Computational and Structural Biotechnology Journal 20 (2022) 2503–2511

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

# Dissecting carbon metabolism of *Yarrowia lipolytica* type strain W29 using genome-scale metabolic modelling

Yufeng Guo [a,b], Liqiu Su [a,b], Qi Liu [a,b], Yan Zhu [c,*], Zongjie Dai [a,b,*], Qinhong Wang [a,b]

[a] Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
[b] TIB-VIB Joint Center of Synthetic Biology, National Center of Technology Innovation for Synthetic Biology, Tianjin 300308, China
[c] Infection Program and Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

ARTICLE INFO

*Article history:*
Received 14 January 2022
Received in revised form 7 May 2022
Accepted 9 May 2022
Available online 16 May 2022

*Keywords:*
*Yarrowia lipolytica*
Genome-scale metabolic model
Carbon metabolism
Theoretical yield

ABSTRACT

*Yarrowia lipolytica* is a widely-used chassis cell in biotechnological applications. It has recently gained extensive research interest owing to its extraordinary ability of producing industrially valuable biochemicals from a variety of carbon sources. Genome-scale metabolic models (GSMMs) enable analyses of cellular metabolism for engineering various industrial hosts. In the present study, we developed a high-quality GSMM *i*Yli21 for *Y. lipolytica* type strain W29 by extensive manual curation with Biolog experimental data. The model showed a high accuracy of 85.7% in predicting nutrient utilization. Transcriptomics data were integrated to delineate cellular metabolism of utilizing six individual metabolites as sole carbon sources. Comparisons showed that 302 reactions were commonly used, including those from TCA cycle, oxidative phosphorylation, and purine metabolism for energy and material supply. Whereas glycolytic reactions were employed only when glucose and glycerol used as sole carbon sources, gluconeogenesis and fatty acid oxidation reactions were specifically employed when fatty acid, alkane and glycerolipid were the sole carbon sources. Further test of 46 substrates for generating 5 products showed that hexanoate outcompeted other compounds in terms of maximum theoretical yield owing to the lowest carbon loss for energy supply. This newly generated model *i*Yli21 will be a valuable tool in dissecting metabolic mechanism and guiding metabolic engineering of this important industrial cell factory.



## 1. Introduction

The growing environmental concerns of fossil-based chemical production urge the development of alternative microbial bioproduction [1,2]. As a non-conventional model oleaginous yeast [3], *Yarrowia lipolytica* has recently gained extensive research interest owing to its extraordinary ability of producing industrially valuable biochemicals such as citrate, erythritol, lipids and terpenoids, from a variety of carbon substrates including plant oils, glycerol, acetate, xylose and even organic waste materials [4]. Notably, *Y. lipolytica* can accumulate lipids up to 90% of its dry weight when grown on cheap substrates such as glucose [4], making it an ideal eukaryotic cell factory for lipid biofuel production. However, most of studies focused on improvement of fermentation process [5-7], whereas the mechanistic understanding of cellular metabolism is limited, thereby significantly hindering the effective engineering of this important industrial yeast for biochemical production. *Y. lipolytica* type strain W29 is the origin of many genetically engineered *Y. lipolytica* strains [8], which has been extensively studied for its physiological metabolism, genome editing and metabolic engineering [9,10]. Compete genome sequence and annotation of W29 are available in NCBI database.

Genome-scale metabolic model (GSMM) and flux balance analysis (FBA) have been widely employed to analyze metabolic phenotypes, stress responses, antimicrobial killing, and engineering

*Abbreviations:* G6P, D-glucose 6-phosphate; G3P, D-glyceraldehyde 3-phosphate; DHAP, dihydroxyacetone phosphate; PEP, phosphoenolpyruvate; OAA, oxaloacetic acid; MAL, malate; CIT, citrate; ICI, isocitrate; AKG, α-ketoglutarate; SUCCCOA, succinyl-CoA; SUCC, succinate; FUM, fumarate; 6PGL, 6-phosphogluconolactone; 6PGC, 6-phospho-D-gluconate; RL5P, D-ribulose 5-phosphate; R5P, D-ribose 5-phosphate; X5P, D-xylulose 5-phosphate; S7P, D-sedoheptulose 7-phosphate; E4P, D-erythrose 4-phosphate; AC-CoA, acetyl-CoA; FA-CoA, fatty acyl-CoA; Mal-CoA, malonyl-CoA.

* Corresponding authors at: Monash University, Clayton, Victoria 3800, Australia. Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

*E-mail addresses:* yan.zhu@monash.edu (Y. Zhu), daizj@tib.cas.cn (Z. Dai).

https://doi.org/10.1016/j.csbj.2022.05.018

**Table 1**
Comparison of *i*Yli21 with previous *Y. lipolytica* models.

| Model | *i*YL619_PCP | *i*NL895 | *i*MK735 | *i*Yali4 | *i*YL_2.0 | *i*YLI647 | *i*Yli21 |
|---|---|---|---|---|---|---|---|
| Year | 2012 | 2012 | 2015 | 2016 | 2017 | 2018 | 2022 |
| No. of Reactions | 1,142 | 2,002 | 1,336 | 1,985 | 1,471 | 1,347 | 2,285 |
| No. of Metabolites | 849 | 1,847 | 1,111 | 1,683 | 1,083 | 1,119 | 1,868 |
| No. of Genes | 596 | 895 | 735 | 901 | 645 | 647 | 1,058 |
| Reference Genome | CLIB122 | CLIB122 | CLIB122 | CLIB122 | CLIB122 | CLIB122 | W29 |

strategies in many prokaryotes and eukaryotes [11-18]. To date, 6 GSMMs have been developed for *Y. lipolytica* [19-24]; none of them were constructed for type strain W29, except model *i*Yali4; and the previous models were validated for a limited number of carbon sources [6,25,26]. Additionally, most flux calculations were purely based on metabolic network [20,24] without any further constraints on either enzyme abundance or activity. Taken together, an experimentally validated, high-quality model is urgently required for strain W29 to accurately predict its metabolic phenotypes.

In this study, with extensive manual curation using Biolog experimental data we developed a high-quality GSMM *i*Yli21 for strain W29, which showed a high accuracy of 85.7% in predicting nutrient utilization. It was then integrated with transcriptomic data to delineate cellular metabolism and analyze gene essentiality. The maximum theoretical biochemical yields were predicted using various carbon sources. The model will be a promising tool in assisting metabolic engineering of this important yeast cell factory.

## 2. Experimental procedures

### 2.1. Strain and media

*Y. lipolytica* W29 was stocked in Yeast Extract Peptone Dextrose (YPD) media (10 $g{\cdot}L^{-1}$ yeast extract, 20 $g{\cdot}L^{-1}$ peptone, 20 $g{\cdot}L^{-1}$ D-glucose) with 20% glycerol at −80℃. Before use, strain W29 was sub-cultured on YPD agar plate at 30℃ overnight. For Biolog nutrient utilization assay, synthetic minimal media (7.5 $g{\cdot}L^{-1}$ ammonium sulfate, 14.4 $g{\cdot}L^{-1}$ potassium dihydrogen phosphate, 0.5 $g{\cdot}L^{-1}$ magnesium sulfate heptahydrate, 20 $g{\cdot}L^{-1}$ glucose, trace metal and vitamin solution, pH 6.0) was used [27,28]. For growth assay, W29 was cultured at 30℃ in 250 mL flasks with 50 mL synthetic minimal media, 3 mL fermentation broth was sampled to measure dry cell weight and extracellular metabolite concentrations. The specific glucose uptake rate and growth rate were calculated using previous methods [28].

### 2.2. Biolog assay

Biolog Phenotype Microarrays (PM1-2, Biolog, Hayward, CA, USA) were employed to examine the utilization of 190 carbon sources with OmniLog incubator (Biolog, Hayward, CA, USA). Microplates were incubated at 30℃ for 24 h with measurement of the optical density at 595 nm every 15 min. Growth curves were recorded and the maximum specific growth rate ($\mu_{max}$) was estimated using R package growthrates [29].

### 2.3. Genome-scale metabolic model

Previous GSMM *i*Yali4 was utilized as a template to expedite model construction [24]. Notably, *i*Yali4 was constructed for strain W29, but it incorrectly used genome annotation of another strain CLIB122 [24]. To correct this error, the CLIB122 genes in model *i*Yali4 were then replaced with their corresponding W29 homologs identified by reciprocal BLASTp (bit-score $\geq$ 200, coverage $\geq$ 50%). Missing W29 reactions were then added to the draft model according to genome annotation, literature and BlastKOALA prediction results [30-32]. BlastKOALA is a tool based on KEGG Ontology and it was used to predict all possible biochemical reactions in W29 cell based on genome annotation. The resulting model was designated *i*Yli21 according to model naming conventions [33].

### 2.4. Flux balance analysis

Flux balance analysis (FBA) was employed to predict aerobic growth on Biolog nutrients with the uptake rate of each nutrient empirically set to 2.43 $mmol{\cdot}gDW^{-1}{\cdot}h^{-1}$ [34].

$$\max v_{biomass}$$

$$s.t. \quad \mathbf{S}\cdot\mathbf{v}=0$$

$$a_j \leq v_j \leq b_j, j=1,2,\cdots,n$$

where stoichiometric matrix **S** has $m$ rows (metabolites) and $n$ columns (reactions). Each flux $v_j$ is constrained by its lower bound $a_j$ and upper bound $b_j$. Correct predictions of growth (true positive, TP) or non-growth (true negative, TN), and incorrect predictions of growth (false positive, FP) or non-growth (false negative, FN) were used to calculate Matthews Correlation Coefficient (MCC) as previously described [18]. The false predictions (FP and FN) were employed for model manual curation with the reference of KEGG [15]. To analyze biochemical production using various carbon sources, objective function of *i*Yli21 was changed to maximizing secretion of citrate, erythritol, oleic acid (C18:1), limonene, 3-dehydroshikimate, respectively. Flux-sum is the total turnover rate of a metabolite and was computed using previous method [35].

### 2.5. Integrative modelling with transcriptomic data

A previous study used single-end RNA-seq to examine the differential gene expression of *Y. lipolytica* W29 growing on 6 different carbon sources: glycerol, glucose, oleic acid, hexadecane, triolein (glyceryl trioleate) and tributyrin (glyceryl tributyrate) [36]. The gene expression data were used to constrain model *i*Yli21 for accurate prediction of metabolic fluxes. First, the raw reads (Accession E-MTAB-939) were downloaded from ArrayExpress database and were subsequently quality-filtered, trimmed, and aligned to W29 reference genome (GenBank Accession GCA_001761485.1) using RSEM [37]. Secondly, the read counts were summarized and the RPKM (Reads Per Kilo base per Million reads) values were calculated for all the genes. Thirdly, RIPTiDe was used to constrain reaction fluxes of *i*Yli21 using the calculated RPKM values [38]. Metabolic solution space was randomly sampled with 500 points for each nutrient condition. *In silico* single-gene deletion [34] was conducted to determine the condition-specific gene essentiality. Finally, the metabolic flux differences under varying nutrient conditions were identified and visualized using R packages.

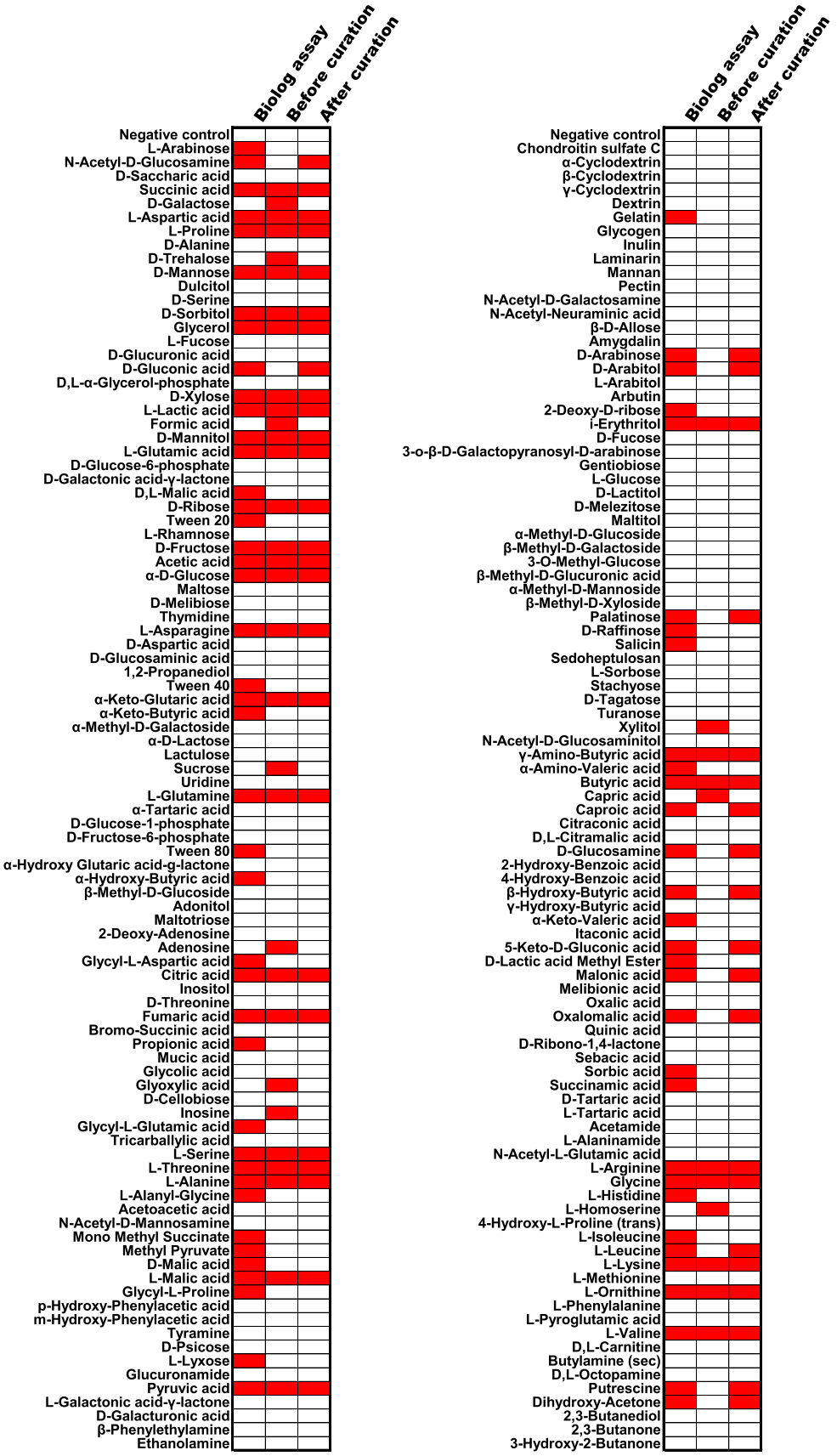


**Fig. 1.** Comparison of Biolog assay (left columns) and model predictions before (middle columns) and after curation (right columns). Red indicates either growth predicted by model *i*Yli21 or nutrient utilization examined by Biolog assays) while white indicates neither growth nor nutrient utilization. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 3. Results

### 3.1. Development of a genome-scale metabolic model for type stain W29

Despite being constructed for strain W29, previous model *i*Yali4 used the genome annotation of another strain CLIB122 [24]. According to NCBI annotation, the genomes of CLIB122 (GCA_000002525) and W29 (GCA_001761485) contain 6,471 and 7,949 protein coding genes, respectively; between them only 4,477 genes (69.2% for CLIB122, and 56.3% for W29) encode 100% identical protein sequences. Model *i*Yali4 contains 901 CLIB122 genes (Table 1). Among them, 886 (98.1%) were substituted with their corresponding W29 homologs while using *i*Yali4 as a template to construct W29-specific model in the present study. CLIB122 unique genes and their associated reactions were also removed from the draft model. For instance, sucrose 6-phosphate hydrolase gene YALI0_E26719g is present in CLIB122 genome but absent in W29, it was thus removed in the draft model. The resulting model shows that W29 is unable to utilize sucrose as a sole carbon source for growth, which is consistent with experimental observation [39]. Furthermore, based on literature and KEGG database, 185 metabolites, 300 reactions, 157 genes were additionally incorporated into the draft model, including those from pathways of fatty acid metabolism, glycerolipid metabolism, amino and nucleotide sugar metabolism, and cross-membrane transport (Supplementary Table 1). The biomass composition was inherited from previous experimentally verified *Y. lipolytica* biomass reaction (biomass_C) [40] and the non-growth associated ATP maintenance was set to 7.86 mmol·gDCW$^{-1}$·h$^{-1}$ according to literature [22]. Specifically, modifications of 72 *i*Yali4 reactions in a previous curation effort were incorporated during model construction [40]. After manual check of reaction directionality, reversibility, mass balance and energetic consistency, a draft GSMM was obtained for type strain W29.

**Table 2**
Comparisons of the predicted and experimentally determined growth rates.

| Glucose uptake rate (mmol·gDCW·h$^{-1}$) | Specific growth rate (h$^{-1}$) ** | | |
|---|---|---|---|
| | **Experiment** | ***i*Yli21** | ***i*Yali4** |
| 2.43* | 0.26* | 0.28 | 0.18 |
| 0.61 [24] | 0.047 [24] | 0.031 | 0.020 |
| 0.64 [24] | 0.048 [24] | 0.036 | 0.022 |

*The specific rates were experimentally determined in this study.
**The non-growth associated ATP maintenance was set to 7.86 mmol·gDCW·h$^{-1}$.

### 3.2. Model curation based on Biolog Phenotype Microarrays

Biolog assay revealed that out of 190 carbon sources, 73 carbon sources including carbohydrates, lipids, amino acids and nucleotides could be utilized by strain W29 (**Supplementary** Fig. 1,**2**). Computation using the aforementioned draft W29 model showed zero *in silico* growth on 41 substrates, potentially due to lack of the corresponding transport reactions (33 substrates) or presence of pathway gaps (8 substrates) (Fig. 1). We further added all necessary reactions to enable the cross-membrane transport of nutrients or fill the pathway gaps. For instance, hexanoate and D-arabinose, both of them were untilizable carbon sources in Biolog experiment but the *in silico* biomass productions were zero before model curation. Model curation indicated that hexanoate activation reactions were absent, acyl-CoA synthetase reaction was added to the draft model and hexanoate can be used as sole carbon source.Similarly, an isomerase reaction converting D-arabinose to D-xylulose 5-phosphate was added to enable the utilization of D-arabinose. After curation, the final model showed an overall accuracy of 85.7% (TP = 46) in predicting utilization of Biolog nutrients (Fig. 1). It was then named *i*Yli21 according to GSMM naming conventions [41]. Further predictions of aerobic growth on glucose show excellent consistency with experimental observations (Table 2). Specifically, with the glucose uptake at 0.61, 0.64 and 2.43 mmol·gDCW$^{-1}$·h$^{-1}$, the growth rates predicted using previous model *i*Yali4 are 0.020, 0.022 and 0.18 h$^{-1}$, respectively; whereas these growth rates predicted using model *i*Yli21 are 0.031, 0.036 and 0.28 h$^{-1}$, much closer to the experimentally determined rates of 0.047, 0.048 and 0.26 h$^{-1}$. Model *i*Yli21 contains 1,868 metabolites, 2,285 reactions and 1,058 genes, and the MCC was increased from 0.44 to 0.71 after curation against Biolog data, suggesting a significant improvement in both component quantity and prediction quality. Notably, model *i*Yli21 contains all the major metabolic pathways of *Y. lipolytica* W29 (Fig. 2). Hence, model *i*Yli21 (Supplementary file) is a comprehensive representation of the metabolic network of this type strain.

The cell metabolism of strain W29 predicted using model *i*Yli21 is different with that of strain CLIB122 using *i*Yali4, *i*MK735, *i*YLI647 and *i*YL_2.0. Across the 190 Biolog carbon sources, *i*Yli21

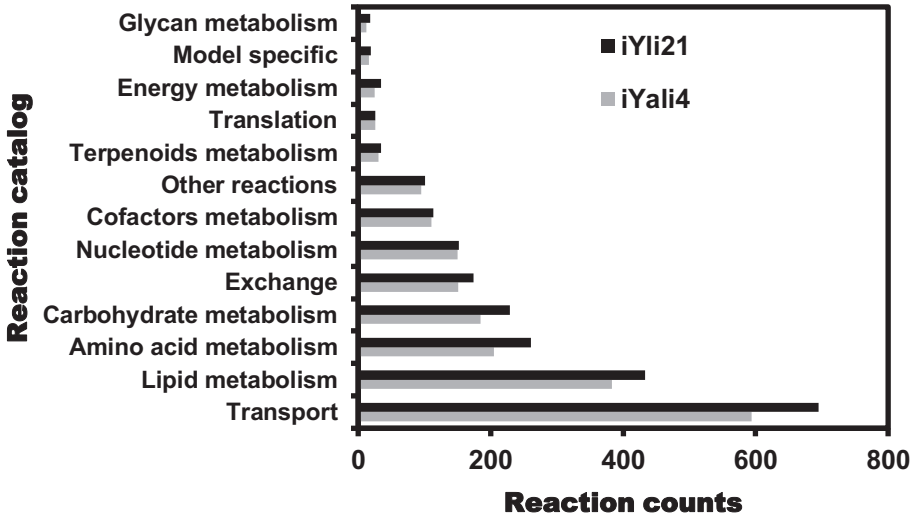


**Fig. 2.** The biochemical reactions involved in *i*Yli21 and *i*Yali4.

predicted that 46 are utilizable, whereas this number is 42, 31, 36 and 30 for *i*Yali4, *i*MK735, *i*YLI647 and *i*YL_2.0, respectively, with 15 substrates in common. The predicted growth rates of W29 grown on these 15 substrates are significantly different with those of CLIB122 (*i*Yli21 *vs* *i*Yali4: $p = 0.0007$, *i*Yli21 *vs* *i*YL_2.0: $p = 0.0114$, Wilcoxon test). The metabolism is also different. For instance, *i*Yli21, *i*Yali4, *i*MK735, *i*YLI647 and *i*YL_2.0 predicted 42, 42, 28, 30 and 36 active pathways for aerobic growth on glucose, with only 18 pathways in common. The unique active pathways of W29 include β-alanine metabolism, propanoate metabolism, and pentose and glucuronate interconversions. Such differences in growth and metabolism may be a result of different quality across models, and/or distinctive metabolic capabilities of W29 compared to CLIB122.

### 3.3. Metabolic changes while utilizing varying carbon sources

*Y. lipolytica* is known for its extraordinary capability of utilizing sugar and lipid substrates. Previous transcriptomic data were integrated with model *i*Yli21 to compare cellular metabolism on polyol (glycerol), sugar (glucose), fatty acid (oleic acid), alkane (hexadecane) and glycerolipids (triolein and tributyrin). Comparisons of predicted fluxomes showed that among totally 516 non-zero fluxes, 302 fluxes were commonly present under all nutrient conditions, whereas 214 fluxes were condition specific (Supplementary Table 2).

Not surprisingly, a complete glycolysis from glucose to pyruvate was used only when glucose was the sole carbon source. Pyruvate was cleaved by mitochondrial pyruvate dehydrogenase to acetyl-CoA, which was further fed to tricarboxylic acid (TCA) cycle to generate NADH and precursors for amino acid and lipid biosynthesis (Fig. 3**A**). When glycerol was the sole carbon source, the upstream glycolytic fluxes from glucose to glyceraldehyde 3-phosphate were reduced to zero (Fig. 3**B**). Glycerol entered in glycolysis (1.83 mmol·gDW·h$^{-1}$) and gluconeogenesis (0.25 mmol·gDW·h$^{-1}$) simultaneously via glyceraldehyde 3-phosphate after phosphorylation and dehydrogenation, respectively. Notably, only when fatty acid compounds were used as sole carbons sources, the cytoplasmic phosphoenolpyruvate carboxykinase (PPCK) flux was activated to start gluconeogenesis. For glycerolipid tributyrin and triolein, glycerol was generated during the initial hydrolysis and then fed to glycolysis and gluconeogenesis as described above. Analysis shows different TCA fluxes under six carbon source conditions. Interestingly, NADH-dependent isocitrate dehydrogenase ICDHxm (Reaction R487) carried no flux during growth on any of oleic acid, triolein and hexadecane, while glyoxylate shunt was utilized. In addition, prediction shows that most TCA reactions were inactive when hexadecane was used as the sole carbon source for growth. The pentose phosphate pathway also shows a distinct flux distribution. Only growth on glucose, hexadecane and oleic acid required the activation of oxidative branch (glucose 6-phosphate dehydrogenase [G6PDH2r, R325], 6-phosphogluconolactonase [PGL, R71], and 6-phosphogluconate dehydrogenase [GND, R639]) to generate NADPH as redox power for biosynthesis. Growth with all six substrates relies on transaldolase TALA (Reaction R764), transketolase TKT1 and TKT2 (Reactions R765 and R766, respectively) in non-oxidative branch, and ribulose 5-phosphate isomerase RPI (Reaction R712) to generate phosphoribosyl pyrophosphate for nucleotide biosynthesis. Fatty acid β-oxidation was commonly used to catabolize hydrocarbon (hexadecane) and lipids (oleic acid, tributyrin and triolein) to acetyl-CoA and NADH. Overall, *Y. lipolytica* showed different metabolism under varying nutrient conditions based on model simulation.

### 3.4. Essential genes for growth

Model *i*Yli21 was subsequently employed to predict gene essentiality during aerobic growth on the aforementioned six carbon sources (Fig. 4, Supplementary Table 3). Overall, 211–227 essential genes were identified; among them, 183 genes from central metabolism were commonly discovered across all conditions. 156 genes were identified under no more than five conditions, suggesting that they are conditionally essential for growth.

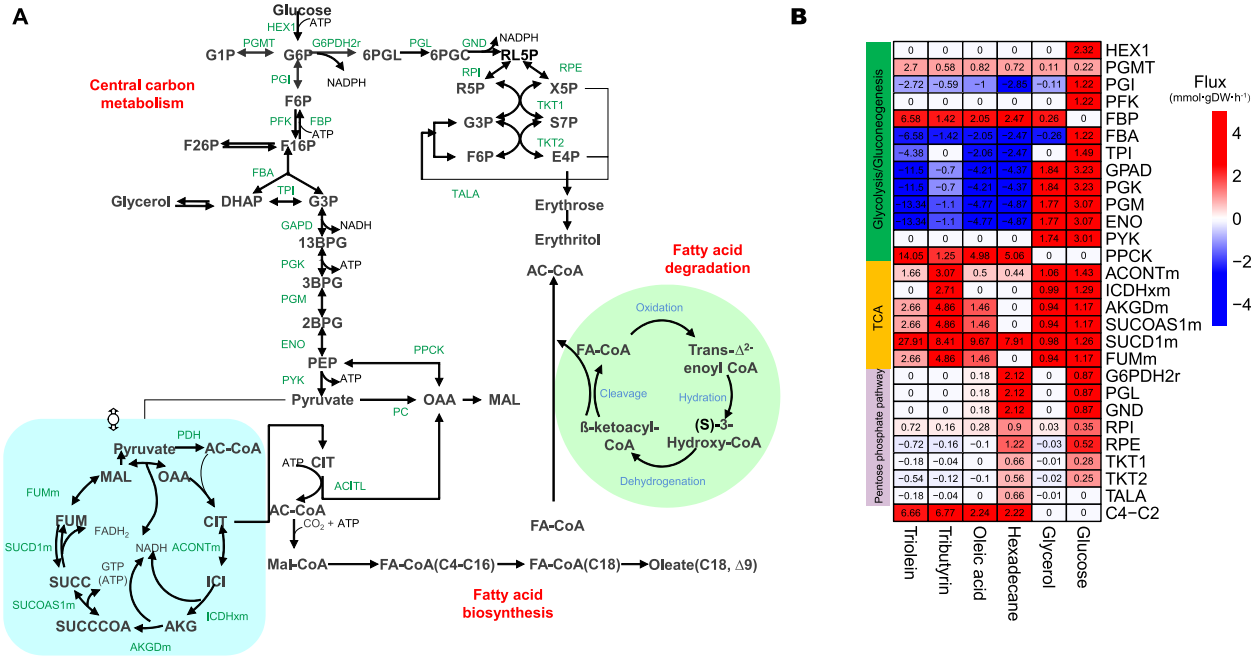


**Fig. 3.** Metabolic variations of *Y. lipolytica* W29 grown on six carbon nutrients. (**A**) Central metabolism including glycolysis/gluconeogenesis, pentose phosphate pathway, mitochondrial TCA cycle (blue shading) and fatty acid biosynthesis (green shading). (**B**) Significantly altered metabolic fluxes under six carbon nutrient conditions shown in mean values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

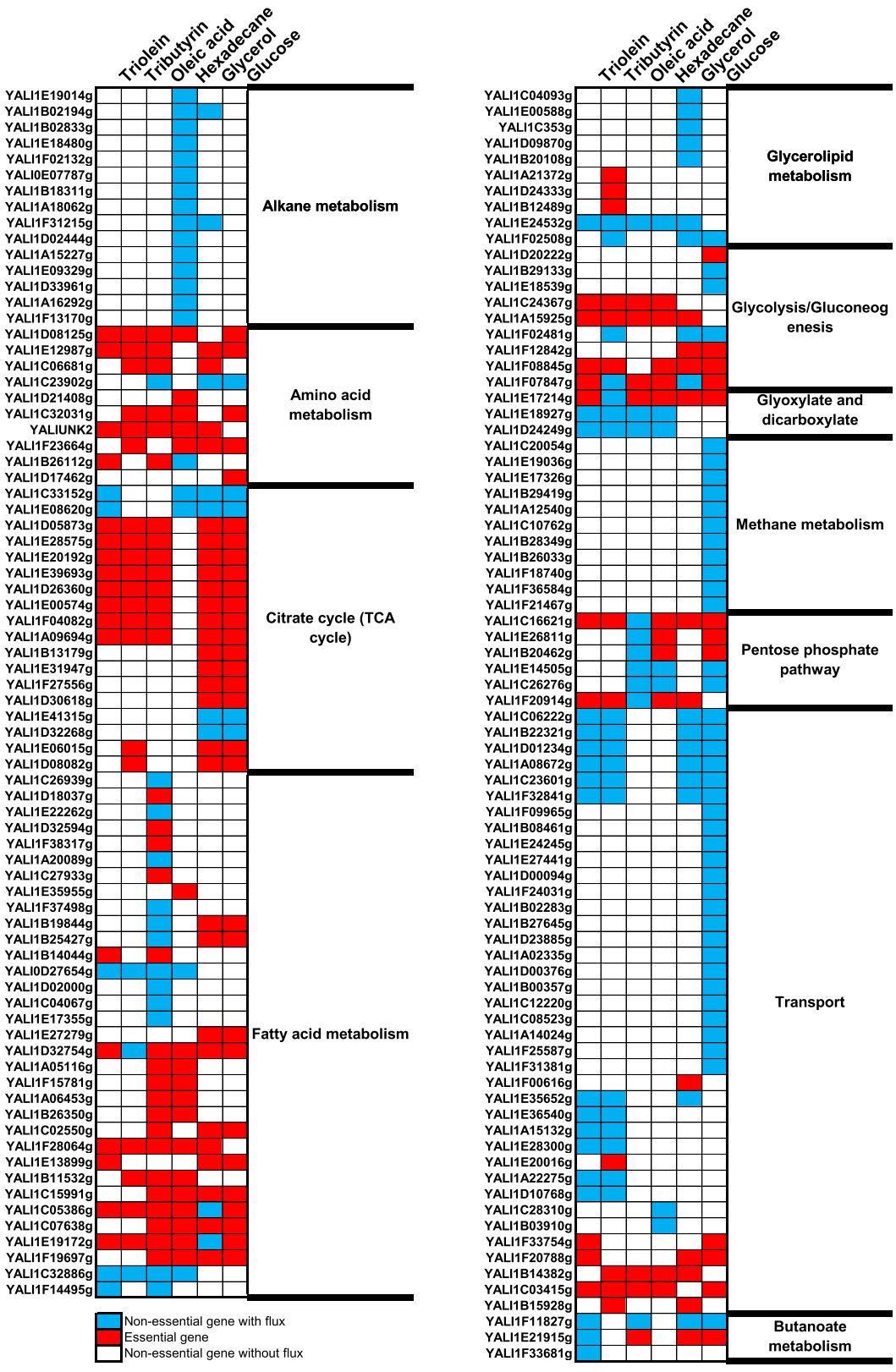


**Fig. 4.** Essential genes predicted under six carbon source conditions. The common essential gene and common non-essential genes of six conditions are not displaying. Red indicates essential genes, while blue indicates non-essential genes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For instance, the genes encoding pyruvate dehydrogenase complex (YALI1B13179g, YALI1E31947g, YALI1F27556g, YALI1D30618g, Reaction R693) are essential only for growth on glucose or glycerol because digestion of either metabolite requires glycolysis and pyruvate dehydrogenase mediated pyruvate cleavage to produce acetyl-CoA in mitochondria [42]. Gluconeogenetic

**Table 3**
Flux-sum analysis of major energy and redox cofactors in *i*Yli21. Values are shown as mean (mmol·gDCW$^{-1}$·h$^{-1}$).

| C ofactor | Triolein | Tributyrin | Oleic acid | Hexadecane | Glycerol | Glucose |
|---|---|---|---|---|---|---|
| ATP | 207.73 | 76.26 | 73.70 | 69.57 | 22.88 | 28.39 |
| ADP | 204.49 | 72.90 | 72.61 | 68.49 | 22.88 | 28.40 |
| AMP | 13.63 | 10.94 | 4.40 | 4.35 | 0.14 | 0.29 |
| NADH | 165.85 | 34.18 | 56.34 | 50.22 | 11.32 | 12.96 |
| NADPH | 10.27 | 5.18 | 3.13 | 7.64 | 2.47 | 2.02 |
| Acetyl-CoA | 120.10 | 27.11 | 40.32 | 35.66 | 1.44 | 2.30 |

genes (phosphoenolpyruvate carboxykinase, YALI1C24367g, Reaction R634, and fructose-bisphosphatase, YALI1A15925g, Reaction R312) are only essential for growth on glycerol and lipid substrates as sugars required for cell envelope and nucleotide biosynthesis can only be generated from gluconeogenesis.

On the other hand, 124–146 genes were categorized as non-essential; among them, 102 non-essential genes are commonly identified across all nutrient conditions, suggesting that they either encode dispensable metabolic functions, or can be replaced by alternative genes. For example, the two hexokinases (YALI1B29133g and YALI1E18539g, Reaction R387) are non-essential with glucose as sole carbon source but are the key enzymes in glucose metabolism. The ATP-citrate lyase genes (ACITL, encoded by YALI1E41315g or YALI1D32268g, R1894) are also non-essential under glucose or glycerol nutrient conditions. The ATP-citrate lyase catalyzes the cytosolic cleavage of citrate to yield oxaloacetic acid and acetyl-CoA [43], while acetyl-CoA is a vital metabolite for biosynthesis of fatty acids, an essential component for biomass formation. The malate synthase (YALI1E18927g, YALI1D24249g, Reaction R536) of glyoxylate shunt is non-essential during growth on tributyrin, triolein, glycerol and hexadecane. The glyoxylate shunt is critical for the fatty acid to enter gluconeogenesis. These genes are non-essential as they are isozymes and functionally replaceable by each other.

### 3.5. Evaluating production yields of biochemicals

*Y. lipolytica* produces industrially valuable biochemicals using a variety of substrates [3,4]. A systematic evaluation of substrates from the production yield perspective is critical for rational design of cell factory. Using model *i*Yli21, we calculated the maximum theoretical yields of citrate (organic acid), erythritol (functional sugar alcohol), oleic acid (unsaturated fatty acid), limonene (terpene) and 3-dehydroshikimate (aromatics) production from the selected 46 carbon sources (Fig. 5**A**). Among these carbon sources, hexanoate (caproic acid in Biolog) shows the highest theoretical yields of 1.64, 0.51, 1.19, 0.96 and 0.47 g/g in production of citrate, oleic acid, erythritol, 3-dehydroshikimate and limonene, respectively. Whereas L-glycine, an amino acid substrate, has the lowest theoretical yields of 0.18, 0.05, 0.15, 0.13, 0.04 g/g in producing the above 5 biochemicals (Fig. 5**A**).

Specifically, when hexanoate was the substrate for production of citrate, it was degraded via β-oxidation to produce NADH and acetyl-CoA (Fig. 5**B**). When the uptake rate of hexanoate is 10 mmol·gDW$^{-1}$·h$^{-1}$, a total of 81 mmol·gDW$^{-1}$·h$^{-1}$ ATP were generated, with 7.86 mmol·gDW$^{-1}$·h$^{-1}$ consumed by non-growth associated maintenance, 10 mmol·gDW$^{-1}$·h$^{-1}$ consumed by the fatty acid activation (acyl-CoA synthetase, Reaction R1898), and the rest 63.14 mmol·gDW$^{-1}$·h$^{-1}$ used for creating a cyclic loop of ATP by Reaction R121 and R603. Hence, hexanoate can produce ample energy and substrate for citrate synthesis (citrate synthase, Reaction R237) compared to other substrates without carbon loss. As a contrast, degradation of L-glycine for citrate production only produces ATP at a rate of 28 mmol·gDW$^{-1}$·h$^{-1}$; 80% L-glycine was fed to TCA, resulting in a significant loss of carbon via $CO_2$ emission the lowest production yields for target biochemicals.

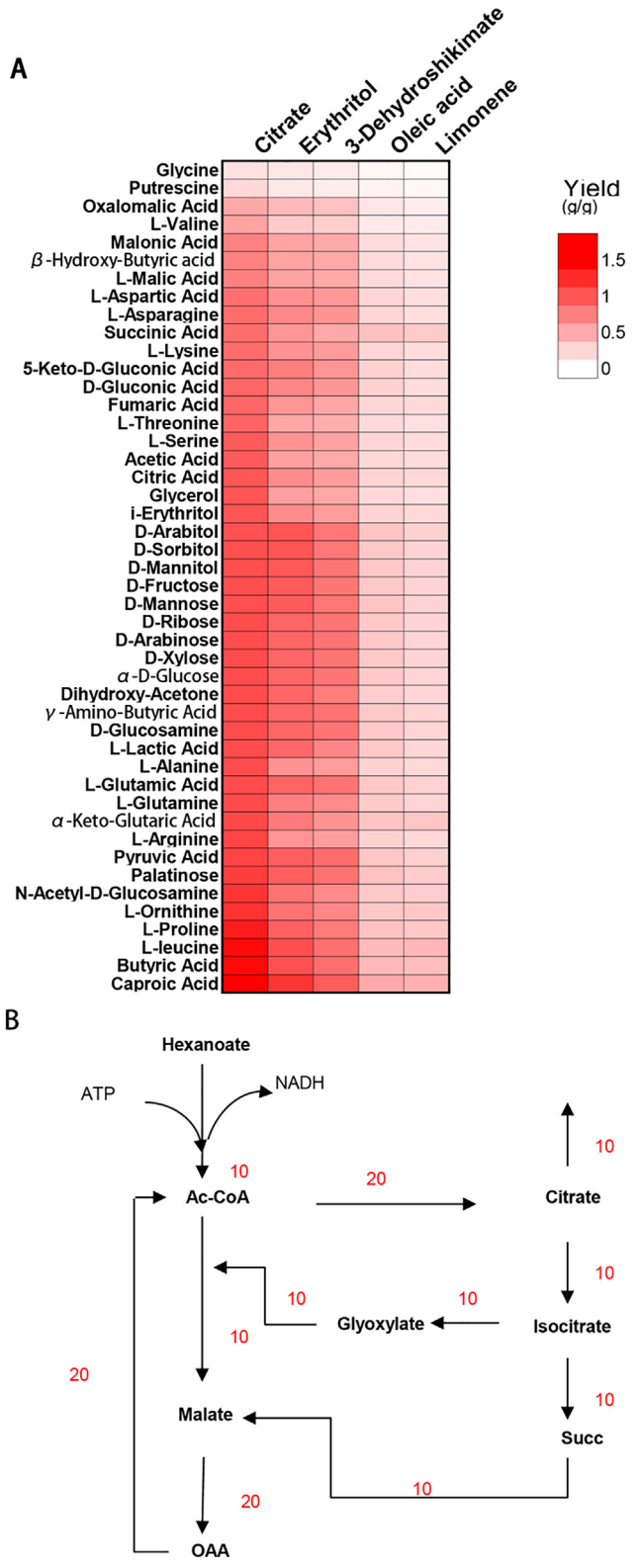


**Fig. 5.** Predicted production of targeted biochemicals using the selected substrates. (**A**) The maximum theoretical yields of six biochemicals using 46 selected carbon sources. (**B**) The major routes of producing citrate from hexanoate. Reaction fluxes are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Generating different end products from the same substrate shows significantly different yields. When using glucose as substrate to produce citrate, the energy consumption includes non-growth associated maintenance and pyruvate carboxylation, which is completely covered by the energy generated by glycolysis. As a result, the maximal yield of citrate from glucose could reach a high level of 1.05 g/g. Whereas for erythritol biosynthesis from glucose, TCA cycle is essential for compensate the energy consumption from generating the precursor F6P and G3P, which led to carbon loss, resulting in a theoretical yield of 0.9 g/g, lower than that of citrate.

Overall, the high-quality GSMM *i*Yli21 for type strain W29 can be used to calculate the maximum theoretical yields of the targeted end products from many substrates and may guide rational design of *Y. lipolytica* cell factory.

## 4. Discussion

*Y. lipolytica* has emerged as an important non-model yeast host for industrial production of many high-value biochemicals including citrate, erythritol, mannitol and lipid from a variety of cheap biomass substrates [4]. This industrial microbe also exhibits a high tolerance to manufacturing stress and an extraordinary capability of utilizing waste materials such as aromatic compounds and urea as substrates [44,45]. However, limited understanding of the complex metabolism has significantly hampered the use of this valuable cell factory. Therefore, we constructed a high-quality GSMM *i*Yli21 for *Y. lipolytica* type strain W29 by extensive manual curation with Biolog experimental results (Fig. 1). Importantly, we have, for the first time, integrated transcriptomic data as constraints to accurately calculate the metabolic changes in response to different carbon nutrients; and systematically compared the production yields using a variety carbon sources as substrates.

Thus far, six GSMMs (*i*YL619_PCP, *i*NL895, *i*MK735, *i*Yali4, *i*YL_2.0, and *i*YLI647), have been developed to delineate the metabolism of *Y. lipolytica* [19-24]; none of them were constructed specifically for strain W29. Models *i*YL619_PCP and *i*NL895 represent the first trials of metabolic modelling for this important oleaginous yeast in 2012. Model *i*MK735 was released in 2015 and it was constructed mostly based on *S. cerevisiae* model *i*ND750 [46]. *i*YL_2.0 is an updated version based on *i*YL619_PCP and *i*YLI647 was derived by integrating components from other models on top of the scaffold *i*MK735. Model *i*Yali4 is virtually a model of another strain because it used CLIB122 genome annotation (Table 1). CLIB122 was derived from a cross between isolates from a Paris sewer (W29) and an American corn processing plant (ATCC 18942) [8]. Hence, the metabolism of these two strains might have significant differences. Indeed, compared to CLIB122, the W29 genome encodes unique enzymes in glycine, serine and threonine metabolism, magnesium transport, and starch and sucrose metabolism. In the present study, we have developed a model *i*Yli21 specific for W29. In addition, the model covers most metabolic pathways in KEGG database (Fig. 2) and has 17%, 11% and 15% increment of genes, metabolites, and reactions, respectively, compared to the previously largest *Y. lipolytica* model *i*Yali4 (Table 1); it represents the most comprehensive GSMM to date for *Y. lipolytica* W29.

Metabolic engineering of *Y. lipolytica* for biochemical production requires comprehensive understanding of several critical metabolic pathways, especially generation of acetyl-CoA (building blocks) and NADPH (redox power). With transcriptomic constraints, we calculated the metabolic flux distributions under six individual carbon sources using model *i*Yli21 (Table 3). The results show that acetyl-CoA is mainly generated via mitochondrial pyruvate dehydrogenase, cytosolic ATP citrate lyase, and lipid β-oxidation (Fig. 3**A**). The former two contribute 88% to overall acetyl-CoA production when grown on glycerol and glucose, while β-oxidation contributes 50% exclusively when grown on lipids. Furthermore, NADPH generation is mainly via the pentose phosphate pathway, NADP-dependent isocitrate dehydrogenase, glycerol dehydrogenase and alkane assimilation. Notably, NADPH production from glycerol exceeds that from glucose because oxidation of 1 mol glycerol (via glycerol dehydrogenase) can generate 1 extra mol NADPH before it enters into glycolysis. Hence, glycerol or glycerol-containing substrate is more preferable for producing a chemical of high reduction degree such as fatty acids.

TCA cycle is a part of central carbon metabolism for energy and precursor supply. However, when using triolein, oleic acid and hexadecane as sole carbon sources, the TCA cycle is not fully activated. The metabolic shift from TCA to the glyoxylate shunt is possibly due to excessive NADH production from β-oxidation. The oxidative part of pentose phosphate pathway is inactivated during growth on triolein, tributyrin and glycerol, and part of NADPH is generated via glycerol dehydrogenase.

Further simulations were conducted to analyze productions of erythritol, citrate, 3-dehydroshikimate, limonene and oleic acid with 46 carbon sources. With glucose as the sole carbon source, the maximum yield of citrate is 1.05 g/g (Fig. 5**A**), even larger than the highest reported value thus far (0.90 g/g) [47]. This is because the model calculation assumes the complete use of glucose in production of citrate without any expense on biomass accumulation; whereas during batch fermentation, part of glucose was inevitably consumed for cell growth and by-product formation, which results in a relatively low yield of citrate. Similarly, the maximum yields of erythritol with glycerol and glucose as substrate are 1.0 g/g and 0.90 g/g, respectively, larger than the highest values (0.77 g/g and 0.37 g/g) previously reported [48,49].

## 5. Conclusions

A GSMM *i*Yli21 for *Y. lipolytica* type strain W29 was built, which contained more genes, metabolites and reactions than any other published *Y. lipolytica* GSMM, and showed a high accuracy in predicting nutrient utilization. This model was used to delineate cellular metabolism, analyze gene essentiality and predict biochemicals production yields on various carbon sources, indicating that it is a high-quality model for *Y. lipolytica* and will be a powerful tool for cell factory design in the future.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by National Key Research and Development Program of China (2021YFA0910600), National Natural Science Foundation of China (32071423, 32161133019), Hundreds of Talents Program of the Chinese Academy of Sciences (Y0J51009) and Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIP-CXRC-002).

## Author contributions

Zongjie Dai and Yan Zhu designed the study. Yufeng Guo performed computations. Yufeng Guo, Liqiu Su and Qi Liu analyzed the data. Yufeng Guo, Zongjie Dai and Yan Zhu wrote the manuscript. Yufeng Guo prepared the dataset and revised the manu-

script. Yufeng Guo, Zongjie Dai, Yan Zhu and Qinhong Wang revised the manuscript.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.05.018.

## References


[1] Kosa M, Ragauskas AJ. Lipids from heterotrophic microbes: advances in metabolism research. Trends Biotechnol 2011;29:53–61.
[2] Jin M, Slininger PJ, Dien BS, Waghmode S, Moser BR, Orjuela A, et al. Microbial lipid-based lignocellulosic biorefinery: feasibility and challenges. Trends Biotechnol 2015;33:43–54.
[3] Markham KA, Alper HS. Synthetic Biology Expands the Industrial Potential of *Yarrowia lipolytica*. Trends Biotechnol 2018;36:1085–95.
[4] Ledesma-Amaro R, Nicaud JM. Metabolic Engineering for Expanding the Substrate Range of *Yarrowia lipolytica*. Trends Biotechnol 2016;34:798–809.
[5] Bilal M, Xu S, Iqbal HMN, Cheng H. *Yarrowia lipolytica* as an emerging biotechnological chassis for functional sugars biosynthesis. Crit Rev Food Sci Nutr 2021;61:535–52.
[6] da Veiga MJ, Jolicoeur M, Schwartz L, Peres S. Fine-tuning mitochondrial activity in *Yarrowia lipolytica* for citrate overproduction. Sci Rep 2021;11:878.
[7] Madzak C, Gaillardin C, Beckerich JM. Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. J Biotechnol 2004;109:63–81.
[8] Madzak C. Yarrowia lipolytica Strains and Their Biotechnological Applications: How Natural Biodiversity and Metabolic Engineering Could Contribute to Cell Factories Improvement. J Fungi (Basel) 2021;7.
[9] Zhang G, Wang H, Zhang Z, Verstrepen KJ, Wang Q, Dai Z. Metabolic engineering of *Yarrowia lipolytica* for terpenoids production: advances and perspectives. Crit Rev Biotechnol 2021:1–16.
[10] Shan L, Dai Z, Wang Q. Advances and opportunities of CRISPR/Cas technology in bioengineering non-conventional yeasts. Front Bioeng Biotechnol 2021;9:765396.
[11] Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. C4GEM, a genome-scale metabolic model to study C4 plant metabolism. Plant Physiol 2010;154:1871–85.
[12] Moreira TB, Shaw R, Luo X, Ganguly O, Kim HS, Coelho LGF, et al. A Genome-Scale Metabolic Model of Soybean (Glycine max) Highlights Metabolic Fluxes in Seedlings. Plant Physiol 2019;180:1912–29.
[13] Kim D, Kim Y, Yoon SH. Development of a Genome-Scale Metabolic Model and Phenome Analysis of the Probiotic *Escherichia coli* Strain Nissle 1917. Int J Mol Sci 2021;22.
[14] Rynearson KD, Ponnusamy M, Prikhodko O, Xie Y, Zhang C, Nguyen P, et al. Preclinical validation of a potent γ-secretase modulator for Alzheimer's disease prevention. J Exp Med 2021;218.
[15] Lu H, Li F, Sanchez BJ, Zhu Z, Li G, Domenzain I, et al. A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. Nat Commun 2019;10(3586).
[16] Zhu Y, Czauderna T, Zhao J, Klapperstueck M, Maifiah MHM, Han ML, et al. Genome-scale metabolic modeling of responses to polymyxins in *Pseudomonas aeruginosa*. GigaScience 2018;7.
[17] Zhu Y, Zhao J, Maifiah MHM, Velkov T, Schreiber F, Li J. Metabolic Responses to Polymyxin Treatment in Acinetobacter baumannii ATCC 19606: Integrating Transcriptomics and Metabolomics with Genome-Scale Metabolic Modeling. mSystems 2019:4.
[18] Zhao J, Zhu Y, Han J, Lin YW, Aichem M, Wang J, et al. Genome-Scale Metabolic Modeling Reveals Metabolic Alterations of Multidrug-Resistant *Acinetobacter baumannii* in a Murine Bloodstream Infection Model. Microorganisms 2020;8.
[19] Loira N, Dulermo T, Nicaud J-M, Sherman DJ. A genome-scale metabolic model of the lipid-accumulating yeast *Yarrowia lipolytica*. BMC Syst Biol 2012;6:35.
[20] Mishra P, Lee NR, Lakshmanan M, Kim M, Kim BG, Lee DY. Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica*. BMC Syst Biol 2018;12:12.
[21] Pan P, Hua Q. Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, *Yarrowia lipolytica*. PLoS One 2012;7:e51535.
[22] Wei S, Jian X, Chen J, Zhang C, Hua Q. Reconstruction of genome-scale metabolic model of *Yarrowia lipolytica* and its application in overproduction of triacylglycerol. Bioresour Bioprocess 2017;4.
[23] Kavscek M, Bhutada G, Madl T, Natter K. Optimization of lipid production with a genome-scale model of *Yarrowia lipolytica*. BMC Syst Biol 2015;9:72.
[24] Kerkhoven EJ, Pomraning KR, Baker SE, Nielsen J. Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. NPJ Syst Biol Appl 2016;2:16005.
[25] Beopoulos A, Chardot T, Nicaud JM. *Yarrowia lipolytica*: A model and a tool to understand the mechanisms implicated in lipid accumulation. Biochimie 2009;91:692–6.
[26] Kim M, Park BG, Kim EJ, Kim J, Kim BG. *In silico* identification of metabolic engineering strategies for improved lipid production in *Yarrowia lipolytica* by genome-scale metabolic modeling. Biotechnol Biofuels 2019;12:187.
[27] Verduyn C, Postma E, Scheffers WA, Van Dijken JP. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 1992;8:501–17.
[28] Dai Z, Huang M, Chen Y, Siewers V, Nielsen J. Global rewiring of cellular metabolism renders *Saccharomyces cerevisiae* Crabtree negative. Nat Commun 2018;9:3059.
[29] Hall BG, Acar H, Nandipati A, Barlow M. Growth rates made easy. Mol Biol Evol 2014;31:232–8.
[30] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2016;45:D353–61.
[31] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.
[32] Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 2016;428:726–31.
[33] Pham N., van Heck RGA, van Dam JCJ, Schaap PJ, Saccenti E., Suarez-Diez M. Consistency, Inconsistency, and Ambiguity of Metabolite Names in Biochemical Databases Used for Genome-Scale Metabolic Modelling. *Metabolites* 2019; 9.
[34] Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol 2013;7:74.
[35] Chung BK, Lee DY. Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. BMC Syst Biol 2009;3:117.
[36] Kabran P, Rossignol T, Gaillardin C, Nicaud JM, Neuveglise C. Alternative splicing regulates targeting of malate dehydrogenase in *Yarrowia lipolytica*. DNA Res 2012;19:231–44.
[37] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 2011;12:323.
[38] Jenior ML, Moutinho Jr TJ, Dougherty BV, Papin JA. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. PLoS Comput Biol 2020;16:e1007099.
[39] Nicaud JM, Fabre E, Gaillardin C. Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. Curr Genet 1989;16:253–60.
[40] Xu Y, Holic R, Hua Q. Comparison and Analysis of Published Genome-scale Metabolic Models of *Yarrowia lipolytica*. Biotechnol Bioprocess Eng 2020;25:53–61.
[41] Carey MA, Drager A, Beber ME, Papin JA, Yurkovich JT. Community standards to facilitate development and address challenges in metabolic modeling. Mol Syst Biol 2020;16:e9235.
[42] Hoshi M, Takashima A, Noguchi K, Murayama M, Sato M, Kondo S, et al. Regulation of mitochondrial pyruvate dehydrogenase activity by tau protein kinase I/glycogen synthase kinase 3beta in brain. Proc Natl Acad Sci USA 1996;93:2719.
[43] Chypre M, Zaidi N, Smans K. ATP-citrate lyase: a mini-review. Biochem Biophys Res Commun 2012;422:1–4.
[44] Brabender M, Hussain MS, Rodriguez G, Blenner MA. Urea and urine are a viable and cost-effective nitrogen source for *Yarrowia lipolytica* biomass and lipid accumulation. Appl Microbiol Biotechnol 2018;102:2313–22.
[45] Bankar AV, Kumar AR, Zinjarde SS. Environmental and industrial applications of *Yarrowia lipolytica*. Appl Microbiol Biotechnol 2009;84:847–65.
[46] Duarte NC, Herrgard MJ, Palsson BO. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 2004;14:1298–309.
[47] Fu GY, Lu Y, Chi Z, Liu GL, Zhao SF, Jiang H, et al. Cloning and Characterization of a Pyruvate Carboxylase Gene from *Penicillium rubens* and Overexpression of the Genein the Yeast *Yarrowia lipolytica* for Enhanced Citric Acid Production. Mar Biotechnol (NY) 2016;18:1–14.
[48] Mirończuk AM, Dobrowolski A, Rakicka M, Rywińska A, Rymowicz W. Newly isolated mutant of *Yarrowia lipolytica* MK1 as a proper host for efficient erythritol biosynthesis from glycerol. Process Biochem 2015;50:61–8.
[49] Mironczuk AM, Kosiorowska KE, Biegalska A, Rakicka-Pustulka M, Szczepanczyk M, Dobrowolski A. Heterologous overexpression of bacterial hemoglobin VHb improves erythritol biosynthesis by yeast *Yarrowia lipolytica*. Microb Cell Fact 2019;18:176.