mass spectra data were processed using Dia-NN software (ver. 1.8.1) (Demichev et al. 2020). The search was carried out against the *Y. lipolytica* proteome database (downloaded from UniProt, Taxon ID 4952, in August 2022) supplemented with manually added relevant heterologous protein sequences. An in silico digest and library-free search mode were used with the default search options. The output was filtered at a 1% false discovery rate (FDR), and the match-between-runs option was enabled. Cysteine carbamidomethylation was set as a fixed modification, and the maximum number of missed cleavages was set to 1. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al. 2025) partner repository with the dataset identifier PXD072100.

Absolute protein abundances were calculated from the mass spectra intensities using the total protein approach (TPA) (Sánchez et al. 2021), assuming the detected proteins composed 99% of the total proteome. Differential expression analysis was performed over biological triplicates, with *p*-values corrected for multiple testing using the Benjamini–Hochberg approach (Benjamini and Hochberg 1995). Principal component analysis (PCA) of the proteomics samples was performed using the PCAtools package (Blighe and Lun 2024). Gene set analysis (GSA) was conducted using the Platform for integrated analysis of *omics* data (Piano) (Väremo et al. 2013) over proteome-normalized proteomics data ($\mu g_{protein} \cdot g_{protein}^{-1}$). Gene ontology (GO) groups with sizes between 5 and 300 genes were included in the analysis.

Genome-scale models

The model *iYali4_corr* (Xu et al. 2020) was chosen as the foundational GEM of *Y. lipolytica*. This model is an update of *iYali4* (Kerkhoven et al. 2016), developed after a comparative analysis with other *Y. lipolytica* GEMs (Xu et al. 2020). *iYali4_corr* was annotated using the CLIB122 (YALI0) strain reference proteome, which ensured direct compatibility with the proteomics dataset generated in this study and improved automated enzyme assignment within the GECKO 3.0 pipeline. Alternative GEMs annotated with non-reference strains were evaluated but showed reduced automated enzyme coverage due to gene identifier mismatches. *iYali4* (Kerkhoven et al. 2016) served as a reference for reaction subsystems, compartment names, and biomass component pseudo-reactions during the curation of the primary model. Extensive manual curation was performed on *iYali4_corr* to

enhance its compatibility with the GECKO 3.0 framework and to improve its overall accuracy following validation against UniProt or scientific literature. These modifications included the addition or removal of reactions, correction of reaction reversibility and bounds, and refinement of gene-protein-relations (GPRs) and naming conventions. Biomass composition was rebalanced, and the growth-associated maintenance (GAM) was recalculated. Modeling reactions were created to enable the prediction of targets for metabolic engineering. These curations resulted in *iYali5*.

A novel functional ecGEM of *Y. lipolytica* (*eciYali5*) was generated from *iYali5* using an adaptation of the semi-automated pipeline proposed in GECKO 3.2.2 (Chen et al. 2024; Kerkhoven et al. 2024a, b). For a given enzyme, k_{cat} values predicted by DLKcat were prioritized over k_{cat} values from BRENDA, unless a higher k_{cat} for a correct substrate from any species was present in the database. Total cell protein content was assumed to be 0.202 $g_{protein} \cdot g_{DCW}^{-1}$, consistent with our own measurements and external literature (Rakicka-Pustułka et al. 2021). The average enzyme saturation factor (σ, 0.49) was calculated using the corresponding GECKO 3 sigma fitter. Glycerol was used as the carbon source for the ecGEM generation. Reactions connected to sensitivity tuning output-limiting k_{cat} values were curated, as described earlier, and/or had their k_{cat} values updated to incorporate higher k_{cat} values from BRENDA. For the generation of *eciYali5*, 32 k_{cat} values were manually curated. These curations were also necessary when a k_{cat} was found to be limiting upon the later addition of proteomics data.

In these cases, k_{cat} reassignment followed a stepwise prioritization strategy. Experimentally determined k_{cat} values from BRENDA corresponding to phylogenetically similar organisms were first considered. If these remained growth-limiting, k_{cat} values associated with the same EC number from other organisms were evaluated. When no suitable experimental value was available, DLKcat predictions were used. As a last resort, the GECKO-calculated standard k_{cat} value was applied and, when required, increased in 10× increments. Additionally, the *Y. lipolytica* k_{cat} values from the BRENDA database (Schomburg et al. 2017) were not automatically detected by the GECKO pipeline and were thus manually incorporated. Finally, k_{cat} values of some putative enzymes without literature support were deliberately decreased to ensure these proteins were not preferentially utilized for FBA unless essential for the solution (De Biaggi 2025).

eciYali5 was updated to better represent each strain's genetic background prior to proteomics data integration. Heterologous enzymes had their respective reactions, compartments, GPRs, k_{cat} values, and molecular weights obtained from the literature, BRENDA, or UniProt metadata. Deletions were modeled by equating the bounds of the reactions catalyzed by their respective enzymes to zero. Biomass composition (including acyl-chain composition), GAM, and experimental fluxes (with 10% variance over experimental averages) were adjusted to match experimental conditions prior to the integration of protein abundances. Growth rates for the nitrogen-limiting phase were assumed to be 0.02 h^{-1} for every strain to ensure that model simulations still represent a turnover of biomass components, even while no or limited significant growth was observed at this sampling timepoint. Subsequently, enzyme abundance bounds were flexibilised until the target experimental growth rate was reached. At this stage, non-growth-associated maintenance (NGAM) was increased, leading to more flexibilization steps, which proved to be essential for some proteomics-constrained ecGEMs ($ecGEM_{prot}$) to achieve the target growth rate. Some conditions also required the unconstraining of the model's protein pool and gas exchange fluxes (O_2 and CO_2), which were later restored after the target growth rate had been achieved. An overview of all the GEM, ecGEM, and $ecGEM_{prot}$ models generated in this study can be found in Table S1.

For metabolic flux analysis, the $ecGEM_{prot}$ models were further constrained with their respective growth rates (10% variance in bounds). Their protein pool reaction was then minimized (with 10% variance in bounds) before being subjected to random sampling (5,000 samples) (Bordel et al. 2010) of the solution space, before taking mean values. ecGEM fluxes (random sampling means) were normalized by the glycerol uptake flux and then mapped to the curated source model. Their fold change (and *p*-values) was calculated as input for gene set analysis, also using the Piano package (Väremo et al. 2013). For this analysis, gene ontology (GO) groups were assigned to reactions by matching the gene names of the model's GPR with their corresponding UniProt GO group assignments. The significance (*p*-values) for each gene set was calculated using the Reporter features algorithm (Patil et al. 2005; Oliveira et al. 2008), which is implemented in Piano. The analysis was performed with gene permutation-based significance estimation, with adjustment for multiple testing using FDR correction. Targets for metabolic engineering were predicted using flux scanning of enforced objective function (FSEOF) (Choi et al. 2010). The *ecFSEOF* function, originally designed for ecGEMs in GECKO 3, was further adapted in this work for GEMs to allow for comparable predictions. Flux control coefficients were calculated as proposed by Nilsson and Nielsen (2016). Briefly, this measured the variation of a flux of interest caused by a 1% increase in the k_{cat} for each enzyme in an ecGEM. Enzyme usage was obtained from $ecGEM_{prot}$ fluxes (random sampling means) and from them high capacity and absolute usage reports were generated using, respectively, the *enzymeUsage*, and *reportEnzymeUsage* functions from GECKO toolbox (Chen et al. 2024; Kerkhoven et al. 2024a, b). Robustness analysis was performed by iteratively

constraining the usage flux of the enzyme involved in the target control reaction across its feasible range, while simultaneously maximizing a desired target objective function at discrete intervals.

The Raven toolbox 2 (v2.10.3) (Wang et al. 2018; Kerkhoven et al. 2024a, b) was largely used for its functions for flux balance analysis (FBA), addition of reactions, and other general GEM manipulations. Gurobi Optimizer (v11.0.3) (Gurobi Optimization, LLC 2024) was used as a solver. The scripts for the generation of the ecGEMs and general analysis have been deposited in the *bioengtaltech/eciYali5-GEM* GitHub repository (De Biaggi 2025). The model repository follows the standard-GEM (Anton et al. 2023) structure. Model quality and consistency for *iYali5* were evaluated using MEMOTE (Lieven et al. 2020), and the full report is provided as Supplementary File 1.

Results

Modified *Y. lipolytica* strains demonstrate typical behavior in bioreactors

The parental (PAR), carotenoid- (CAR), and lipid-producing (OBE) *Y. lipolytica* were cultured in triplicates in glycerol minimal medium (C/N 86) in aerobic pH-controlled batch bioreactors. All strains exhibited two distinct growth phases: an initial phase of unrestricted exponential growth, followed by a phase of growth under nitrogen limitation. Proteomics samples were collected during and following the exponential growth phase (Fig. 1). Strains PAR and CAR exhibited the most rapid growth, with no statistically significant difference observed between their specific growth rates (μ). Furthermore, similar biomass yields (Ysx) and specific glycerol uptake rates (r_{gly}) were recorded for these strains (Table 2). The respective growth rates of PAR and CAR strains, 0.45 h^{-1} and 0.42 h^{-1}, respectively, are comparable to values reported for *Y. lipolytica* DSM 21175 (0.39 to 0.47 h^{-1}) when cultured in defined nitrogen-limited media (C/N = 70) (Egermeier et al. 2017), representing one of the highest specific growth rates documented in the literature. Strain OBE demonstrated a faster growth rate in glycerol minimal medium (μ = 0.31 h^{-1}, C/N = 86) (Table 2) compared to non-limited glucose minimal medium (μ = 0.24 h^{-1}) (Sagnak et al. 2018). This observation is consistent with the previously demonstrated *Y. lipolytica* preference for glycerol over glucose (Workman et al. 2013).

The nitrogen-limited (N_{lim}) lipid content was higher in CAR (9.9% DCW) when compared to PAR (4.1% DCW) (Fig. 2a, Table 2). This could be due to the overexpression of the truncated hydroxymethylglutaryl CoA reductase (tHMGR, *YALI0E04807g*) (Table 1) in CAR, which likely increased the upstream acetyl-CoA pool (Fig. 5).

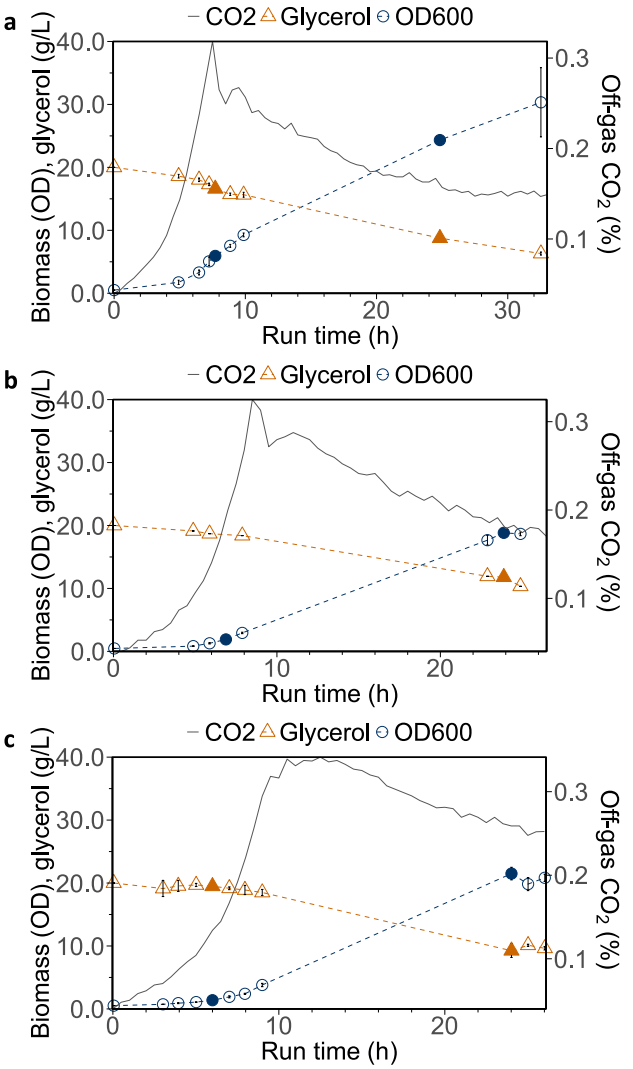

Fig. 1 Aerobic batch cultivation profiles of *Y. lipolytica* PAR (**a**), CAR (**b**), and OBE (**c**), in stirred tank reactors at 30 °C and pH 5.5. The profiles display biomass (OD_{600}), glycerol concentration, and off-gas CO_2 (%) over time. Filled symbols indicate time points at which samples were collected for subsequent proteomics, FAME, and carotenoid quantification (for CAR)

The carotenoid contents by this strain were 16.6 ± 4.3 and 11.6 ± 0.8 mg·g_{DCW} (Table 2), for the exponential and N_{lim} phases, respectively. Those values are within the same order of magnitude achieved by Larroude et al. (2018) after transforming *Y. lipolytica* Po1d with plasmid car-cassette (Table 1).

As anticipated, OBE was the strain with the highest lipid content in both growth phases (Fig. 2a, Table 2), indicating that it could accumulate more lipids even during its exponential growth phase, which is typically regarded as suboptimal for lipid accumulation. The lipid content of the obese strain during the N_{lim} phase (20.9% DCW) was comparable to the one obtained in a previous work growing the same strain in

Table 2 Key physiological parameters of *Y. lipolytica* strains from aerobic batch bioreactor cultivations. This table summarizes the maximum specific growth rate (μ_{max}), overall biomass yield on glycerol (Y_{XS}), specific glycerol uptake rate, specific CO_2 production rate, and lipid and protein content for strains PAR, CAR, and OBE. All values are reported as the average of triplicate independent bioreactor cultivations ± standard deviations. Cultivation conditions are detailed in the "Methods" section

Strain	μ_{max}(h^{-1})	Y_{XS}[†]	Uptake or production rate (mmol g_{DCW}^{-1} h^{-1})[†]		Lipid content (%)		Protein content (%)		Carotenoid content (mg/g_{DCW})	
			Glycerol	CO_2	exp	N_{lim}	exp	N_{lim}	exp	N_{lim}
PAR (control)	0.45 ± 0.02	0.46 ± 0.02	−10.7 ± 1.1	8.8 ± 0.4	3.8 ± 0.7	4.1 ± 0.3	20.1 ± 1.0	10.3 ± 0.7	NA	NA
CAR	0.42 ± 0.02	0.47 ± 0.00	−9.7 ± 0.6	16.1 ± 1.2[*]	3.8 ± 0.2	9.9 ± 0.5[**]	23.4 ± 3.9[‡]	11.0 ± 1.1[‡]	16.6 ± 4.3	11.6 ± 0.8
OBE	0.31 ± 0.01[*]	0.38 ± 0.04	−9.0 ± 1.0	14.0 ± 0.5	13.7 ± 0.6[**]	20.9 ± 1.7[**]	26.8 ± 1.7[**]	11.8 ± 0.9	NA	NA

[*] $p \leq 0.05$; [**] $p \leq 0.01$

[†] Global values

[‡] Calculated from the average protein content of the other strains

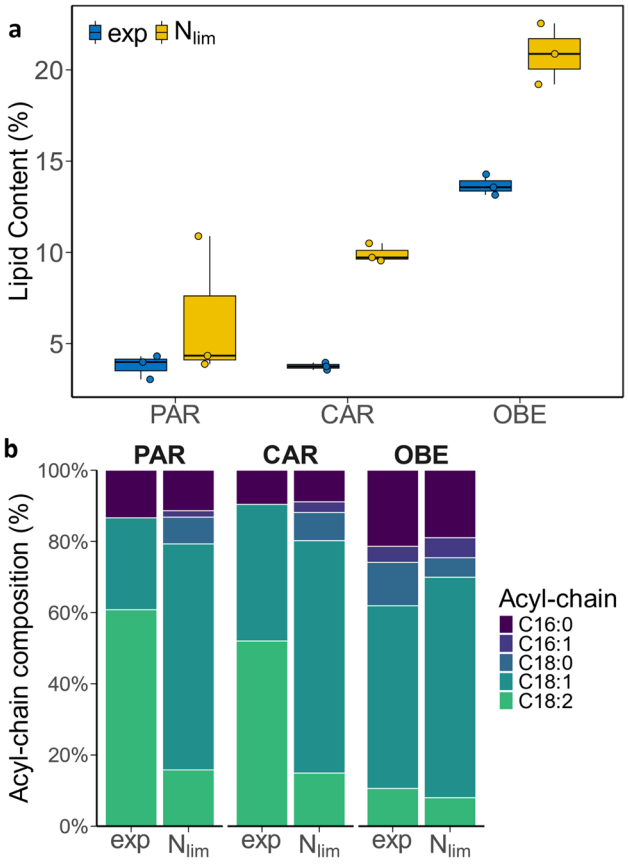

Fig. 2 Lipid (**a**) content (%) and fatty acyl-chain (**b**) distribution (%) from *Y. lipolytica* strains PAR, CAR and OBE. Samples were collected during the exponential (exp) and nitrogen-limiting (N_{lim}) growth phases of aerobic batch cultivations conducted as described in the Methods section. Boxplots show the distribution across biological replicates, with individual replicate values overlaid as points

glycerol media (C/N = 80) in shake flasks (~20% after 24 h) (Hambalko et al. 2021). Strain OBE also exhibited a more balanced distribution between C16 and C18 fatty acids, with total C16 species accounting for approximately 25% (mol/mol) on average across growth phases, compared to 13% in PAR and 11% in CAR (Fig. 2b). Correspondingly, total C18 species represented ~ 75% in OBE, whereas PAR and CAR contained ~ 87% and ~ 89%, respectively (Fig. 2b). Moreover, the average C18:2 fraction in OBE (~ 9%) was substantially lower than in PAR (~ 38%) and CAR (~ 34%) (Fig. 2b). A similar fatty-acid distribution was reported for JMY3501 (OBE) grown on glucose, fructose, or sucrose, where total C16 species accounted for approximately 34–35% of the fatty acids and C18:2 remained low (~ 6%) (Lazar et al. 2014). This suggests an influence of the strain's genetic background (Table 1) on its fatty acid (FA) composition, as JMY3501 carries deletions to reduce lipid turnover (Δpox1–6, Δtgl4) together with a push–pull module for TAG engineering (DGA2 and GPD1 overexpression). The protein contents of PAR and OBE during the N_{lim} phase decreased by at least two-fold compared to the exponential phase (Table 2), which is compatible with the reduced protein synthesis in the stationary phase.

Regarding other extracellular by-products, citrate and ethanol were not detected in significant concentrations throughout the entire cultivation, meaning that they were neither produced nor consumed (data not shown). This finding was further verified by a carbon balance for the PAR cultivation, which yielded a $C_{\_output/input}$ of 1.07. Furthermore, in other bioreactor batch cultivations of JMY3501 (OBE), citrate build-up was observed only after 24 h or even 72 h of cultivation (Sagnak et al. 2018; Lazar et al. 2014). Although cell viability was not directly quantified, no decrease in biomass concentration or unexpected accumulation of extracellular metabolites was observed during the nitrogen-limited phase.

Proteomics highlights downregulated sterol synthesis in modified strains

Proteomics analysis was performed to elucidate the physiological states of each strain across the studied growth

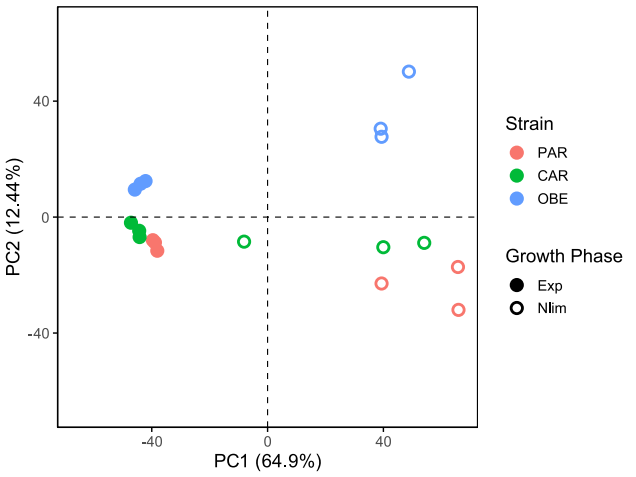

Fig. 3 PCA of proteomic data of *Y. lipolytica* strains. Samples (in triplicate) were obtained from aerobic batch cultivations in the exponential (Exp) and nitrogen-limiting (N_{lim}) growth phases, with experimental details provided in the Methods section

phases and assess the impact of the genetic modifications on the production of the target molecules. Absolute protein abundances were calculated from the intensities provided by the mass spectrometry analysis using the TPA method (Sánchez et al. 2021). An average of 4329 unique proteins was detected. Principal component analysis was performed on those protein abundances to inspect the variability among replicates and samples (Fig. 3). The greatest variability was observed between samples from the exponential and N_{lim} phases (PC1—65.1%). PC2 (12.43%), on the other hand, showed a less distinct pattern but appeared to correlate with biomass yield (Fig. 3, Table 2). One of the replicates of CAR $_{Nlim}$ was visually distinct in the PCA plot and was therefore excluded from the subsequent protein abundance calculations and differential expression analysis.

The physiological state among the strains and their growth phases was further compared using differential expression and gene set analysis (GSA) (Figure S1, Table 3). GO groups related to ribosomes and translation were downregulated in the N_{lim} phase compared to the exponential phase in all strains (Figure S1), indicating, as expected, that translation was more intense during logarithmic growth. A similar trend has been observed when comparing *Y. lipolytica* grown in glucose YNB at C/N ratios of 10 and 150 (Pomraning et al. 2016), and when comparing different growth phases of *Rhodotorula toruloides*, another oleaginous yeast, in minimal media with different carbon sources (Reķēna et al. 2023). This trend is also consistent with the higher protein content observed for all strains during exponential growth (Table 2).

Beyond the expected decrease in translation between the exp and N_{lim} phases, we specifically assessed whether the high-lipid-producing strain (OBE) showed an additional reduction in ribosome/translation-related processes

Table 3 Differentially expressed GO terms from proteomics data in modified versus parental strains. GO terms derived from proteomics data, comparing strains CAR and OBE with the parental strain (PAR) during the exponential growth phase. Arrows denote significantly (*p*-adj mix.dir < 0.05) up- (⇑) or down- (⇓) regulated processes. Dots (●) indicate differentially expressed processes without a significant directional change

GO terms	CAR	OBE
septin complex [GO:0031105]	⇓	⇓
septin cytoskeleton [GO:0032156]	⇓	⇓
ergosterol biosynthetic process [GO:0006696]	⇓	⇓
cell cortex [GO:0005938]	⇓	⇓
septin ring [GO:0005940]	⇓	⇓
cytoskeleton-dependent cytokinesis [GO:0061640]	●	⇓
endoplasmic reticulum [GO:0005783]	⇓	⇓
sterol biosynthetic process [GO:0016126]	⇓	⇓
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor [GO:0016616][†]		⇑⇓
ribosome [GO:0005840]	⇑	
plasma membrane [GO:0005886]	⇓	⇓
proteasome regulatory particle assembly [GO:0070682]	●	
oxidoreductase activity [GO:0016491]		⇑⇓
lipid droplet [GO:0005811]	⇓	●
cell cortex of cell tip [GO:0051285]	●	
GTPase activator activity [GO:0005096]		●
pyridoxal phosphate binding [GO:0030170]	●	
intracellular calcium ion homeostasis [GO:0006874]	●	
cell division site [GO:0032153]	⇓	
actin binding [GO:0003779]	⇓	
protein N-linked glycosylation via asparagine [GO:0018279]		⇓
mitochondrial large ribosomal subunit [GO:0005762]	⇑	
myosin complex [GO:0016459]	●	
acyltransferase activity [GO:0016746]		●
actin filament binding [GO:0051015]	⇓	
structural constituent of ribosome [GO:0003735]	⇑	
mating projection tip [GO:0043332]	⇓	
medial cortex [GO:0031097]	●	

[†]This gene set exhibited subsets of 121 genes significantly upregulated and 93 significantly downregulated

compared to PAR within the same growth-phase sampling point. Pomraning et al. (2016) suggested a potential positive correlation between lipid accumulation and the downregulation of ribosome biosynthesis and translation; however, in our dataset, proteome-based GSA did not indicate a significant (*p-adj* > 0.05) differential regulation of ribosome/translation GO terms in OBE relative to PAR either in the exponential or N_{lim} phases (Table S2), despite OBE accumulating 5.1 times more lipids than PAR under N_{lim} (Table 2).

During the exponential phase, CAR and OBE exhibited 20 and 16 differentially expressed GO groups, respectively, compared to PAR ($p < 0.05$) (Table S2). Nine of

these groups were common to both strains and were primarily associated with the septin complex (GO:0031105), endoplasmic reticulum (GO:0005783), and plasma membrane (GO:0005886). GO groups uniquely differentially expressed in CAR were linked to ribosomes (GO:0003735, GO:0005840, GO:0005762) and actin-mediated processes (GO:0003779, GO:0051015, GO:0043332, GO:0051285), whereas those unique to OBE were related to oxidoreductase activity (GO:0016491, GO:0016616) and acyl-chain processes (GO:0016746). In the N_{lim} phase, no differentially expressed GO groups were identified for CAR relative to PAR, indicating a high degree of physiological similarity between these two strains. In contrast, OBE displayed differentially expressed groups related to oligopeptide transport (Liu et al. 2022) (GO:0015833, GO:0035672, and GO:0035673).

Ergosterol biosynthesis proteins were downregulated in the exponential phase in both modified strains compared to the parental strain (PAR) (Table 3). The proteins encoded by *YALI0B15818g* (Q6CEG5) and *YALI0D15334g* (Q6C906), both involved in ergosterol biosynthesis regulation (Maguire et al. 2014), were also downregulated in the modified strains (log_2FC: −0.36 and −0.63 for *YALI0B15818g*; −2.13 and −2.35 for *YALI0D15334g* in the CAR and OBE strains, respectively) (Table S3). Previous studies have demonstrated an inverse relationship between sterol biosynthesis and both lipid and carotenoid production in *Y. lipolytica* (M. Liu et al. 2022). M. Liu et al. (2022) reported that knocking out ergosterol transcriptional regulation resulted in increases of 96% and 112% in carotenoid and lipid contents, respectively. Our results support the inverse relationship between lipid/terpene and sterol biosynthesis reported by Liu et al. (2022). This effect, however, was not observed in the N_{lim} phase.

Overall, proteomics results were consistent with the effects anticipated from the genetic modifications made to PAR. In CAR_{Exp}, the top differentially upregulated proteins were present in the carotenoid overexpression cassettes (plasmids HMG1-Nat and car-cassette, Table 1), namely, geranylgeranyl diphosphate synthase (GGS1, *YALI0D17050g*) and HMGR (Table S3). Notably, all the overexpressed proteins in CAR ranked in the third quartile of abundance for both growth phases, confirming their high expression levels (Fig. 4). Moreover, an increase in the relative abundance was also seen for all the overexpressed native proteins compared to PAR (Fig. 4), which demonstrates the success of this overexpression. OBE_{Exp}, in turn, exhibited differentially expressed acyltransferase activity (GO:0016746) and lipid droplet (GO:0005811) components when compared to PAR_{Exp}, which could be related to its increased lipid accumulation (Table 3). Diacylglycerol acyltransferase (DGA2, *YALI0D07986g*), which was overexpressed in OBE (Table 1), was also the most differentially expressed protein in OBE_{Exp} (Table S3). Furthermore, acyl-CoA desaturase (FAD2, *YALI0B10153g*) was slightly downregulated in OBE_{Exp} compared to PAR_{Exp} (Table S3), which aligned with the reduced C18:2 contents observed in the FAME analysis (Fig. 2). Similar to CAR, an increase in the relative abundance of the genes overexpressed in OBE was observed, most notably DGA2, which was enriched from the 17th percentile in PAR to the 87th percentile in the obese strain in the exponential phase (Fig. 4).

eciYali5 applications

Exploring insights from ecGEM analysis of experimental fluxes

Following the reconstruction of the $ecGEM_{prot}$, the proteomics-constrained model for each strain/growth phase pair was further constrained with the experimental fluxes calculated from the bioreactor experiments described in Sect. "Media composition", namely specific growth rate, glycerol uptake rate, O_2 uptake rate, and CO_2 production rate. No intracellular reaction fluxes were directly constrained. Subsequently, the glycerol-uptake normalized flux distributions of each condition were subjected to gene set analysis.

Similarly to other yeasts (Kito et al. 2016), *Y. lipolytica,* when grown on glycerol, utilizes upper gluconeogenesis to divert carbon to the Pentose phosphate pathway (PPP) and lower glycolysis to direct carbon to the TCA cycle (Fig. 5, Table S4). Within this context, the glycolytic process (GO:0006096) was upregulated in the modified strains compared to the parental strain during the exponential phase (Table S5). Simulations indicated that $ecPAR_{prot\_exp}$ diverted more carbon to the PPP than the other two models (Table S4), thereby reducing carbon availability for lower glycolysis. This diversion was likely necessary to produce more NADPH, which was required in greater amounts by glutamate dehydrogenase (GDH1, *YALI0F17820g*). The flux predicted for this enzyme by $ecPAR_{prot\_exp}$ ($0.335\ mmol \cdot g_{DCW}^{-1} \cdot h^{-1}$) was 4.0- and 5.8-fold higher than those predicted by $ecCAR_{prot\_exp}$ and $ecOBE_{prot\_exp}$, respectively. In *Y. lipolytica*, this enzyme participates in nitrogen uptake, particularly during nitrogen starvation, and has been speculated to have a regulatory role in this yeast's lipid and citrate production (Trotter et al. 2020).

Beyond nitrogen-related flux adjustments, proteome-constrained simulations also identified potential bottlenecks within central carbon metabolism in CAR strain. Enolase (ENO, *YALI0F16819g*) was predicted to be rate-limiting in both exponential and N_{lim} phases, with predicted capacity usages of 99.8% and 100%, respectively (Table S6). To further explore this, robustness analysis was performed with this enzyme using $ecCAR_{prot\_exp}$ *and* $ecCAR_{prot\_Nlim}$, where carotenoid production was evaluated

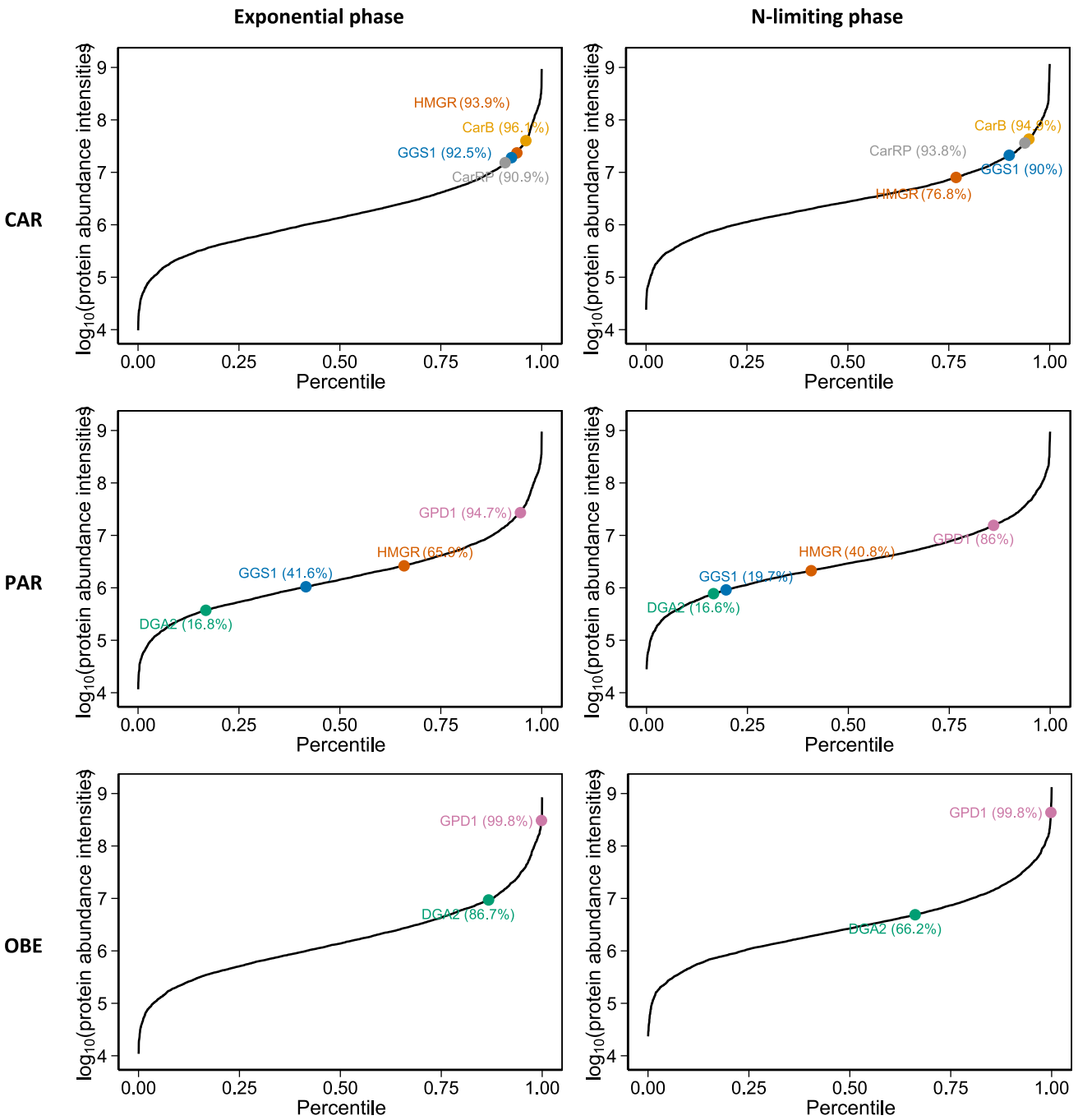

Fig. 4 Empirical cumulative distribution functions (ECDFs) of protein abundance for *Y. lipolytica* strains PAR, CAR, and OBE during the exponential and nitrogen-limiting growth phases. ECDFs of $\log_{10}$-transformed protein abundance intensities are shown. Proteins of interest are highlighted with distinct colors and labeled with their names and percentiles, illustrating their relative abundance within each group

across the feasible range of ENO's usage reaction fluxes. Carotenoid production increased with ENO availability until a plateau was reached, suggesting that additional metabolic constraints emerge once ENO limitation is relieved (Figure S2).

ecGEMs provide growth-phase specific targets for metabolic engineering

Targets for further metabolic engineering of *Y. lipolytica* for carotenoid and lipid accumulation were predicted using ecFSEOF (Chen et al. 2024) applied with $ecCAR_{\_(exp\ or\ Nlim)}$

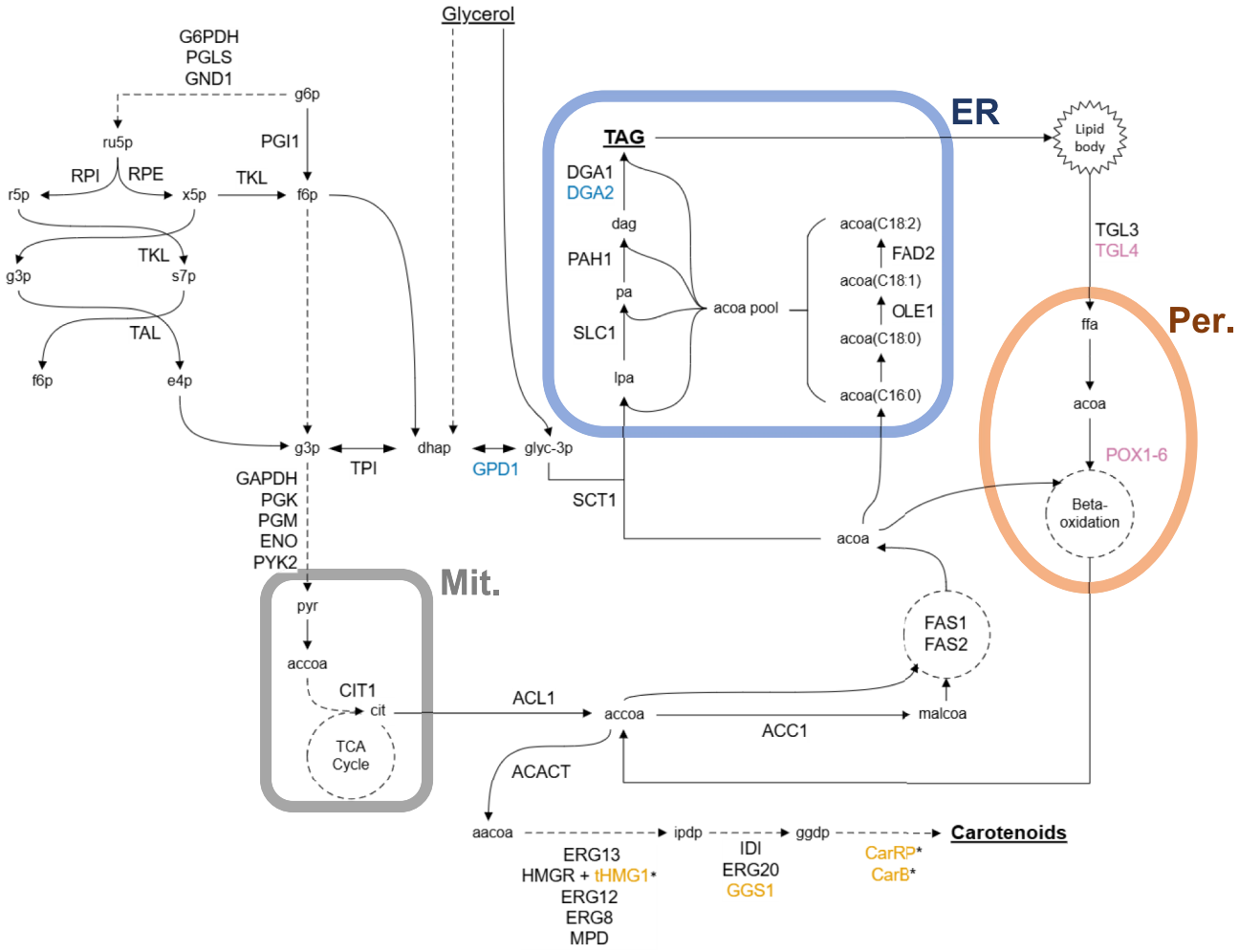

Fig. 5 Schematic overview of central carbon metabolism in *Y. lipolytica* for lipid and carotenoid production. Solid arrows represent a single reaction, while dashed arrows indicate multiple reaction steps. Enzymes marked with an asterisk (*) are heterologous. Enzyme coloring denotes strain-specific genetic modifications: orange indicates overexpression in CAR, blue indicates overexpression in OBE, pink indicates deletion in OBE. Only key enzymes and those mentioned in the main text are labeled with their respective acronyms to enhance figure clarity. Abbreviations: ACACT, acetyl-CoA C-acetyltransferase; ACC1, acetyl-coenzyme A carboxylase; ACL1, ATP-citrate lyase; CarB, phytoene dehydrogenase; CarRP, phytoene synthase/lycopene cyclase; CIT1, citrate synthase; DGA1-2, diacylglycerol acyltransferase; DPGM, diphosphoglyceromutase; ENO, enolase; ERG8, phosphomevalonate kinase; ERG12, mevalonate kinase; ERG13, hydroxymethylglutaryl-CoA synthase; ERG20, Farnesyl pyrophosphate synthase; FAD2, acyl-CoA desaturase; FAS1-2, fatty acid synthase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; G6PDH, glucose 6-phosphate dehydrogenase; GGS1, Geranylgeranyl pyrophosphate synthase; GND1, phosphogluconate dehydrogenase; GPD1, glycerol-3-phosphate dehydrogenase; HMGR, hydroxymethylglutaryl-CoA reductase; IDI, isopentenyl-diphosphate D-isomerase; MCS, 2-methylcitrate synthase; MPD, mevalonate pyrophosphate decarboxylase; OLE1, stearoyl-CoA desaturase; PAH1, phosphatidate phosphatase; PGM, phosphoglycerate mutase; PGI1, glucose-6-phosphate isomerase; PGK, phosphoglycerate kinase; PGLS, 6-phosphogluconolactonase; POX1-6, acyl-coenzyme A oxidase; PYK2, pyruvate kinase; RPE, ribulose 5-phosphate 3-epimerase; RPI, ribose-5-phosphate isomerase; SCT1, glycerol-3-phosphate O-acyltransferase; SLC1, 1-acyl-sn-glycerol-3-phosphate acyltransferase; TAL, transaldolase; TGL3-4, triacylglycerol lipase; TKL, transketolase; tHMG1, hydroxymethylglutaryl-CoA reductase (truncated). Organelle abbreviations: ER, endoplasmic reticulum; Mit., mitochondria; Per., peroxisome

and $ecPAR_{\_(exp\ or\ Nlim)}$ or $ecOBE_{\_(exp\ or\ Nlim)}$, respectively. Results for both lipid and carotenoid overexpression targets encompassed enzymes from pathways closely related to the target molecules, as well as enzymes from central carbon metabolism (Table 4).

Unlike GEM FSEOF predictions, where targets remained identical across exponential or N_{lim} phases, some putative overexpression targets from ecGEMs were growth-phase-specific (Table 4). These predictions were consistent with in vivo observations, as carotenogenic yeasts naturally

Table 4 FSEOF-predicted overexpression targets for increased carotenoid and lipid accumulation using regular GEMs and ecGEMs. This table presents the predicted overexpression targets identified by FSEOF analysis for increasing: (**a**) carotenoid accumulation in *Y. lipolytica* CAR and (**b**) lipid accumulation in *Y. lipolytica* strains PAR and OBE

(a)

Gene	Associated reactions	CAR_{GEM}		ecCAR		Reference
		exp	N_{lim}	exp	N_{lim}	
YALI0F05632g	MPD	●	●		●	Gao et al. (2017)†
YALI0F04015g	IDI	●	●		●	Gao et al. (2017)†
YALI0E06193g	ERG8	●	●		●	Gao et al. (2017)†
YALI0E04807g	HMGR	●	●		●	Gao et al. (2017)†
YALI0F05214g	TPI	●	●			
YALI0F09185g	PYK2	●	●			
YALI0B02728g	PGM	●	●			
YALI0C06369g	GPD1	●	●			
YALI0F16819g	ENO	●	●			
YALI0D12400g	PGK	●	●			
YALI0B16038g	ERG12	●	●			Gao et al. (2017)†
YALI0E05753g	ERG20	●	●		●	Gao et al. (2017)†
YALI0F30481g	ERG13	●	●			Gao et al. (2017)†
YALI0E11099g	ACACT	●	●			Gao et al. (2017)
MUCCIDRAFT_31317	CarB	●	●		●	Gao et al. (2017)
YALI0D17050g	GGS1	●	●		●	Gao et al. (2017)
CARRP	CarRP	●	●		●	Gao et al. (2017)

(b)

Gene	Associated reactions	PAR_{GEM}		OBE_{GEM}		ecPAR		ecOBE		Reference
		exp	N_{lim}	exp	N_{lim}	exp	N_{lim}	exp	N_{lim}	
YALI0C05951g	FAD2	●	●	●	●	●	●	●	●	Yan et al. (2020)
YALI0E06479g	TKL	●	●	●	●	●	●	●	●	Dobrowolski and Mirończuk (2020)
YALI0C11880g	RPE	●	●	●	●	●	●	●	●	Dobrowolski and Mirończuk (2020)†
YALI0B15598g	GND1	●	●	●	●	●	●	●	●	Dobrowolski and Mirończuk (2020)†, Silverman et al. (2016)†
YALI0E22649g	G6PDH	●	●	●	●	●	●	●	●	Dobrowolski and Mirończuk (2020)†, Yuzbasheva et al. (2017) ‡
YALI0C00209g	SCT1	●	●	●	●	●	●	●	●	Silverman et al. (2016)†
YALI0D27016g	PAH1	●	●	●	●	●	●	●	●	Silverman et al. (2016)†
YALI0C06369g	GAPDH	●	●	●	●		●	●	●	T. Dulermo and Nicaud (2011)
YALI0B02728g	PGM	●	●	●	●		●	●	●	
YALI0F09185g	PYK2	●	●	●	●		●	●	●	Mierzejewska and Chreptowicz (2016)§
YALI0D12400g	PGK	●	●	●	●		●	●	●	
YALI0F16819g	ENO	●	●	●	●		●	●	●	

Table 4 (continued)

YALI0E18964g	SLC1	●	●	●	●		●		●	Silverman et al. (2016)
YALI0E00638g	CIT1	●	●	●	●	●		●		J. Yang et al. (2019)[§]
YALI0E32769g	DGA1	●	●	●	●		●		●	Tai and Stephanopoulos (2013)
YALI0D07986g	DGA2	●	●	●	●	●		●		Silverman et al. (2016)
YALI0F15587g	TAL	●	●	●	●		●		●	Dobrowolski and Mirończuk (2020)[†]
YALI0C19085g	PGLS	●	●	●	●		●		●	Silverman et al. (2016)
YALI0E02684g	MCS	●	●	●	●		●		●	J. Yang et al. (2019)[§]
YALI0E11671g	PGL	●	●	●	●	●		●		Silverman et al. (2016)
YALI0F19514g	SLC1	●	●	●	●			●		Silverman et al. (2016)[†]
YALI0F07711g	PGI1					●	●	●	●	
YALI0B06941g	RPI						●	●	●	

Dots (●) indicate reactions where flux linearly increased when the models were linearly shifted from biomass maximization towards 90% enforcement of the target product. References are provided for targets that have been experimentally verified. Note: the symbols and abbreviations listed below apply to both Table 4a and Table 4b

ACACT acetyl-CoA C-acetyltransferase, *ACC1* acetyl-coenzyme A carboxylase, *ACL1* ATP-citrate lyase, *CarB* phytoene dehydrogenase, *CarRP* phytoene synthase/lycopene cyclase, *CIT1* citrate synthase, *DGA1-2* diacylglycerol acyltransferase, *DPGM* diphosphoglyceromutase, *ENO* enolase, *ERG8* phosphomevalonate kinase, *ERG12* mevalonate kinase, *ERG13* hydroxymethylglutaryl-CoA synthase, *ERG20* Farnesyl pyrophosphate synthase, *FAD2* acyl-CoA desaturase, *FAS1-2* fatty acid synthase, *GAPDH* glyceraldehyde-3-phosphate dehydrogenase, *G6PDH* glucose 6-phosphate dehydrogenase, *GGS1* Geranylgeranyl pyrophosphate synthase, *GND1* phosphogluconate dehydrogenase, *GPD1* glycerol-3-phosphate dehydrogenase, *HMGR* hydroxymethylglutaryl-CoA reductase, *IDI* isopentenyl-diphosphate D-isomerase, *MCS* 2-methylcitrate synthase, *MPD* mevalonate pyrophosphate decarboxylase, *OLE1* stearoyl-CoA desaturase, *PAH1* phosphatidate phosphatase, *PGM* phosphoglycerate mutase, *PGI1* glucose-6-phosphate isomerase, *PGK* phosphoglycerate kinase, *PGLS* 6-phosphogluconolactonase, *POX1-6* acyl-coenzyme A oxidase, *PYK2* pyruvate kinase, *RPE* ribulose 5-phosphate 3-epimerase, *RPI* ribose-5-phosphate isomerase, *SCT1* glycerol-3-phosphate O-acyltransferase, *SLC1* 1-acyl-sn-glycerol-3-phosphate acyltransferase, *TAL* transaldolase, *TGL3-4* triacylglycerol lipase, *TKL* transketolase

[†]Did not lead to significant improvements

[‡]Led to improvement only when co expressed with other modifications

[§]Tested in other species other than *Y. lipolytica*

accumulate carotenoids significantly during the N_{lim} phase (Freitas et al. 2014), and our ecGEM model predicted overexpression targets for carotenoid production exclusively during this same phase (Table 4a). No central carbon metabolism overexpression (OE) targets were predicted for *ecCAR_(exp or Nlim)* under any condition (Table 4a). In fact, we found no evidence of the central carbon metabolism genes listed in Table 4a having ever been overexpressed in yeast for increased carotenoid production. Nonetheless, other central carbon metabolism genes, such as those involved in acetyl-CoA supply (a precursor for carotenoid biosynthesis), have been successfully employed for this purpose (Q. Guo et al. 2024).

A similar trend was also observed when predicting lipid overexpression targets with *ecPAR_Nlim* (Table 4a), which also accumulated more lipids in the N_{lim} phase (Fig. 2a, Table 2). In contrast, *ecOBE_(exp or Nlim)* had a similar number of OE targets for both growth phases (Table 4b), which is consistent with its superior lipid accumulation in the exponential phase (Fig. 2a). For lipid ecFSEOF, two targets were predicted exclusively with the ecGEMs: glucose-6-phosphate isomerase (PGI1, *YALI0F07711g*), ribose-5-phosphate isomerase (RPI, *YALI0B06941g*) (Table 4b).

Notably, no OE targets were predicted by *ecCAR_exp* (Table 4a). Inspection of the FSEOF function revealed that although the carotenoid production linearly increased as growth decreased, as per the FSEOF principle, the slopes of the OE targets did not pass the function's upper quartile cut-off (data not shown). This indicates that, according to the simulations, the enforcement of carotenoid production only slightly redirects fluxes from biomass production, suggesting that the exponential phase is not optimal for carotenoid production.

The complete output of the ecFSEOF function includes not only OE targets, but also putative targets for knockdown (KD) and knockout (KO). In fact, many KD candidates were predicted across all strains and conditions, while *ecCAR_Nlim* was the only simulation that yielded a limited number of KO candidates (Table S7). However, gene repression approaches are not yet as well established in *Y. lipolytica* as in model microorganisms, and KO targets were identified for only a single condition. Therefore, the discussion was centered on OE targets, which represent more established and readily implementable engineering strategies in this organism. All targets predicted by ecFSEOF for all strains and conditions, as well as their predicted essentiality for biomass growth, are available in Table S7.

In addition to predicting gene overexpression targets, flux control coefficients (FCC) for carotenoid and lipid production were also calculated. Similar to ecFSEOF, FCCs also encompassed enzymes both closely related to the target molecules and enzymes from the central carbon metabolism (Fig. 6a,b). Phytoene dehydrogenase (CarB, *MUCCIDRAFT_31317*) was identified as the major bottleneck in carotenoid production (Fig. 6a), exhibiting an FCC two orders of magnitude higher than the second highest. This is likely attributable to the comparatively low k_{cat} for this enzyme used in the model (4.38×10^{-4} s^{-1}).

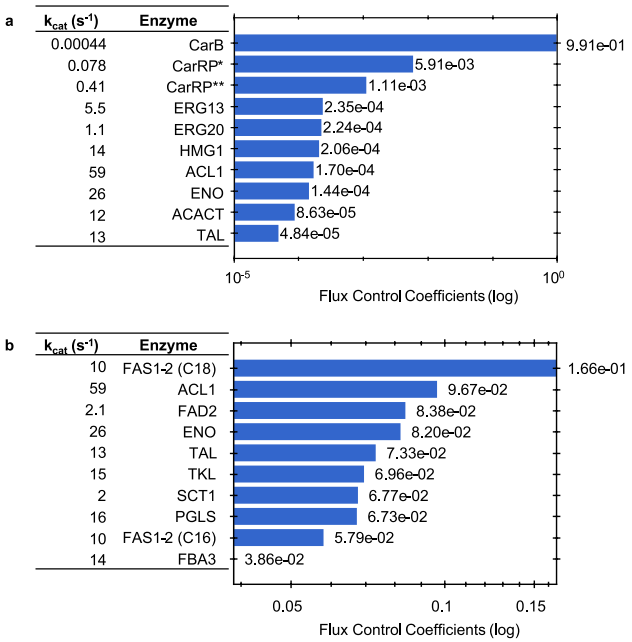

Fig. 6 Flux Control Coefficients for carotenoid and lipid biosynthesis. **a** FCCs for carotenoid biosynthesis from the ecGEM ecCAR. **b** FCCs for lipid biosynthesis from the ecGEM ecOBE. Both charts display the 10 reactions with the highest FCC from every reaction in the model. Those coefficients were obtained by constraining the model's glycerol uptake rate to 1 $mmol.g_{DCW}^{-1}.h^{-1}$ and maximizing either the **a** phytoene dehydrogenase reaction or **b** triacylglycerol production as the objective function. Refer to the Methods section for details on the FCC calculation, and to our GitHub page (De Biaggi 2025) for associated model files and scripts

For lipid production, the reaction with the highest FCC was the C18 formation step of the fatty acyl-CoA synthase complex (FAS1-2; *YALI0B15059g* and *YALI0B19382g*), which corresponded to the most abundant fatty acid in the FAME analysis (Fig. 2b) of OBE for both growth phases (Fig. 6b). Other key enzymes for lipid production, such as ATP-citrate lyase (ACL, *YALI0E34793g* and *YALI0D24431g*), also exhibited a high FCC over this molecule's biosynthesis. However, the magnitude difference between the FCCs was not as pronounced for lipid production. Given that lipid production pathways are more branched than the linear carotenoid biosynthesis, it is unlikely that a single enzyme would exert a major effect in controlling the flux of this pathway, as was shown for the latter. Protein engineering for increased catalytic efficiency was suggested as a strategy to improve enzymes that exert high control over a pathway (Sánchez et al. 2017). This approach, combined with other insights from metabolic modeling, would provide strategies

for further improving the strains used in this work, as discussed in the next section.

Discussion

Genome-scale modeling was able to predict several overexpression targets that had already been independently validated, attesting to their predictive abilities. The integration of enzymatic constraints further enhanced these predictions by identifying growth-phase-specific expression requirements (Table 4a) and suggesting targets omitted by standard GEMs (Table 4b). Furthermore, the calculation of FCCs – uniquely enabled by ecGEMs–augments FSEOF analysis by identifying which targets exert the greatest flux control over the target pathway. Lastly, metabolic bottlenecks can be identified by inspecting predicted enzyme usage (Section "Exploring insights from ecGEM analysis of experimental fluxes") and by evaluating the k_{cat} values of enzymes with high FCCs. The former insights depend on the incorporation of k_{cat} constraints and proteomics data, and would not be obtained from GEMs lacking enzymatic limitations.

These insights provide systematic guidance for metabolic engineering. Carotenoid overexpression usually involves the overexpression of genes in the mevalonate pathway (MVA), tHMG1 being the most effective, and the carotenoid biosynthesis pathway (Gao et al. 2017; Larroude et al. 2018; Du et al. 2025). Higher titers are typically achieved by adding extra copies of these genes (Gao et al. 2017). Our results highlight the effectiveness of this strategy: MVA and carotenoid biosynthesis genes were predominant both in FSEOF and FCC analyses (Table 4a, Fig. 6a). CAR would likely benefit from modification by this approach. In addition to confirming established strategies, we believe that the modeling provided insights for strategies not yet tested.

For instance, the fact that targets within the MVA and carotenoid biosynthesis pathways were predicted to have an impact on carotenoid production only in the N_{lim} phase (Table 4a). It suggests that either inducible promoters, or promoters of genes strongly expressed in the stationary phase, could be used to drive the expression of these genes when they are most effective in establishing the carotenoid production. This strategy has proven successful in the yeast *Kluyveromyces marxianus*, where improved carotenoid production was achieved by using a xylose-inducible promoter to trigger its accumulation when xylose-rich medium was added as feed after a biomass accumulation phase in a fed-batch cultivation (X. Yang et al. 2022). *Y. lipolytica* has several promoters induced by copper (Xiong and Chen 2020), erythritol (Park et al. 2019), or fatty acids (Shabbir Hussain et al. 2017), which could be employed for this end.

As briefly mentioned previously, another strategy to increase carotenoid production based on modeling results involves improving the catalytic efficiency (k_{cat}) of enzymes exerting the biggest flux control. According to our results, this enzyme is phytoene dehydrogenase (CarB) (Fig. 6a). Considering that this is one of the enzymes expressed in multiple copies to achieve higher carotenoid titers, mutating it for an improved turnover number could reduce the number of copies needed to achieve similar results, thereby reducing the metabolic burden caused by the synthesis of extra proteins. Enzyme engineering approaches, random or rational, have already been validated to improve carotenoid production in yeast. Strategies such as directed evolution and molecular docking analysis led to increases of 2.4-fold in yield (Zhou et al. 2017) and up to 36.8% in titer (Du et al. 2025) of astaxanthin in *S. cerevisiae*, respectively.

The overexpression of enzymes in the central carbon metabolism for push–pull strategies has also proven successful for increasing carotenoid production. For instance, Qiang et al. (2020) increased beta-carotene content in *Y. lipolytica* by 98% solely by overexpressing hexokinase (HXK1, *YALI0B22308g*). ecFSEOF predictions suggested enolase (ENO, *YALI0F16819g*) as a potential target for pushing the flux towards carotenoid production (Table 4a), which was also the central carbon metabolism enzyme with the highest FCC (Fig. 6a). Furthermore, $ecCAR_{prot}$ predicted that ENO operated close to maximum catalytic capacity in both growth phases of carotenoid production (Table S6), and robustness analysis showed that increasing its availability enhances carotenoid production until control shifts to another metabolic step (Figure S2). To our knowledge, ENO has not been attempted to be overexpressed for increased carotenoid production in any yeast species.

Combining these insights, an optimal strategy for increasing carotenoid production in CAR would involve overexpressing ENO under a constitutive promoter and multiple copies of the carotenoid-producing and MVA enzymes already present (Table 1) under an inducible promoter provided that CarB is first engineered for higher catalytic efficiency.

Similarly, the genes overexpressed in OBE (Table 1) and other independent strategies for increased lipid production were also suggested by FSEOF (Table 4b). This demonstrated the predictive capacities of $ecPAR_{\_(exp\ or\ Nlim)}$ and $ecOBE_{\_(exp\ or\ Nlim)}$.

The general approach of expressing multiple copies of the genes identified by FSEOF, utilizing central carbon metabolism enzymes for push–pull effects, and engineering enzymes with high FCCs, as outlined for carotenoid overexpression, could also be applied to lipid production. Most of these strategies have been individually successful in previous works. For example, expressing extra copies of DGA2 (up to three) in *Y. lipolytica* Q4 increased lipid content by up to 32% (Gajdoš et al. 2015). OBE was originally designed to achieve a push–pull effect towards lipid

production through the overexpression of DGA2 and glycerol-3-phosphate dehydrogenase (GPD1, *YALI0B02948g*) (Lazar et al. 2014). In the same study, adding an extra copy of native *hxk*1 enhanced the "push" effect, leading to higher lipid contents when these strains were grown in glucose, fructose, or sucrose.

Protein engineering of lipid metabolism enzymes has been reported in yeasts, for example to relieve regulation (Shi et al. 2014) or to alter substrate specificity (Gajewski et al. 2017). However, we found limited evidence of studies explicitly improving catalytic efficiency (i.e., kinetic constants such as k_{cat}) of such enzymes; thus, we propose this as a topic for future research, as FCC analysis showed that FAS1–2 complex (type I) exerted the biggest flux control over lipid production (Table 4b, Fig. 6b). However, its native structure, composed of six α (FAS2, *YALI0B19382g*) and six β (FAS1, *YALI0B15059g*) subunits (Köttig et al. 1991), might make this challenging to perform. Therefore, expressing heterologous fatty acyl-CoA synthase presents as an alternative. This strategy has proven successful in *S. cerevisiae*, where replacing FAS complex genes with a thirteen-gene hybrid transcription unit comprising ten *Escherichia coli* and three *Arabidopsis thaliana* FAS type II genes led to a lipid content three times higher than the parental strain (Pozdniakova et al. 2023). A similar approach could be employed in *Y. lipolytica*, initially by expressing a heterologous type II FAS transcriptional unit. This strategy is intended to debottleneck the pathway, shifting flux control to other metabolic reactions, which may then become new targets for k_{cat} optimization. Adding a new FAS system might also be desirable for fine-tuning the production of specialized fatty acids, as shown in Pozdniakova et al. (2023).

Despite the availability of inducible promoters for *Y. lipolytica*, we have not found any publication mentioning their use for increased lipid accumulation. In principle, we found that in *R. toruloides*, another oleaginous yeast, the expression of DGA1 under the LDP1 promoter, which is highly expressed during nitrogen limitation, led to a 21% higher lipid content compared to the same gene expressed under the constitutive promoter, pGPD1 (Y. Liu et al. 2016). However, considering that OBE already exhibited satisfactory lipid accumulation in the exponential phase of growth (Fig. 2), and that it did not show a clear growth phase distinction between its FSEOF-predicted targets (Table 4b), we conclude that inducible promoters might not offer significant gain for further improving lipid accumulation in this strain.

Therefore, based on our results, one possible strategy to further increase lipid production in OBE would involve expressing more copies of the GPD1 and DGA2 transcriptional unit already present in this strain and expressing a heterologous type II FAS transcriptional unit, which could later be engineered for higher catalytic efficiency.

Despite the thorough analysis of the model's predictions, we acknowledge that the suggestions made here require experimental validation, which we propose as a lead for future studies on improving lipid and carotenoid production in *Y. lipolytica*. Moreover, several other strategies could be explored to meet this goal when considering regulatory and kinetic layers, which are not considered in genome-scale metabolic modeling.

Conclusion

The updated genome-scale model with enzymatic constraints of *Y. lipolytica* (*eciYali5*), generated in this work using GECKO 3.0, proved to be a suitable tool for analyzing bioreactor cultivation and proteomics data, predicting metabolic engineering overexpression targets, and identifying metabolic bottlenecks for further improvement of existing lipid- or carotenoid-accumulating strains. The combination of these predictions allowed us to rationally identify targets for multi-copy expression, growth-phase specific induction, or improvement of catalytic activity through enzyme engineering. These insights were then combined into strain-specific strategies to maximize future modification efforts. Supporting this conclusion, cultivation and proteomics observations were consistent with independent work on Y. *lipolytica*. Simultaneously, our models independently predicted several successful metabolic engineering targets already reported and tested in the literature. The capabilities of our model (De Biaggi 2025) make it a useful tool for providing insights for the design and learn tasks in a Design-Build-Test-Learn (DBTL) cycle for improving microbial cell factories, or for enhancing the general understanding of *Y. lipolytica* physiology.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s00253-026-13791-4.

Acknowledgements We would like to acknowledge that Mauricio Pesantes participated in strain construction, Dr. Isma Belouah for her assistance in bioreactor cultivations, and to Dr. Ivar Ilves and other staff from University of Tartu's Proteomics Core Facility for the assistance with the proteomics methodology.

Author contribution PJL and RLA designed study. JSB and YKP performed experiments and wrote the manuscript. JSB and PJL analyzed the data. PJL, EJK and RLA supervised the study. All authors read and approved the manuscript.

Funding This research was supported by the European Cooperation in Science and Technology (COST) program, Action No. CA18229 Yeast-4Bio (Non-Conventional Yeasts for the Production of Bioproducts) (JSB, PJL), the Estonian Research Council grant PRG1101 (JSB and PJL), the European Union's Horizon Europe Teaming for Excellence program under grant agreement No. 101060066 (DIGIBIO) (JSB and PJL), the Novo Nordisk Foundation under grant no. NNF20CC0035580 (EJK), and the Bio-Based Industries Joint Undertaking under the EU's

Horizon 2020 research and innovation program (grant agreement No. 101022370 - PERFECOAT) (JSB, YKP, RLA, and PJL).

Data availability The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files, as well as publicly available on the GitHub repository *bioengtaltech/eciYali5-GEM* and the ProteomeXchange repository under the accession number *PXD072100*.

Declarations

Competing interests The authors declare no competing interests.

References

Anton M, Almaas E, Benfeitas R, Benito-Vaquerizo S, Blank LM, Dräger A, Hancock JM, Kittikunapong C, König M, Li F, Liebal UW, Lu H, Ma H, Mahadevan R, Mardinoglu A, Nielsen J, Nogales J, Pagni M, Papin JA, Patil KR, Price ND, Robinson JL, Sánchez BJ, Suarez-Diez M, Sulheim S, Svensson LT, Teusink B, Vongsangnak W, Wang H, Zeidan AA, Kerkhoven EJ (2023) Standard-GEM: standardization of open-source genome-scale metabolic models. Preprint, bioRxiv, March 23. https://doi.org/10.1101/2023.03.21.512712

Barth G, Gaillardin C (1996) *Yarrowia lipolytica*. In: Wolf K (ed) Nonconventional yeasts in biotechnology: a handbook. Springer. https://doi.org/10.1007/978-3-642-79856-6_10

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc: Ser B (Methodol) 57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blighe K, Lun A (2024) PCAtools: PCAtools: Everything principal components analysis. https://doi.org/10.18129/B9.bioc.PCAtools

Bordel S, Agren R, Nielsen J (2010) Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. PLoS Comput Biol 6(7):e1000859. https://doi.org/10.1371/journal.pcbi.1000859

Cao L, Li J, Yang Z, Hu X, Wang P (2023) A review of synthetic biology tools in *Yarrowia lipolytica*. World J Microbiol Biotechnol 39(5):129. https://doi.org/10.1007/s11274-023-03557-9

Chao R, Mishra S, Si T, Zhao H (2017) Engineering biological systems using automated biofoundries. Metab Eng 42:98–108. https://doi.org/10.1016/j.ymben.2017.06.003

Chen Y, Gustafsson J, Tafur Rangel A, Anton M, Domenzain I, Kittikunapong C, Li F, Yuan L, Nielsen J, Kerkhoven EJ (2024) Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO toolbox 3.0. Nat Protoc. https://doi.org/10.1038/s41596-023-00931-7

Choi HS, Lee SY, Kim TY, Woo HM (2010) In silico identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76(10):3097–3105. https://doi.org/10.1128/AEM.00115-10

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

De Biaggi JS (2025) Bioengtaltech/eciYali5-GEM: 1.0.1. V. v1.0.1. Zenodo, released October. https://doi.org/10.5281/zenodo.17434109

Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods 17(1):1. https://doi.org/10.1038/s41592-019-0638-x

Dobrowolski A, Mirończuk AM (2020) The influence of transketolase on lipid biosynthesis in the yeast *Yarrowia lipolytica*. Microb Cell Fact 19(1):138. https://doi.org/10.1186/s12934-020-01398-x

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):1. https://doi.org/10.1038/s41467-022-31421-1

Du J, Bao Y, Zhu J, Pang X, Ren D, Yin X, Zhou P (2025) Engineering of β-carotene hydroxylase for enhanced astaxanthin production in *Saccharomyces cerevisiae*. Biochem Eng J 219:109722. https://doi.org/10.1016/j.bej.2025.109722

Dulermo R, Gamboa-Meléndez H, Dulermo T, Thevenieau F, Nicaud J-M (2014) The fatty acid transport protein Fat1p is involved in the export of fatty acids from lipid bodies in Yarrowia lipolytica. FEMS Yeast Res 14(6):883–896. https://doi.org/10.1111/1567-1364.12177

Dulermo T, Nicaud J-M (2011) Involvement of the G3P shuttle and β-oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. Metab Eng 13(5):482–491. https://doi.org/10.1016/j.ymben.2011.05.002

Egermeier M, Russmayer H, Sauer M, Marx H (2017) Metabolic flexibility of *Yarrowia lipolytica* growing on glycerol. Front Microbiol 8. https://doi.org/10.3389/fmicb.2017.00049

Freitas C, Nobre B, Gouveia L, Roseiro J, Reis A, da Lopes Silva T (2014) New at-line flow cytometric protocols for determining carotenoid content and cell viability during *Rhodosporidium toruloides* NCYC 921 batch growth. Process Biochem 49(4):554–562. https://doi.org/10.1016/j.procbio.2014.01.022

Gajdoš P, Nicaud J-M, Rossignol T, Čertík M (2015) Single cell oil production on molasses by *Yarrowia lipolytica* strains overexpressing DGA2 in multicopy. Appl Microbiol Biotechnol 99(19):8065–8074. https://doi.org/10.1007/s00253-015-6733-8

Gajewski J, Pavlovic R, Fischer M, Boles E, Grininger M (2017) Engineering fungal de novo fatty acid synthesis for short chain fatty acid production. Nat Commun 8(1):14650. https://doi.org/10.1038/ncomms14650

Gao S, Tong Y, Zhu L, Ge M, Zhang Y, Chen D, Jiang Y, Yang S (2017) Iterative integration of multiple-copy pathway genes in *Yarrowia Lipolytica* for heterologous β-carotene production. Metab Eng 41:192–201. https://doi.org/10.1016/j.ymben.2017.04.004

Groenewald M, Boekhout T, Neuvéglise C, Gaillardin C, Van Dijck PWM, Wyss M (2014) *Yarrowia lipolytica* : safety assessment of an oleaginous yeast with a great industrial potential. Crit Rev Microbiol 40(3):187–206. https://doi.org/10.3109/1040841X.2013.770386

Guo Q, Peng Qian-Qian, Li Ya-Wen, Yan Fang, Wang Yue-Tong, Ye Chao, Shi T-Q (2024) Advances in the metabolic engineering of

Saccharomyces cerevisiae and *Yarrowia lipolytica* for the production of β-carotene. Crit Rev Biotechnol 44(3):337–351. https://doi.org/10.1080/07388551.2023.2166809

Guo Y, Liqiu Su, Qi Liu, Zhu Y, Dai Z, Wang Q (2022) Dissecting carbon metabolism of *Yarrowia lipolytica* type strain W29 using genome-scale metabolic modelling. Comput Struct Biotechnol J 20:2503–2511. https://doi.org/10.1016/j.csbj.2022.05.018

Gurobi Optimization LLC (2024) Gurobi Optimizer Reference Manual. https://www.gurobi.com

Hambalko J, Gajdoš P, Nicaud J-M, Ledesma-Amaro R, Tupec M, Pichová I, Čertík M (2021) Production of long chain fatty alcohols found in bumblebee pheromones by *Yarrowia lipolytica*. Front Bioeng Biotechnol. https://doi.org/10.3389/fbioe.2020.593419

Holkenbrink C, Dam MI, Kildegaard KR, Beder J, Dahlin J, Doménech Belda D, Borodina I (2018) EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. Biotechnol J 13(9):1700543. https://doi.org/10.1002/biot.201700543

Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB, Krijgsveld J (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc 14(1):68–85. https://doi.org/10.1038/s41596-018-0082-x

Ishchuk OP, Domenzain I, Sánchez BJ, Muñiz-Paredes F, Martínez JL, Nielsen J, Petranovic D (2022) Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 119(30):e2108245119. https://doi.org/10.1073/pnas.2108245119

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(1):72. https://doi.org/10.1186/s12918-015-0217-4

Kerkhoven E, 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(1):1. https://doi.org/10.1038/npjsba.2016.5

Kerkhoven E, Sánchez B, Domenzain I, johan-gson, Rangel AT, Anton M, Redestig H, Robinson J, Beber ME, Sulheim S, Badger TG (2024a) SysBioChalmers/GECKO: GECKO 3.2.2. Version v3.2.2. Zenodo, October. https://doi.org/10.5281/zenodo.13934828

Kerkhoven E, Wang H, Marcišauskas S, Robinson J, Sánchez B, Anton M, johan-gson, Kittikunapong C, Domenzain I, Ågren R, danieljcook, Rangel AT, Pfau T, Väremo L, Badger TG, SciLifeLab S@, xkzhang (2024b) SysBioChalmers/RAVEN: V2.10.3. Zenodo, December 21. https://doi.org/10.5281/zenodo.14541827

Kito K, Ito H, Nohara T, Ohnishi M, Ishibashi Y, Takeda D (2016) Yeast interspecies comparative proteomics reveals divergence in expression profiles and provides insights into proteome resource allocation and evolutionary roles of gene duplication*. Mol Cell Proteomics 15(1):218–235. https://doi.org/10.1074/mcp.M115.051854

Köttig H, Rottner G, Beck K-F, Schweizer M, Schweizer E (1991) The pentafunctional FAS1 genes of *Saccharomyces cerevisiae* and *Yarrowia lipolytica* are co-linear and considerably longer than previously estimated. Mol Gen Genet 226(1):310–314. https://doi.org/10.1007/BF00273618

Kumar LR, Tyagi RD, Drogui P (2023) Economic analysis for simultaneous production of microbial lipid and citric acid by oleaginous yeast cultivated on purified crude glycerol. Biomass Convers Biorefin 13(10):9141–9154. https://doi.org/10.1007/s13399-021-01772-8

Larroude M, Celinska E, Back A, Thomas S, Nicaud J-M, Ledesma-Amaro R (2018) A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β-carotene. Biotechnol Bioeng 115(2):464–472. https://doi.org/10.1002/bit.26473

Larroude M, Park Y-K, Soudier P, Kubiak M, Nicaud J-M, Rossignol T (2019) A modular Golden Gate toolkit for *Yarrowia lipolytica* synthetic biology. Microb Biotechnol 12(6):1249–1259. https://doi.org/10.1111/1751-7915.13427

Lazar Z, Dulermo T, Neuvéglise C, Crutz-Le Coq A-M, Nicaud J-M (2014) Hexokinase—a limiting factor in lipid production from fructose in *Yarrowia lipolytica*. Metab Eng 26:89–99. https://doi.org/10.1016/j.ymben.2014.09.008

Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, Kerkhoven EJ, Nielsen J (2022) Deep learning-based Kcat prediction enables improved enzyme-constrained model reconstruction. Nat Catal 5(8):8. https://doi.org/10.1038/s41929-022-00798-z

Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell JA, Blank LM, Chauhan S, Correia K, Diener C, Dräger A, Ebert BE, Edirisinghe JN, Faria JP, Feist AM, Fengos G, Fleming RMT, García-Jiménez B, Hatzimanikatis V, van Helvoirt W, Henry CS, Hermjakob H, Herrgård MJ, Kaafarani A, Kim HU, King Z, Klamt S, Klipp E, Koehorst JJ, König M, Lakshmanan M, Lee D-Y, Lee SY, Lee S, Lewis NE, Liu F, Ma H, Machado D, Mahadevan R, Maia P, Mardinoglu A, Medlock GL, Monk JM, Nielsen J, Nielsen LK, Nogales J, Nookaew I, Palsson BO, Papin JA, Patil KR, Poolman M, Price ND, Resendis-Antonio O, Richelle A, Rocha I, Sánchez BJ, Schaap PJ, Malik Sheriff RS, Shoaie S, Sonnenschein N, Teusink B, Vilaça P, Vik JO, Wodke JAH, Xavier JC, Yuan Q, Zakhartsev M, Zhang C (2020) Memote for standardized genome-scale metabolic model testing. Nat Biotechnol 38(3):272–276. https://doi.org/10.1038/s41587-020-0446-y

Liu M, Zhang J, Liu X, Hou J, Qi Q (2022) Rapid gene target tracking for enhancing β-carotene production using flow cytometry-based high-throughput screening in *Yarrowia lipolytica*. Appl Environ Microbiol 88(19):e01149-e1222. https://doi.org/10.1128/aem.01149-22

Liu Y, Yap SA, Koh CMJ, Ji L (2016) Developing a set of strong intronic promoters for robust metabolic engineering in oleaginous *Rhodotorula* (*Rhodosporidium*) yeast species. Microb Cell Fact 15(1):200. https://doi.org/10.1186/s12934-016-0600-x

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(1):35. https://doi.org/10.1186/1752-0509-6-35

Maguire SL, Wang C, Holland LM, Brunel F, Neuvéglise C, Nicaud J-M, Zavrel M, White TC, Wolfe KH, Butler G (2014) Zinc finger transcription factors displaced SREBP proteins as the major sterol regulators during *Saccharomycotina evolution*. PLoS Genet 10(1):e1004076. https://doi.org/10.1371/journal.pgen.1004076

Mierzejewska J, Chreptowicz K (2016) Lack of Maf1 enhances pyruvate kinase activity and fermentative metabolism while influencing lipid homeostasis in *Saccharomyces cerevisiae*. FEBS Lett 590(1):93–100. https://doi.org/10.1002/1873-3468.12033

Mishra P, Lee N-R, Lakshmanan M, Kim M, Kim B-G, Lee D-Y (2018) Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica*. BMC Syst Biol 12(2):12. https://doi.org/10.1186/s12918-018-0542-5

Nilsson A, Nielsen J (2016) Metabolic trade-offs in yeast are caused by F1F0-ATP synthase. Sci Rep 6(1):1. https://doi.org/10.1038/srep22264

Oliveira AP, Patil KR, Nielsen J (2008) Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. BMC Syst Biol 2(1):17. https://doi.org/10.1186/1752-0509-2-17

Park Y-K, Korpys P, Kubiak M, Celińska E, Soudier P, Trébulle P, Larroude M, Rossignol T, Nicaud J-M (2019) Engineering the architecture of erythritol-inducible promoters for regulated and

enhanced geneexpression in *Yarrowia lipolytica*. FEMS Yeast Res 19(1):foy105. https://doi.org/10.1093/femsyr/foy105

Park Y-K, Ledesma-Amaro R (2023) What makes *Yarrowia lipolytica* well suited for industry? Trends Biotechnol 41(2):242–254. https://doi.org/10.1016/j.tibtech.2022.07.006

Patil KR, Rocha I, Förster J, Nielsen J (2005) Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6(1):308. https://doi.org/10.1186/1471-2105-6-308

Perez-Riverol Y, Bandla C, Kundu DJ, Kamatchinathan S, Bai J, Hewapathirana S, John NS, Prakash A, Walzer M, Wang S, Vizcaíno JA (2025) The PRIDE database at 20 years: 2025 update. Nucleic Acids Res 53(D1):D543–D553. https://doi.org/10.1093/nar/gkae1011

Pinheiro MJ, Bonturi N, Belouah I, Miranda EA, Lahtvee P-J (2020) Xylose metabolism and the effect of oxidative stress on lipid and carotenoid production in *Rhodotorula toruloides*: insights for future biorefinery. Front Bioeng Biotechnol. https://doi.org/10.3389/fbioe.2020.01008

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(1):138. https://doi.org/10.1186/s12864-016-2471-2

Pozdniakova TA, Cruz JP, Silva PC, Azevedo F, Parpot P, Domingues MR, Carlquist M, Johansson B (2023) Optimization of a hybrid bacterial/*Arabidopsis thaliana* fatty acid synthase system II in *Saccharomyces cerevisiae*. Metab Eng Commun 17:e00224. https://doi.org/10.1016/j.mec.2023.e00224

Qiang S, Wang J, Xiong XC, Qu YL, Liu L, Hu CY, Meng YH (2020) Promoting the synthesis of precursor substances by overexpressing hexokinase (Hxk) and hydroxymethylglutaryl-CoA synthase (Erg13) to elevate β-carotene production in engineered *Yarrowia lipolytica*. Front Microbiol. https://doi.org/10.3389/fmicb.2020.01346

Rakicka M, Rywińska A, Lazar Z, Rymowicz W (2017) Two-stage continuous culture – technology boosting erythritol production. J Clean Prod 168:420–427. https://doi.org/10.1016/j.jclepro.2017.09.060

Rakicka-Pustułka M, Miedzianka J, Jama D, Kawalec S, Liman K, Janek T, Skaradziński G, Rymowicz W, Lazar Z (2021) High value-added products derived from crude glycerol via microbial fermentation using *Yarrowia* clade yeast. Microb Cell Fact 20(1):195. https://doi.org/10.1186/s12934-021-01686-0

Reķēna A, Pinheiro MJ, Bonturi N, Belouah I, Tammekivi E, Herodes K, Kerkhoven EJ, Lahtvee P-J (2023) Genome-scale metabolic modeling reveals metabolic trade-offs associated with lipid production in *Rhodotorula toruloides*. PLoS Comput Biol 19(4):e1011009. https://doi.org/10.1371/journal.pcbi.1011009

Rymowicz W, Fatykhova AR, Kamzolova SV, Rywińska A, Morgunov IG (2010) Citric acid production from glycerol-containing waste of biodiesel industry by *Yarrowia lipolytica* in batch, repeated batch, and cell recycle regimes. Appl Microbiol Biotechnol 87(3):971–979. https://doi.org/10.1007/s00253-010-2561-z

Sagnak R, Cochot S, Molina-Jouve C, Nicaud J-M, Guillouet SE (2018) Modulation of the glycerol phosphate availability led to concomitant reduction in the citric acid excretion and increase in lipid content and yield in *Yarrowia lipolytica*. J Biotechnol 265:40–45. https://doi.org/10.1016/j.jbiotec.2017.11.001

Sánchez BJ, Lahtvee P-J, Campbell K, Kasvandik S, Yu R, Domenzain I, Zelezniak A, Nielsen J (2021) Benchmarking accuracy and precision of intensity-based absolute quantification of protein abundances in *Saccharomyces cerevisiae*. Proteomics 21(6):2000093. https://doi.org/10.1002/pmic.202000093

Sánchez BJ, Zhang C, Nilsson A, Lahtvee P-J, Kerkhoven EJ, Nielsen J (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol Syst Biol 13(8):935. https://doi.org/10.15252/msb.20167411

Schomburg I, Jeske L, Ulbrich M, Placzek S, Chang A, Schomburg D (2017) The BRENDA enzyme information system–from a database to an expert system. Journal of Biotechnology, Bioinformatics Solutions for Big Data Analysis in Life Sciences Presented by the German Network for Bioinformatics Infrastructure 261(November):194–206. https://doi.org/10.1016/j.jbiotec.2017.04.020

Shabbir Hussain M, Wheeldon I, Blenner MA (2017) A strong hybrid fatty acid inducible transcriptional sensor built from *Yarrowia lipolytica* upstream activating and regulatory sequences. Biotechnol J 12(10):1700248. https://doi.org/10.1002/biot.201700248

Shi S, Chen Y, Siewers V, Nielsen J (2014) Improving production of malonyl coenzyme A-derived metabolites by abolishing Snf1-dependent regulation of Acc1. Mbio 5(3):10.1128/mbio.01130-14. https://doi.org/10.1128/mbio.01130-14

Silverman AM, Qiao K, Xu P, Stephanopoulos G (2016) Functional overexpression and characterization of lipogenesis-related genes in the oleaginous yeast *Yarrowia lipolytica*. Appl Microbiol Biotechnol 100(8):3781–3798. https://doi.org/10.1007/s00253-016-7376-0

Tai M, Stephanopoulos G (2013) Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. Metab Eng 15(January):1–9. https://doi.org/10.1016/j.ymben.2012.08.007

Tammekivi E, Vahur S, Kekišev O, Werf ID van der, Toom L, Herodes K, Leito I (2019) Comparison of derivatization methods for the quantitative gas chromatographic analysis of oils. Anal Methods 11(28):3514–3522. https://doi.org/10.1039/C9AY00954J

Tammekivi E, Vahur S, Vilbaste M, Leito I (2021) Quantitative GC–MS analysis of artificially aged paints with variable pigment and linseed oil ratios. Molecules 26(8):8. https://doi.org/10.3390/molecules26082218

Tomás-Pejó E, Morales-Palomo S, González-Fernández C (2023) *Cutaneotrichosporon curvatum* and *Yarrowia lipolytica* as key players for green chemistry: efficient oil producers from food waste via the carboxylate platform. Bioengineered 14(1):2286723. https://doi.org/10.1080/21655979.2023.2286723

Trotter PJ, Juco K, Le HT, Nelson K, Tamayo LI, Nicaud J-M, Park Y-K (2020) Glutamate dehydrogenases in the oleaginous yeast *Yarrowia lipolytica*. Yeast 37(1):103–115. https://doi.org/10.1002/yea.3425

Väremo L, Nielsen J, Nookaew I (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic Acids Res 41(8):4378–4391. https://doi.org/10.1093/nar/gkt111

Verduyn C, Postma E, Scheffers WA, Van Dijken JP (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8(7):501–517. https://doi.org/10.1002/yea.320080703

Vijayakumar S, Rahman PKSM, Angione C (2020) A hybrid flux balance analysis and machine learning pipeline elucidates metabolic adaptation in cyanobacteria. iScience. https://doi.org/10.1016/j.isci.2020.101818

Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, Nielsen J, Kerkhoven EJ (2018) RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. PLoS Comput Biol 14(10):e1006541. https://doi.org/10.1371/journal.pcbi.1006541

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(1):51. https://doi.org/10.1186/s40643-017-0180-6

Workman M, Holt P, Thykaer J (2013) Comparing cellular performance of *Yarrowia lipolytica* during growth on glucose and glycerol in submerged cultivations. AMB Express 3(1):58. https://doi.org/10.1186/2191-0855-3-58

Xiong X, Chen S (2020) Expanding toolbox for genes expression of *Yarrowia lipolytica* to include novel inducible, repressible, and hybrid promoters. ACS Synth Biol 9(8):2208–2213. https://doi.org/10.1021/acssynbio.0c00243

Xu Y, Holic R, Hua Q (2020) Comparison and analysis of published genome-scale metabolic models of *Yarrowia lipolytica*. Biotechnol Bioprocess Eng 25(1):53–61. https://doi.org/10.1007/s12257-019-0208-1

Yan FX, Dong GR, Qiang S, Niu YJ, Hu CY, Meng YH (2020) Overexpression of △12, △15-desaturases for enhanced lipids synthesis in *Yarrowia lipolytica*. Front Microbiol. https://doi.org/10.3389/fmicb.2020.00289

Yang J, Li S, Khan MAK, Garre V, Vongsangnak W, Song Y (2019) Increased lipid accumulation in *Mucor circinelloides* by overexpression of mitochondrial citrate transporter genes. Ind Eng Chem Res 58(6):2125–2134. https://doi.org/10.1021/acs.iecr.8b05564

Yang X, Wang D, Hong J (2022) Carotenoid production from Nondetoxified xylose mother liquid or corncob hydrolysate with engineered *Kluyveromyces marxianus*. Bioresour Technol 364:128080. https://doi.org/10.1016/j.biortech.2022.128080

Yuzbasheva EY, Mostova EB, Andreeva NI, Yuzbashev TV, Laptev IA, Sobolevskaya TI, Sineoky SP (2017) Co-expression of Glucose-6-Phosphate dehydrogenase and Acyl-CoA binding protein enhances lipid accumulation in the yeast *Yarrowia lipolytica*. New Biotechnol 39:18–21. https://doi.org/10.1016/j.nbt.2017.05.008

Zhou P, Xie W, Li A, Wang F, Yao Z, Bian Q, Zhu Y, Yu H, Ye L (2017) Alleviation of metabolic bottleneck by combinatorial engineering enhanced astaxanthin synthesis in *Saccharomyces cerevisiae*. Enzyme Microb Technol 100:28–36. https://doi.org/10.1016/j.enzmictec.2017.02.006