

RESEARCH

Open Access



# Reconstruction of genome-scale metabolic model of *Yarrowia lipolytica* and its application in overproduction of triacylglycerol

Songsong Wei<sup>1</sup>, Xingxing Jian<sup>1</sup>, Jun Chen<sup>1</sup>, Cheng Zhang<sup>1,2\*</sup> and Qiang Hua<sup>1,3\*</sup>

## Abstract

**Background:** *Yarrowia lipolytica* is widely studied as a non-conventional model yeast owing to the high level of lipid accumulation. Therein, triacylglycerol (TAG) is a major component of liposome. In order to investigate the TAG biosynthesis mechanism at a systematic level, a novel genome-scale metabolic model of *Y. lipolytica* was reconstructed based on a previous model iYL619\_PCP published by our lab and another model iYali4 published by Kerkhoven et al.

**Results:** The novel model iYL\_2.0 contains 645 genes, 1083 metabolites, and 1471 reactions, which was validated more effective on simulations of specific growth rate. The precision of 29 carbon sources utilities reached up to 96.6% when simulated by iYL\_2.0. In minimal growth medium, 111 genes were identified as essential for cell growth, whereas 66 essential genes were identified in yeast extract medium, which were verified by database of essential genes, suggesting a better prediction ability of iYL\_2.0 in comparison with other existing models. In addition, potential metabolic engineering targets of improving TAG production were predicted by three in silico methods developed in-house, and the effects of amino acids supplementation were investigated based on model iYL\_2.0.

**Conclusions:** The reconstructed model iYL\_2.0 is a powerful platform for efficiently optimizing the metabolism of TAG and systematically understanding the physiological mechanism of *Y. lipolytica*.

**Keywords:** *Yarrowia lipolytica*, Genome-scale metabolic model, iYL\_2.0, Triacylglycerol, Gene-level prediction targets

## Background

*Yarrowia lipolytica* is a non-conventional model oleaginous yeast, which is capable of accumulating substantial amounts of neutral lipids, even above than 50% of its dry cell weight (DCW) (Blazek et al. 2014; Papanikolaou et al. 2002). Today, the chemical industry has been widely developing technology to synthesize multiple kinds of products using oils and fats, such as surfactants, polymers, and biofuels (Ledesma-Amaro and Nicaud 2016). Triacylglycerol (TAG) was discovered as the major constituent of liposome and predominant storage of metabolic energy in microorganism cells, which can also be widely

used as valuable resources for dietary consumption and biofuels in industrial application (Yen et al. 2008).

In recent years, great potential of lipid synthesis in *Y. lipolytica* attracts much attention to study its physiological mechanism and explore its application in biofuels. Many efforts, including fermentation and genetic manipulation strategies, have been made for improving TAG production in *Y. lipolytica* (Papanikolaou and Aggelis 2010). It has been reported that the yield of TAG could reach up to 40% of dry cell weight when using glycerol or industrial fats as carbon source (Papanikolaou and Aggelis 2002). In *Y. lipolytica*, TAG can be biosynthesized followed by the Kennedy pathway, and the delta-9 stearyl-CoA desaturase (SCD) was identified as a rate-limiting enzyme in biosynthesis pathway (Dulermo and Nicaud 2011; Kennedy 1957). Besides, genes YALI0E32769g (DGA1) and YALI0E16797g (LRO1), encoding major triacylglycerol synthases, play important

\*Correspondence: cheng.zhang@scilifelab.se; qhua@ecust.edu.cn

<sup>1</sup> State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, People's Republic of China

Full list of author information is available at the end of the article

roles in the biosynthesis pathway (Athenstaedt 2011). Additionally, the repression of TGL3 and TGL4, encoding triacylglycerol lipases involved in the degradation of TAG, could lead to high levels of TAG accumulation (Morin et al. 2011). Although much improvement of TAG production and lipid accumulation in *Y. lipolytica* has been achieved, it is still necessary to rationally discover possible limiting factors involved in TAG biosynthesis at a systems level, and in silico analyze its metabolic characteristics under a wide range of genotypic and environmental conditions.

A genome-scale metabolic model (GSMM) is a powerful platform for investigation of microorganism's metabolic processes and rational strain design, as a GSMM may represent the metabolic genotype–phenotype relationships in the microorganism (Feist et al. 2009). GSMMs have been widely used in guidance of strain design, analysis of complex biological phenomena, and contextualization of high-throughput data (Jian et al. 2016a; Loira et al. 2012; Oberhardt et al. 2009; Zhang and Hua 2015). So far, there are three versions of genome-scale metabolic network models of *Y. lipolytica*, namely iYL619\_PCP (Pan and Hua 2012), iNL895 (Loira et al. 2012), and iYali4 (Kerkhoven et al. 2016), where yeast consensus network was used as a reference model when iNL895 and iYali4 were reconstructed. Besides, the model iNL895 was integrated automatically, which may cause a series of network gaps when simulating cell growth or the production of several compounds. However, the knowledge in *Y. lipolytica* physiology and metabolism is being updated continuously over time, and an up-to-date metabolic model is most likely required to improve the accuracy of prediction. On the other hand, large quantities of in silico optimization methods have been developed and both gene-level and reaction-level strategies can be identified based on constraint-based flux analysis. Normally, gene-level targets are more feasible and efficient than reaction-level ones according to the complex associations between genes and reactions in metabolic model. For instance, our previously published algorithms, OptGeneKnock (Zhang et al. 2015a), IdealKnock (Gu et al. 2016), and APGC (Jian et al. 2016b), which could provide potential gene-level strategies based on logic transformation of model (LTM) (Zhang et al. 2015b), are more feasible for in vivo implementation.

In this study, to systematically investigate the metabolism of *Y. lipolytica* and gain novel insight into TAG biosynthesis of this strain, a new genome-scale metabolic model of *Y. lipolytica*, named iYL\_2.0, was reconstructed. The utilizations of different carbon sources were simulated based on this newly developed metabolic model followed by the validation by experimental data. Also, analysis of gene essentiality of *Y. lipolytica* cells growth

in minimal growth (MG) medium or yeast extract (YE) medium was conducted, and the essential genes predicted based on four *Y. lipolytica* models in YE medium were further verified by database of essential genes (DEG). Additionally, potential genetic strategies were identified by three recent in silico simulation algorithms in order to enhance the biosynthesis of TAG in *Y. lipolytica*, and furthermore, the effects of amino acid supplementation on TAG production were investigated based on the novel model of iYL\_2.0.

## Results and discussions

### Characteristics of the updated genome-scale metabolic model of *Y. lipolytica*

The reconstructed genome-scale metabolic model iYL\_2.0 (Additional file 1) included 645 genes, 1083 metabolites, and 1471 reactions (1008 metabolic reactions, 326 transport reactions, and 137 exchange reactions) (Table 1). Compared with iYL619\_PCP, the model size of iYL\_2.0 was more extensive, but less than iNL895 and iYali4 for the reason that models iNL895 and iYali4 were constructed automatically and several genes, metabolites, and reactions in iNL895 and iYali4 were quoted directly from the metabolic model of *Saccharomyces cerevisiae* (Heavner et al. 2012). The directionalities or reversibility of all metabolic reactions were defined based on thermodynamics, therefore, there exist 266 reversible and 741 irreversible enzymatic reactions in the model of iYL\_2.0. Besides, two additional compartments (peroxisome and endoplasmic reticulum) were introduced in iYL\_2.0 compared with its previous version of iYL619\_PCP. The peroxisome compartment is mainly associated to glyoxylate and dicarboxylate metabolism,

**Table 1 Features of four genome-scale metabolic models of *Y. lipolytica***

Features	iYL619_PCP	iNL895	iYali4	iYL_2.0
Compartment	4	16	16	6
Gene	619	895	901	645
Metabolite	843	1847	1683	1083
Reaction	1142	2002	1985	1471
Metabolism reaction	781	1406	1236	1008
Reversible reaction	563	1113	268	266
Irreversible reaction	218	293	968	742
Transport reaction	236	425	595	326
Exchange reaction	125	171	154	137
Biomass	<i>Y. lipolytica</i>	<i>S. cerevisiae</i>	<i>S. cerevisiae</i>	<i>Y. lipolytica</i>

while endoplasmic reticulum compartment is related to glycerolipid and glycerophospholipid metabolism.

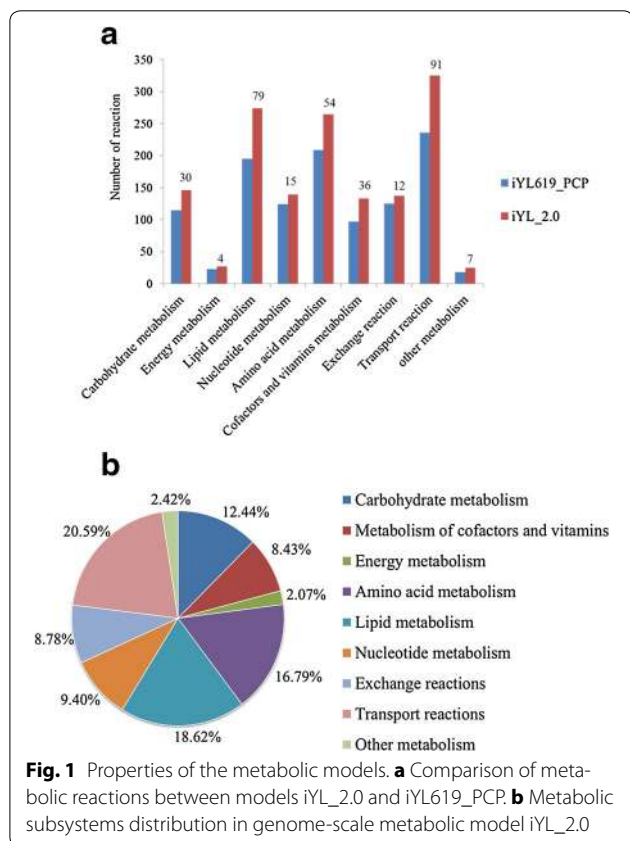
In addition, 1471 reactions were classified into nine different subsystems based on KEGG pathway database. These subsystems include carbohydrate metabolism, energy metabolism, lipid metabolism, nucleotide metabolism, amino acid metabolism, cofactors and vitamins metabolism, exchange reactions, transport reactions, and other metabolisms (terpenoid backbone biosynthesis and *N*-glycan biosynthesis). Compared to model iYL619\_PCP, the number of reactions in each subsystem of iYL\_2.0 was increased (Fig. 1a). The additional reactions in iYL\_2.0 were mainly associated with lipid metabolism (79 reactions), amino acid metabolism (54 reactions), and transport reactions (91 reactions). Specifically, lipid metabolism, the largest subsystem in iYL\_2.0, accounts for 18.6% of the total reactions (Fig. 1b), in comparison with 16.7% in iYL619\_PCP and 14.2% in iYali4.

**Analysis and verification of model iYL\_2.0**

In order to quantitatively demonstrate the advantage of iYL\_2.0, we firstly simulate the in silico cell growth characteristics in minimal growth medium based on three previously constructed models and the new model iYL\_2.0 using flux balance analysis (FBA) (Orth et al.

2010; Toya and Shimizu 2013). The maximum specific growth rate obtained from iYL\_2.0 was 0.0378 h<sup>-1</sup>, which agreed very well with the experimental value (0.0352 h<sup>-1</sup>) (Oberhardt et al. 2009), better than the results predicted by iYL619\_PCP (0.0439 h<sup>-1</sup>), iNL895 (3.4367 h<sup>-1</sup>), and iYali4 (1.7454 h<sup>-1</sup>) (Table 2). The inaccuracy of simulation result by model iNL895 was largely due to the biomass equation used in iNL895 which was integrated automatically from the yeast consensus model version 4.36 (Herrgard et al. 2008) and might lack of essential gap filling. Similarly, the inconsistency between iYali4 and experimental data might also attribute to the fact that the biomass equation used in iYali4 was quoted directly from the yeast consensus network version 5 (Heavner et al. 2012).

In addition, these models were employed to investigate in silico cell growth on 29 different carbon sources by FBA, in which each substance was used as a sole carbon source and cell growth was optimized. The utilization of different substrates by *Y. lipolytica* has previously been reported (Sitepu et al. 2014; Wehrspann and Fullbrandt 1985). According to the above literatures, *Y. lipolytica* could grow on 11 carbon sources, while cell growth was not detected when other 18 compounds were used as carbon sources (Table 3). However, the utilization of 5, 9, and 7 out of 29 carbon sources were not correctly predicted by using models iYL619\_PCP, iNL895, and iYali4, respectively. In comparison, the newly developed model iYL\_2.0 could be employed to simulate the growth characteristics on 28 carbon sources (expected for hexadecane). Models iYL619\_PCP, iNL895, and iYali4 showed false positive growth on trehalose, while iYL\_2.0, with trehalose glucohydrolase removed after careful consideration, showed no growth and were consistent with the experimental data. Moreover, iYL619\_PCP, the earlier version of iYL\_2.0, failed to show the positive growth on D-mannitol and D-glucitol, which was then fixed by adding two reactions (D-mannitol:NAD<sup>+</sup> 2-oxidoreductase and D-glucitol:NAD<sup>+</sup> 2-oxidoreductase) enabling the conversion of D-mannitol or D-sorbitol to beta-D-fructose (Scolnick and Lin 1962). On the other hand, both iNL895 and iYali4 exhibited predictions of positive



**Table 2 Comparison of in silico and in vivo biomass growth rates of *Y. lipolytica* grown in MG medium**

Medium condition (mmol/gDCW/h)	In vivo	Growth rate (h <sup>-1</sup> )			
		In silico			
		iYL619_PCP	iNL895	iYali4	iYL_2.0
Glucose (ν = 20)	0.0352	0.0439	3.4367	1.7454	0.0378

In vivo, experimental results; in silico, simulation results. Biomass reaction in GSMM *Y. lipolytica* was set as the objective function

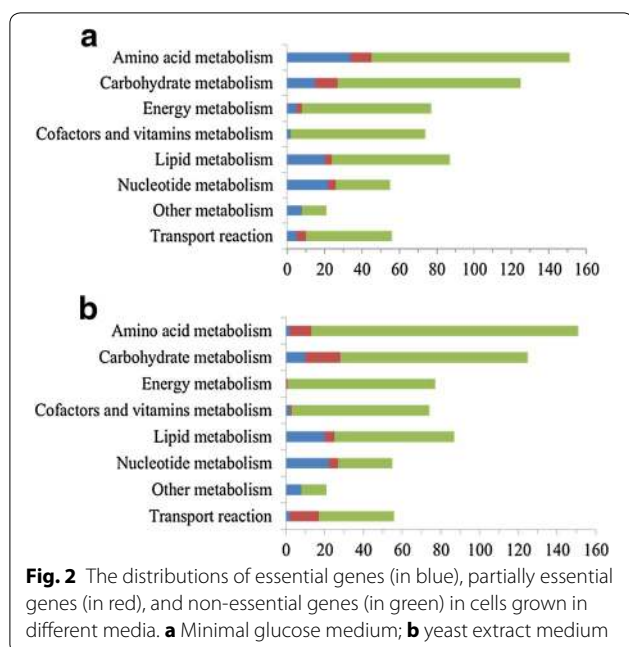
**Table 3 Comparison between simulated results by four models and experimental results on the utilization of different carbon sources in *Y. lipolytica***

Substrate	<i>in vivo</i>	<i>in silico</i>			
		iYL619_PCP	iNL895	iYali4	iYL_2.0
D-glucose	+	■	■	■	■
Ethanol	+	■	■	■	■
Glycerol	+	■	■	■	■
N-acetyl-D-glucosamine	+	■	■	■	■
Citrate	+	■	■	■	■
D-lactate	+	■	■	■	■
L-lactate	+	■	■	■	■
Succinate	+	■	■	■	■
Hexadecane	+	■	■	■	■
D-mannitol	+	■	■	■	■
D-sorbitol	+	■	■	■	■
Sucrose	-	■	■	■	■
Cellobiose	-	■	■	■	■
D-glucosamine	-	■	■	■	■
Lactose	-	■	■	■	■
Inositol	-	■	■	■	■
Trehalose	-	■	■	■	■
Methanol	-	■	■	■	■
L-rhamnose	-	■	■	■	■
Galactitol	-	■	■	■	■
Maltose	-	■	■	■	■
Melibiose	-	■	■	■	■
Salicin	-	■	■	■	■
D-xylose	-	■	■	■	■
L-arabinose	-	■	■	■	■
D-arabinose	-	■	■	■	■
Raffinose	-	■	■	■	■
Methyl-D-glucoside	-	■	■	■	■
Inulin	-	■	■	■	■

In vivo: experimental results, In silico: simulation results. '+' represents that the carbon source could be utilized and '-' represents not growth on that carbon source. The green color indicates that in silico result was consistent with in vivo data, whereas red color for the inconsistency. The uptake rate of each carbon source was set at 20 mmol/gDCW/h, and uptake rates of sulfate, phosphate, ammonia, O<sub>2</sub>, H<sub>2</sub>O, CO<sub>2</sub> and H<sup>+</sup> were set as 1000 mmol/gDCW/h

growth of *Y. lipolytica* on sucrose, cellobiose, D-glucosamine, and D-xylose, and iNL895 also showed positive growth of the strain on maltose and melibiose, which were, however, inconsistent with the in vivo experimental result. The false positive growths on sucrose, cellobiose and D-glucosamine probably attributed to a direct quotation of yeast reactions in iNL895 and iYali4, while the false positive growths on D-xylose, melibiose, and maltose may be activated by intention for simulation purpose (Forster et al. 2007; Wang et al. 2014). Interestingly, all four models considered in this study showed false positive growth on hexadecane. We found that when the annotated genes encoding the required enzymes were added to four models, hexadecane could not be utilized as a sole carbon source, which might be due to the absence of the corresponding regulatory mechanism in the four models.

In addition, the essentialities of individual gene of *Y. lipolytica* grown in both minimal growth medium and yeast extract medium were investigated based on the four above models. All genes involved in iYL\_2.0 were classified into three categories: essential genes, partially essential genes, and non-essential genes. As concluded in Additional file 2, 111 (17.21% of the total) and 66 (10.23% of the total) genes were found as essential cells grown in MG medium and YE medium, respectively. The simulated 66 essential genes in YE medium were further examined via BLASTp with the database of essential genes (DEG) (Gao et al. 2015), where 35 genes were found to have a high homology with *S. cerevisiae* (Giaever et al. 2002). For instance, UMP phosphotransferase (YALI0F09339g) is involved in the synthesis of CMP and UMP, which was essential to cell growth. Similarly, thioredoxin reductase (YALI0D27126g) is essential in maintaining the balance of intracellular NADP<sup>+</sup>/NADPH, which is vital to the whole metabolic process of *Y. lipolytica*. In comparison, only 31 of 71, 25 of 143, and 29 of 112 of essential genes obtained from the simulation of iYL619\_PCP, iNL895, and iYali4 were confirmed by DEG, respectively, indicating a better prediction performance of essential genes by iYL\_2.0. Further, the essentialities of genes involved in eight subsystems of iYL\_2.0 are shown in Fig. 2. Most essential genes were found distributed in amino acid metabolism, lipid metabolism, and nucleotide metabolism in case of MG medium. When analyzing gene essentiality for cells grown in YE medium, only approximately 3% genes categorized in amino acid metabolism were considered essential for cell growth due to the fact that yeast extract contains a mix of peptides, several free amino acids, purine, and pyrimidine bases. For transport reaction, the essential genes associated



with amino acids metabolism were mainly identified in cells grown in MG medium as the uptake of amino acids needs transport reactions. In addition, for genes categorized in other metabolism, the proportions of essential genes in this category were similar when cells were cultivated in both MG medium and YE medium, implying fewer alternatives routes for biosynthesis of some terpenoids and *N*-glycan in cell metabolic network.

### Prediction of metabolic engineering strategies for improving triacylglycerol production

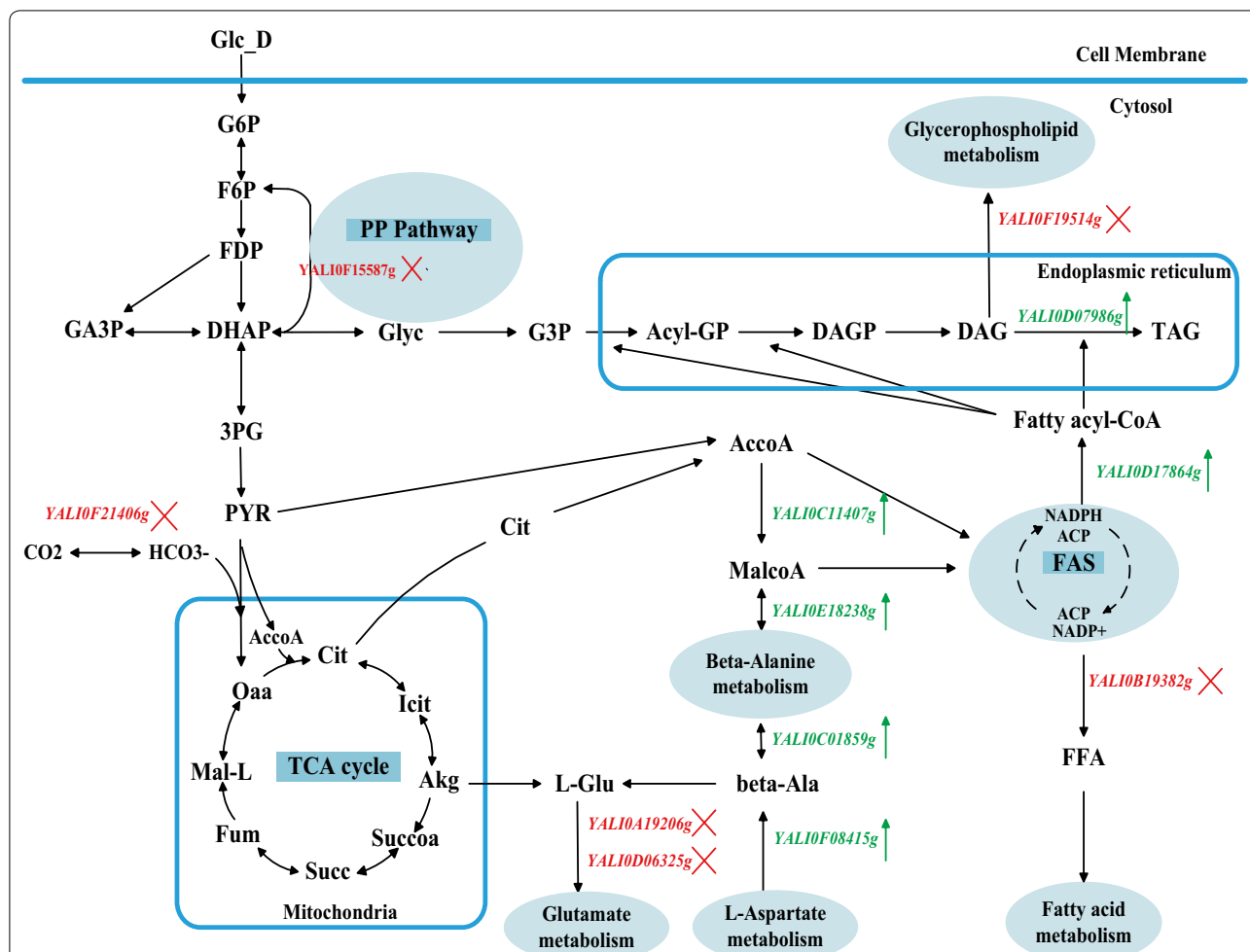
*Yarrowia lipolytica* is capable of producing TAG by the Kennedy pathway (Fig. 3). As an acyl donor, fatty acyl-CoA is mostly supplied by fatty acid synthase system (FAS) (Schweizer et al. 1988; Wang et al. 2011). In addition, acetyl-CoA and malonyl-CoA are essential carbon donor molecules for TAG biosynthesis. In model *iYL\_2.0*, there are multiple pathways that could supply the above essential molecules, including citrate shuttle and conversion from the TCA cycle, amino acids degradation, and other acetyl-CoA generation process (Beopoulos et al. 2011). In addition to TAG production, acetyl-CoA and malonyl-CoA are also used for fatty acid synthesis, amino acid synthesis, and glycerophospholipid synthesis.

To gain additional insight into TAG biosynthesis in *Y. lipolytica*, our previously published simulation methods OptGeneKnock (Zhang et al. 2015a), IdealKnock (Gu et al. 2016), and APGC (Jian et al. 2016b), were employed to identify metabolic engineering strategies to enhance TAG production in cells grown MG medium. The OptGeneKnock algorithm was modified from OptKnock

algorithm through LTM (Zhang et al. 2015), where OptKnock is able to screen the best potential reaction knockout strategies based on a bilevel programming framework. However, IdealKnock algorithm is a top-down framework that scans for the mutants of interest firstly, followed by the identification of the knockout target by iterative searching process. The best single deletion strategy identified by OptGeneKnock was the knockout of carbonate hydrolyase (YALIOF21406g) from the metabolic network, leading to an increase of specific TAG production rate from 0.0126 to 0.0810 mmol/gDCW/h together with the optimal cell growth. The carbonate hydrolyase associated with enzymatic reaction R0450 is involved in energy metabolism and nitrogen metabolism of *Y. lipolytica*. By comparison of the in silico metabolic flux distribution before and after the deletion of YALIOF21406g, we observed that more carbon fluxes were distributed into acetyl-CoA and malonyl-CoA metabolism while fluxes of amino acid synthesis pathway and synthesis pathways of other compounds decreased. Moreover, prediction of double deletion strategy with OptGeneKnock indicated that in addition to YALIOF21406g, further deletion of gene YALIOF15587g (*D*-glyceraldehyde-3-phosphate glycerontransferase, R0487) could improve TAG production rate. Actually, TAG synthesis was reported to be indirectly enhanced together with the decrease in pentose phosphate pathway flux and increase in glycerol formation when R0487 was removed from *Y. lipolytica* (Tai and Stephanopoulos 2013).

Meanwhile, another quadruple knockout target (YALIOF19514g, YALIOB19382g, YALIOA19206g, and YALIOD06325g) was identified by IdealKnock program of the COBRA Toolbox v2.0 (Additional file 3). *O*-Acyltransferase (YALIOF19514g) is a significant enzyme responsible for the synthesis of phosphatidylcholine from 1-acyl-*sn*-glycero-3-phosphocholine, which was essential in the last step of TAG biosynthesis (Makri et al. 2010). The deficiency of *O*-acyltransferase might decrease the flux of phosphatidylcholine to glycerophospholipid synthesis and therefore improve the TAG production directly. The gene YALIOB19382g is associated with three reactions (R0161, R0164, and R0176) in the FAS pathway. The in silico deletion of YALIOB19382g could suppress the fatty acid biosynthesis from fatty acyl-CoA, and the TAG synthesis was enhanced accordingly. Furthermore, the removal of YALIOA19206g and YALIOD06325g responsible for glutamate degradation might contribute to the activation of acetyl-CoA carboxylase and the formation of malonyl-CoA due to increased availability of intracellular glutamate (Yu et al. 2003).

To further investigate gene targets of overexpression for enhanced TAG production in MG medium, we



**Fig. 3** Overview of different pathways involved in TAG synthesis in model iYL\_2.0. Metabolite abbreviations are shown in black and genes in red or green. The genes in red are knockout targets identified by OptGeneKnock and IdealKnock, the genes in green are overexpression targets identified by APGC. Metabolite abbreviations: Glc\_D, D-glucose; G6P, D-glucose 6-phosphate; F6P, D-glucose 6-phosphate; FDP, D-fructose 1,6-bisphosphate; DHAP, Glycerone phosphate; GA3P, D-glyceraldehyde 3-phosphate; 3PG, D-glycerate 3-phosphate; PYR, pyruvate; Oaa, oxaloacetate; Cit, citrate; Icit, isocitrate; Akg, 2-oxoglutarate; Succoa, succinyl-CoA; Succ, succinate; Fum, fumarate; Mal-L, L-malate; AccoaA, Acetyl-CoA; MalcoA, Malonyl-CoA; beta-Ala, beta-alanine; L-Glu, L-glutamate; Glyc, glycerol; G3P, sn-glycerol 3-phosphate; Acyl-GP, 1-acyl-sn-glycerol 3-phosphate; DAGP, 1,2-diacyl-sn-glycerol 3-phosphate; DAG, diacylglycerol; TAG, triacylglycerol; ACP, acyl-carrier protein

performed APGC algorithm with the use of newly constructed model iYL\_2.0. The prediction results showed that 6 genes were identified as potential targets for overexpression (Additional file 3). All these gene targets play important roles in TAG synthetic pathway and these genes are involved in amino acids metabolism, degradation of fatty acids, glycerolipid metabolism, and the growth of biomass. For example, the overexpression of acetyl-CoA:carbon-dioxide ligase (YAL10C11407g) led to an increase in malonyl-CoA synthesis, which hence improved the availability of acetyl-CoA for TAG biosynthesis (Qiao et al. 2015). Three potential gene targets (YAL10F08415g, YAL10C01859g, and YAL10E18238g)

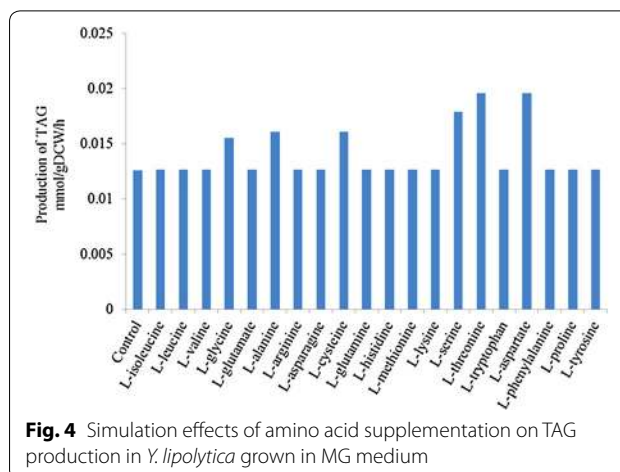
categorized in amino acids metabolism are responsible for the conversion of amino acids into acetyl-CoA and malonyl-CoA. In addition, reaction R0310 (YAL10D07986g) is the last step of TAG biosynthetic pathway, and overexpressing this step might contribute directly to the production enhancement of TAG, which was reported by Yen et al. (2005). The simulation of the overexpression of the above 6 genes showed a 34.1% improvement of TAG production rate compared to the wide-type strain (from 0.0126 to 0.0169 mmol/gDCW/h).

Amino acid metabolism always plays a significant role in lipid metabolism. During the biosynthesis of TAG, additional supplementation of amino acids might be able

to promote the biomass growth and TAG production. For this purpose, the effects of individual amino acid additions in MG medium on TAG production were simulated (Additional file 4). The uptake rate of each amino acid was set at 20 mmol/gDCW/h. As shown in Fig. 4, the supplementation of L-glycine, L-alanine, L-cysteine, L-serine, L-threonine, and L-aspartate improved TAG production through enhancing the supply of acetyl-CoA, similar to the results reported for *Schizochytrium limacinum* (Ye et al. 2015a). The TAG production rate was increased by 23.3, 27.7, 27.8, 42.0, 55.5, and 55.5%, respectively. The uptake of above-mentioned amino acids might contribute to the enhanced supply of acetyl-CoA and malonyl-CoA, which therefore increased the production rate of TAG. One example is that additional L-threonine in the medium might be converted to intracellular acetyl-CoA via acetaldehyde:NAD<sup>+</sup>/NADP<sup>+</sup> oxidoreductase (R0361/R0362) and acetate:CoA ligase (R0358). In addition, via L-aspartate ligase (R0887) and 2-(Nomega-L-arginino) succinate arginine-lyase (R0049), L-aspartate could be converted into either fumarate to enhance the TCA cycle or malonyl-CoA for TAG biosynthesis.

## Conclusions

In this study, a novel GSMM of *Y. lipolytica*, named iYL\_2.0, was reconstructed based on two previous models and metabolic information from updated databases and recent literatures. The new model contains 645 genes, 1083 metabolites, and 1471 reactions across different compartments. This model was found superior to the two previous models when employed to simulate cell growth characteristics on glucose or other carbon sources. Gene essentiality analysis resulted in 111 essential genes for cells grown in MG medium and 66 essential genes when cells cultivated in YE medium. Using model iYL\_2.0, double and quadruple gene knockout targets that could facilitate TAG production were identified by OptGeneKnock and IdealKnock, respectively. In addition, 6 gene amplification targets were predicted by APGC framework, which could lead to a 34.1% improvement of TAG production for increasing the availability of important precursors of acetyl-CoA and malonyl-CoA. Further simulation with iYL\_2.0 indicated that the additional supplementation of six amino acids, namely L-glycine, L-alanine, L-cysteine, L-serine, L-threonine, and L-aspartate, might largely contribute to the enhancement of TAG production, where the addition of either L-threonine or L-aspartate resulted in a 55.5% increase in TAG production rate. In summary, the newly developed genome-scale metabolic model iYL\_2.0 may serve as a powerful platform for better understanding metabolic viability and guiding rational metabolic engineering strategies for *Y. lipolytica*.



**Fig. 4** Simulation effects of amino acid supplementation on TAG production in *Y. lipolytica* grown in MG medium

## Methods

### Reconstruction of *Y. lipolytica* GSMM

The draft model was reconstructed by amassing reactions from two reference models of *Y. lipolytica*, iYL619\_PCP (Pan and Hua 2012), and iYali4 (Kerkhoven et al. 2016). To update and expand the draft model, the absent metabolic reactions were added from KEGG database (Kanehisa et al. 2008). Three databases UniProt (Apweiler et al. 2004), CHEBI (Degtyarenko et al. 2008), and PubChem (Wang et al. 2009) were used to characterize the annotation of metabolites based on their chemical formulas and neutral charges. Additionally, the reaction reversibilities were judged by BioPath (Schreiber 2002) database and thermodynamic data of metabolites. Compared with iYL619\_PCP, two new compartments, the peroxisome and endoplasmic reticulum, were determined according to the other reference model iYali4 and subcellular localization prediction tool CELLO (Yu et al. 2004). Transport reactions and exchange reactions were added for metabolic network connectivity. Moreover, to refine the draft model, metabolic gaps were identified by literature data and the GapFind (Satish Kumar et al. 2007) program in COBRA Toolbox v2.0 (Schellenberger et al. 2011).

### Simulation and constraint-based flux analysis

The verification and analysis included optimal cell growth in minimal growth (MG) medium (Lanciotti et al. 2005) and utilization of different carbon sources. Meanwhile, essential genes were analyzed by *singlegene-deletion* function in COBRA Toolbox v2.0 in MG medium and YE medium (Ye et al. 2015b). The MG medium only contains basic elements, such as C, H, O, N, P, and S. The glucose uptake rate was set as 20 mmol/gDCW/h, and uptake rates of sulfate, O<sub>2</sub>, phosphate, ammonia, H<sub>2</sub>O,

CO<sub>2</sub>, and H<sup>+</sup> were set as 1000 mmol/gDCW/h. The YE medium contains not only basic elements, but also 20 amino acids, and the uptake rates of amino acids were set as 5 mmol/gDCW/h. Biomass equation was set as the objective function for the different conditions, and optimization was performed by GLPK solver.

The in silico simulations were performed by flux balance analysis (FBA), which is a constraint-based flux analysis method (Orth et al. 2010; Toya and Shimizu 2013). Meanwhile, under a pseudo-steady-state metabolic model, FBA determines metabolic flux distribution based on the assumption that organisms could endeavor to optimize growth or production. Mathematically, the optimization of objective functions subjected to stoichiometric and capacity constraints can be formulated as follows:

$$\begin{aligned} \max \quad & Z = \sum_j c_j v_j \\ \text{s.t.} \quad & \sum_j S_{ij} v_j = 0; \quad \forall \text{ metabolite } i \\ & v_j^{\min} \leq v_j \leq v_j^{\max}; \quad \forall \text{ reaction } j. \end{aligned} \quad (1)$$

$Z$  represents the cellular objective function of all the metabolic reactions where the relative weights are determined by the coefficient  $c_j$ .  $S_{ij}$  corresponds to the stoichiometric coefficient of the metabolite  $i$  involved in reaction  $j$ .  $v_j$  denotes the flux or specific rate of metabolic reaction  $j$ .  $v_j^{\min}$  and  $v_j^{\max}$  refer to the lower and upper bounds of the flux of reaction  $j$ , respectively.

#### Identification of potential targets through gene-level algorithms

For analysis of TAG production, two novel knockout algorithms OptGeneKnock (Zhang et al. 2015a) and IdealKnock (Gu et al. 2016) and one overexpression algorithm, analysis of production and growth coupling (APGC) (Jian et al. 2016b) were carried out for cells grown in MG medium. The APGC program was applied to identify amplification targets for improving production, and OptGeneKnock and IdealKnock were used to identify knockout targets. The OptGeneKnock algorithm was performed by applying logic transformation of model (LTM) (Zhang et al. 2015a) to OptKnock (Burgard et al. 2003). GLPK solver and Gurobi solver were used for linear programming and quadratic programming, respectively. The synthesis reaction of TAG, R0310, was set as objective function to identify the potential targets, and the minimum specific growth rate of biomass was set as 0.0378 h<sup>-1</sup> in MG medium.

## Additional files

**Additional file 1.** Detailed information about metabolic model iYL\_2.0.

**Additional file 2.** Gene essentiality analysis in minimal growth medium and yeast extract medium by flux balance analysis.

**Additional file 3.** Potential gene-level targets identified by OptGeneKnock, IdealKnock and APGC to improve TAG production.

**Additional file 4.** Simulation results of amino acid supplementation for TAG production.

#### Abbreviations

TAG: triacylglycerol; DCW: dry cell weight; SCD: delta-9 stearoyl-CoA desaturase; GSMM: genome-scale metabolic model; MG: minimal growth; YE: yeast extract; FBA: flux balance analysis; DEG: database of essential genes; FAS: fatty acid synthase system; LTM: logic transformation of model; APGC: analysis of production and growth coupling.

#### Authors' contributions

QH and CZ developed and designed the study. SW carried out the reconstruction of iYL\_2.0, performed the statistical analysis, and prepared the manuscript. XJ provided help for implementation of the in silico algorithms. QH and CZ revised the manuscript. All authors read and approved the final manuscript.

#### Author details

<sup>1</sup> State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, People's Republic of China. <sup>2</sup> Science for Life Laboratory, KTH-Royal Institute of Technology, Stockholm SE-171 21, Sweden. <sup>3</sup> Shanghai Collaborative Innovation Center for Biomanufacturing Technology, 130 Meilong Road, Shanghai 200237, China.

#### Acknowledgements

We thank Prof. Jens Nielsen for providing the *Y. lipolytica* model of iYali4 and Prof. David James Sherman for providing the *Y. lipolytica* model of iNL895.

#### Competing interests

The authors declare that they have no competing interests.

#### Availability of data and materials

Not applicable.

#### Consent for publication

Not applicable.

#### Ethics approval and consent to participate

Not applicable.

#### Funding

This study was financially supported by National Natural Science Foundation of China (21776081, 21576089).

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 September 2017 Accepted: 22 November 2017

Published online: 02 December 2017

#### References

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M et al (2004) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 32:D115–D119



- Athenstaedt K (2011) YALI0E32769g (DGA1) and YALI0E16797g (LRO1) encode major triacylglycerol synthases of the oleaginous yeast *Yarrowia lipolytica*. *Biochem Biophys Acta* 1811:587–596
- Beopoulos A, Nicaud JM, Gaillardin C (2011) An overview of lipid metabolism in yeasts and its impact on biotechnological processes. *Appl Microbiol Biotechnol* 90:1193–1206
- Blažek J, Hill A, Liu LQ, Knight R, Miller J, Pan A, Otoupal P, Alper HS (2014) Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* 5:3131
- Burgard AP, Pharkya P, Maranas CD (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 84:647–657
- Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcantara R, Darsow M, Guedj M, Ashburner M (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 36:D344–D350
- Dulermo T, Nicaud JM (2011) Involvement of the G3P shuttle and beta-oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. *Metab Eng* 13:482–491
- Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7:129–143
- Forster A, Aurich A, Mauersberger S, Barth G (2007) Citric acid production from sucrose using a recombinant strain of the yeast *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 75:1409–1417
- Gao F, Luo H, Zhang CT, Zhang R (2015) Gene essentiality analysis based on DEG 10, an updated database of essential genes. *Methods Mol Biol* 1279:219–233
- Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B et al (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418:387–391
- Gu D, Zhang C, Zhou S, Wei L, Hua Q (2016) Idealknock: a framework for efficiently identifying knockout strategies leading to targeted overproduction. *Comput Biol Chem* 61:229–237
- Heavner BD, Smallbone K, Barker B, Mendes P, Walker LP (2012) Yeast 5—an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Syst Biol* 6:55
- Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M et al (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 26:1155–1160
- Jian X, Li N, Zhang C, Hua Q (2016a) In silico profiling of cell growth and succinate production in *Escherichia coli* NZN111. *Bioresour Bioprocess* 3:48
- Jian X, Zhou S, Zhang C, Hua Q (2016b) In silico identification of gene amplification targets based on analysis of production and growth coupling. *Bio Syst* 145:1–8
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T et al (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36:D480–D484
- Kennedy EP (1957) Metabolism of lipides. *Annu Rev Biochem* 26:119–148
- Kerkhoven EJ, Pomraning KR, Baker SE, Nielsen J (2016) Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. *NPJ Syst Biol Appl* 2:16005
- Lancioti R, Gianotti A, Baldi D, Angrisani R, Suzzi G, Mastrocola D, Guerzoni ME (2005) Use of *Yarrowia lipolytica* strains for the treatment of olive mill wastewater. *Bioresour Technol* 96:317–322
- Ledesma-Amaro R, Nicaud JM (2016) *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. *Prog Lipid Res* 61:40–50
- Loira N, Dulermo T, Nicaud JM, Sherman DJ (2012) A genome-scale metabolic model of the lipid-accumulating yeast *Yarrowia lipolytica*. *BMC Syst Biol* 6:35
- Makri A, Fakas S, Aggelis G (2010) Metabolic activities of biotechnological interest in *Yarrowia lipolytica* grown on glycerol in repeated batch cultures. *Bioresour Technol* 101:2351–2358
- Morin N, Cescut J, Beopoulos A, Lelandais G, Le Berre V, Uribelarrea JL, Molina-Jouve C, Nicaud JM (2011) Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast *Yarrowia lipolytica*. *PLoS ONE* 6:e27966
- Oberhardt MA, Palsson BO, Papin JA (2009) Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 5:320
- Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? *Nat Biotechnol* 28:245–248
- Pan P, Hua Q (2012) Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, *Yarrowia lipolytica*. *PLoS ONE* 7:e51535
- Papanikolaou S, Aggelis G (2002) Lipid production by *Yarrowia lipolytica* growing on industrial glycerol in a single-stage continuous culture. *Bioresour Technol* 82:43–49
- Papanikolaou S, Aggelis G (2010) *Yarrowia lipolytica*: a model microorganism used for the production of tailor-made lipids. *Eur J Lipid Sci Technol* 112:639–654
- Papanikolaou S, Chevalot I, Komaitis M, Marc I, Aggelis G (2002) Single cell oil production by *Yarrowia lipolytica* growing on an industrial derivative of animal fat in batch cultures. *Appl Microbiol Biotechnol* 58:308–312
- Qiao K, Imam Abidi SH, Liu H, Zhang H, Chakraborty S, Watson N, Kumaran Ajikumar P, Stephanopoulos G (2015) Engineering lipid overproduction in the oleaginous yeast *Yarrowia lipolytica*. *Metab Eng* 29:56–65
- Satish Kumar V, Dasika MS, Maranas CD (2007) Optimization based automated curation of metabolic reconstructions. *BMC Bioinform* 8:212
- Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S et al (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6:1290–1307
- Schreiber F (2002) High quality visualization of biochemical pathways in BioPath. *In Silico Biol* 2:59–73
- Schweizer E, Kottig H, Regler R, Rottner G (1988) Genetic control of *Yarrowia lipolytica* fatty acid synthetase biosynthesis and function. *J Basic Microbiol* 28:283–292
- Scolnick EM, Lin EC (1962) Parallel induction of D-arabitol and D-sorbitol dehydrogenases. *J Bacteriol* 84:631–637
- Sitepu I, Selby T, Lin T, Zhu S, Boundy-Mills K (2014) Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. *J Ind Microbiol Biotechnol* 41:1061–1070
- Tai M, Stephanopoulos G (2013) Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. *Metab Eng* 15:1–9
- Toya Y, Shimizu H (2013) Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnol Adv* 31:818–826
- Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 37:W623–W633
- Wang JJ, Zhang BR, Chen SL (2011) Oleaginous yeast *Yarrowia lipolytica* mutants with a disrupted fatty acyl-CoA synthetase gene accumulate saturated fatty acid. *Process Biochem* 46:1436–1441
- Wang W, Wei H, Alahuhta M, Chen X, Hyman D, Johnson DK, Zhang M, Himmel ME (2014) Heterologous expression of xylanase enzymes in lipogenic yeast *Yarrowia lipolytica*. *PLoS ONE* 9:e111443
- Wehrspann P, Fullbrandt U (1985) Report of a case of *Yarrowia lipolytica* (Wickerman et al.) van der Walt & von Arx isolated from a blood culture. *Mykosen* 28:217–222
- Ye C, Qiao W, Yu X, Ji X, Huang H, Collier JL, Liu L (2015a) Reconstruction and analysis of the genome-scale metabolic model of *Schizochytrium limacinum* SR21 for docosahexaenoic acid production. *BMC Genomics* 16:799
- Ye C, Xu N, Chen H, Chen YQ, Chen W, Liu L (2015b) Reconstruction and analysis of a genome-scale metabolic model of the oleaginous fungus *Mortierella alpina*. *BMC Syst Biol* 9:1
- Yen CLE, Monetti M, Burri BJ, Farese RV (2005) The triacylglycerol synthesis enzyme DGAT1 also catalyzes the synthesis of diacylglycerols, waxes, and retinyl esters. *J Lipid Res* 46:1502–1511
- Yen CL, Stone SJ, Koliwad S, Harris C, Farese RV Jr (2008) Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. *J Lipid Res* 49:2283–2301
- Yu LJ, Qin WM, Lan WZ, Zhou PP, Zhu M (2003) Improved arachidonic acids production from the fungus *Mortierella alpina* by glutamate supplementation. *Bioresour Technol* 88:265–268
- Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci* 13:1402–1406
- Zhang C, Hua Q (2015) Applications of genome-scale metabolic models in biotechnology and systems medicine. *Front Physiol* 6:413
- Zhang C, Ji B, Mardinoglu A, Nielsen J, Hua Q (2015) Logical transformation of genome-scale metabolic models for gene level applications and analysis. *Bioinformatics* 31:2324–2331