**Genome-scale metabolic modeling of *Mucor circinelloides* and comparative analysis
with other oleaginous species**

Wanwipa Vongsangnak^{1,2}, Amornpan Klunchui³, Iyarest Tawornsamretkit⁴, Witthawin
Tatiyaborwornchai⁵, Kobkul Laoteng^{6*}, Asawin Meechai^{4,5*}

¹Department of Zoology, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.

²Center for Systems Biology, Soochow University, Suzhou 215006, China.

³Biological Engineering Program, Faculty of Engineering, King Mongkut's University of
Technology Thonburi, 126 Pracha Uthit Rd, Thung Khru, Bangkok, 10140, Thailand.

⁴Systems Biology Research Group, King Mongkut's University of Technology Thonburi,
Bangkok 10140, Thailand.

⁵Department of Chemical Engineering, Faculty of Engineering, King Mongkut's University
of Technology Thonburi, Bangkok 10140, Thailand.

⁶Bioprocess Technology Laboratory, Bioresources Technology Unit, National Center for
Genetic Engineering and Biotechnology, National Sciences and Technology Development
Agency, Khong Luang, Pathum Thani, 12120, Thailand.

*Corresponding authors

E-mails: kobkul@biotec.or.th (KL), asawin.mee@kmutt.ac.th (AM)

Abstract

We present a novel genome-scale metabolic model *iWV1213* of *M. circinelloides*, which is an oleaginous fungus for industrial applications. The model contains 1,213 genes, 1,413 metabolites and 1,326 metabolic reactions across different compartments. We demonstrate that *iWV1213* is able to accurately predict the growth rates of *M. circinelloides* on various nutrient sources and culture conditions using Flux Balance Analysis and Phenotypic Phase Plane analysis. Comparative analysis of three oleaginous genome-scale models, including *M. circinelloides* (*iWV1213*), *Mortierella alpina* (*iCY1106*) and *Yarrowia lipolytica* (*iYL619_PCP*) revealed that *iWV1213* possesses a higher number of genes involved in carbohydrate, amino acid, and lipid metabolisms that might contribute to its versatility in nutrient utilization. Moreover, the identification of unique and common active reactions among the Zygomycetes oleaginous models using Flux Variability Analysis unveiled a set of gene/enzyme candidates as metabolic engineering targets for cellular improvement. Thus, *iWV1213* offers a powerful metabolic engineering tool for multi-level omics analysis, enabling strain optimization as a cell factory platform of lipid-based production.

Keywords: Lipid, *Mucor circinelloides*, Genome-scale metabolic modeling, Oleaginous species, Metabolic engineering

1. Introduction

The field of lipid biotechnology is rapidly expanding as a result of the diverse industrial applications of lipids across several sectors. Functional lipids involved in vital metabolic processes for improved health and well-being are of commercial interest. In addition, there is a growing demand for bio-oils as potential feedstocks in the oleochemical industries [1, 2]. Indeed, particular attention has been given to discovering potential sources of specialty and commodity oils. The so-called oleaginous microorganisms, which accumulate lipid at over 20% of their biomass [3], are considered powerful workhorses for single-cell-oil (SCO) production. Particular yeast, *Yarrowia lipolytica*, is able to store lipid at up to 70-80% of its biomass [4, 5]. Some oleaginous fungi, including *Mucor* sp. and *Mortierella* sp., are also able to synthesize specific lipids, such as high-value polyunsaturated fatty acids (PUFAs). Several approaches for SCO production have been investigated to improve the economic feasibility in an array of applications, including personal care products, cosmetics, food, animal feed, pharmaceutical, and other biotechnological products. Recent advances in emerging technologies, such as omics, systems biology, metabolic engineering, synthetic biology, and bioprocess engineering, have accelerated bio-production methods of SCO, including strain improvement and product development [6].

Certain oleaginous species display advantageous cultivation properties, including the ability to utilize a wide range of substrates (carbon and nitrogen sources), and to tolerate environmental stress (e.g. acidic pH). The dimorphic oleaginous fungus, *Mucor circinelloides*, which belongs to the class Zygomycetes, is a well-known lipid producer capable of accumulating a high level of triacylglycerol containing γ -linolenic acid (GLA, C18:3 n -6).

Accordingly, it is considered a promising strain for biodiesel production [7, 8]. Recent progress in the omics era has simplified genetic analysis of *M. circinelloides*, making it possible to interconnect different levels of omics information and create genome-scale metabolic networks that can be exploited to improve both the strain and the cultivation process. The genome-scale metabolic network of *M. circinelloides* was first presented in 2013 by Vongsangnak et al. [9]. This metabolic network demonstrated an association between overall metabolic reactions and processes in *M. circinelloides*, in particular those processes related to cell growth and lipid accumulation. However, in order to apply this existing knowledge for the purposes of metabolic engineering and strain improvement, a development of metabolic model is required to predict cell behavior accurately.

Here, we use a range of protein and pathway databases combined with prediction tools to reconstruct the genome-scale metabolic model of *M. circinelloides* with extensively improved annotation. To explore metabolic characteristics and behaviors of this oleaginous fungus, the biomass equation is formulated as well as transport and exchange reactions are added to reconstruct a model capable of simulating growth. This oleaginous model is then validated by Flux Balance Analysis (FBA) and Phenotypic Phase Plane (PhPP) analysis.

Metabolic and phenotypic features of *M. circinelloides* and other related oleaginous species, including *M. alpina* and *Y. lipolytica* were investigated using comparative models analysis and Flux Variability Analysis (FVA) to determine important common and unique pathways.

The *M. circinelloides* model presented here can be used as a powerful tool for integrative analysis of omics data in systems biology, as well as for the development of efficient

microbial cell factories, particularly in the production of lipid-derived products for both food and non-food applications.

2. Materials and Methods

2.1 Improving reconstruction of genome-scale metabolic network of *M. circinelloides*

An earlier metabolic network of *M. circinelloides* [9] was used as a basis to gain better reconstruction of genome-scale metabolic network for this study. To investigate unidentified genes and relevant enzyme functions and provide novel putative reactions to enhance the metabolic network of *M. circinelloides*, the SEquence annotaTOR (SEQTOR) [10] and various protein and pathway databases, including KEGG [11] via GhostKOALA (www.kegg.jp/ghostkoala/) and BlastKOALA (www.kegg.jp/blastkoala/), JGI (genome.jgi-psf.org/Mucci2/Mucci2.home.html), Interproscan [12], KOGs [13], Pfam[14], and Conserved Domain Database (CDD) [15] were used for improved annotation. Compartmentalization information was determined using the subcellular localization prediction tool, e.g., CELLO [16]. The metabolite names and reversibility of metabolic reactions were curated according to MetaCyc [17] (metacyc.org/) and KEGG databases (www.genome.jp/kegg/pathway.html). Moreover, transport and exchange reactions were added for network connectivity.

2.2 Formation of *M. circinelloides* genome-scale metabolic model

In addition to the reconstructed genome-scale metabolic network, a metabolic model of *M. circinelloides* at a genome-scale was formulated using COBRA toolbox version 2 [18] of

MATLAB platform (The Mathworks Inc.) with the Systems Biology Markup Language (SBML). The biomass composition of *M. circinelloides* was calculated from a range of sources, including biochemical literature and genome databases, and added into the model (Supplementary File 1). The contents of protein, nucleotide, lipid, and carbohydrate were derived from *Mucor lusitanicus* [19]. The average DNA and RNA contents were calculated from the extracted dsDNA of dry cell weight of *Mucor mucedo* [20]. The compositions of amino acid, DNA and RNA were taken from the total amino acid, DNA and RNA sequence information of *M. circinelloides* from the JGI genome database (genome.jgi-psf.org/Mucci2/Mucci2.home.html). The summarized biomass composition is shown in Table 1. The full biomass composition is available in Supplementary File 2. For energetic parameters, the ATP requirement for non-growth associated purposes (m_{ATP}), the ATP requirement for synthesis of biomass from macromolecules (K_{ATP}), and the operational Phosphate/Oxygen (P/O) ratio were considered. These parameters could not be determined independently, but can be estimated from the experimental data where one parameter is known. The operational P/O ratio was assumed to be 1.5 [21], the m_{ATP} ($\text{mmol gDW}^{-1}\text{h}^{-1}$) was estimated to be 1.9, and the K_{ATP} (mmol gDW^{-1}) was estimated by fitting model simulation with experimental data obtained at a specific growth rate of 0.11 h^{-1} [22] with glucose as a sole carbon source. The value of K_{ATP} was hereby estimated to be 72 mmol gDW^{-1} .

2.3 *i*WV1213 model simulation and validation using FBA

The constraint-based flux simulation was performed using FBA and linear programming solver, GLPK (www.gnu.org/software/glpk), provided by the COBRA toolbox version 2 [18]

running through MATLAB. To estimate the optimal flux distributions under the maximized cell growth, the biomass equation was formulated and set as the objective function for model simulation under aerobic growth conditions for the two given carbon sources, glucose and galactose (Figure 1). The prediction results were validated using published data from independent experiments [22-24]. *iWV1213* was also used to assess growth capability using different substrates, including carbon and nitrogen sources (Table 2).

2.4 Characterizing metabolic phenotype of *M. circinelloides* using PhPP analysis

PhPP is a technique used for sensitivity analysis to describe growth characteristics as a function of dual variables [25]. In this study, PhPP was performed for two simulation conditions using the COBRA toolbox version 2 [18]. The first simulation was carried out by setting the boundaries of nitrogen (i.e. NH_3) and glucose consumption rates and iteratively calculating the growth rates. The second simulation was performed by setting the boundaries of oxygen and glucose consumption rates and iteratively calculating the growth rates.

2.5 Comparative models analysis of *M. circinelloides* and other oleaginous species

A comparison of the general and metabolic features of our reconstructed *M. circinelloides* model and the published genome-scale metabolic models of *M. alpina* and *Y. lipolytica* was performed. For this purpose, the SBML files of both oleaginous models of *M. alpina* (iCY1106) and *Y. lipolytica* (iYL619_PCP) were retrieved from previous reports of Ye et al. [26], and Pan and Hua [27], respectively and compared for growth phenotypes. Besides, the SBML files were converted to EXCEL format and assessed for the number of genes, EC numbers, metabolic reactions, metabolites, and cellular compartments.

2.6 Comparative FVA of the *M. circinelloides* model and other oleaginous model

FVA was performed to determine the robustness of the reconstructed metabolic models. FVA examines the range of flux values (v) for individual reactions through a metabolic network to satisfy a given constraints set, optimizing for a particular objective function [28]. In this study, a maximum biomass production rate was set for the objective function and first computation [29]. Each reaction in the network was consequently set as the objective function, and the maximum and minimum flux values through each reaction were calculated by FBA with the constraints corresponding to an optimal solution [28]. The calculated minimum and maximum flux values for each reaction were defined as the flux variability. *iWV1213* and *iCY1106* were used to perform FVA using COBRA toolbox version 2 [18] to identify active metabolic routes for growth and lipid production.

3. Results

3.1 Characteristics of the genome-scale metabolic model of oleaginous *M. circinelloides*

The reconstructed genome-scale metabolic model *iWV1213* includes 1,213 genes from a total of 11,719 protein-coding genes (10.4%) presented in the *M. circinelloides* genome. *iWV1213* consists of 1,413 metabolites and 1,326 metabolic reactions across five compartments, including the plasma membrane, cytoplasm, mitochondria, peroxisome, and extracellular space (Table 3). Of these metabolic reactions, 1,065 reactions (80.3%) were gene-associated. Compared to the previous published metabolic network [9], *iWV1213* contains a higher degree of Gene-Protein-Reaction (GPR) information (Supplementary File 1) with full formation of biomass equation, including biomass compositions (Table 1 and Supplementary

File 2), representing the true cellular biomass in the *in silico* simulation. In addition, the improved metabolic network reconstruction was relied on the improved annotation data (Supplementary File 3). All possible putative genes and relevant enzyme functions involved in metabolic pathways, which do not exist in the earlier reconstructed metabolic network [9], were continually added into the metabolic network of *M. circinelloides*. Obviously, the putative genes and enzyme functions involved in lipoic acid metabolism, fatty acid elongation, glycosylphosphatidyl inositol (GPI)-anchor biosynthesis, arachidonic acid, nicotinate and nicotinamide metabolism, and N-glycan biosynthesis. Considering on network connectivity, a total of 178 transport and exchange reactions were curated and added into the model.

A comparative summary of the genome-scale metabolic model of oleaginous *M. circinelloides* with that of other oleaginous species is presented in Table 3. The Systems Biology Markup Language (SBML) model of *iWV1213* is also available in Supplementary File 4 via www.sbi.kmutt.ac.th/MCModel/S4_File.xml.

3.2 Growth simulation and validation of *iWV1213* on various substrates

To examine the metabolic capability of *iWV1213*, the simulated growth rate was validated with experimental data using FBA and linear programming through the COBRA toolbox version 2 [18]. The *in silico* growth rate was derived by maximizing the biomass with various constraints, for example by constraining measurable uptake rates of extracellular metabolites, such as different carbon sources e.g., glucose and galactose. The biomass equation (or flux), which was the objective function, was then fine-tuned according to the fungal growth

conditions. As shown in Figure 1, the predicted growth rates derived from the model when using glucose as a carbon source with four different uptake rates of 1.52, 2.15, 3.04, and 4.74 mmol gDW⁻¹h⁻¹ (millimole per gram dry weight per hour) were 0.1102, 0.1578, 0.2250, and 0.3533 h⁻¹, respectively, which agree with published experimental data [22-24]. Particularly, as the reports presented by Bredenkamp et al. [22] and Lübbehüsen et al. [24], the model showed closely fit to experimental data with the computed percentage error average for 0.56%. Similarly, using galactose uptake rates of 1.68 and 3.98 mmol gDW⁻¹h⁻¹, the predicted growth rates were 0.1223 and 0.2959 h⁻¹, respectively, which also agree well to the experimental data [24] with the computed percentage error average for 8.13%.

Thus, the model *iWV1213* can be used to predict fungal growth using glucose or galactose as a sole carbon source. A detailed example of the metabolic flux distribution analysis on glucose with an uptake rate of 1.52 mmol gDW⁻¹h⁻¹ is available in Supplementary File 5. To further exploit the metabolic capability of *iWV1213*, various carbon and nitrogen sources were subjected to analysis of growth prediction. Using FBA, the model correctly predicted the growth capability of *M. circinelloides* on different carbon sources (e.g., xylose, maltose, lactose, sucrose, cellulose, cellobiose, glycerol or starch) and nitrogen sources (e.g., NH₄Cl, NH₄NO₃, (NH₄)₂SO₄, KNO₃, or urea) in most cases, as shown in Table 2.

In addition, to investigate the effect of dual substrate utilization, including glucose and NH₃, and glucose and oxygen, on the cellular behavior of *M. circinelloides*, PhPP analysis was performed on *iWV1213*. As shown in Figure 2A, the biomass growth rate correlated well with the uptake rates of the substrates (glucose and NH₃). Notably, the glucose uptake rate (GUR) had a more pronounced effect on cell growth rate than the nitrogen uptake rate (NUR).

However, very high GURs inhibited cell growth, which was not observed in cultures running with high NURs. The biomass growth rate was also dependent on the oxygen uptake rate (OUR); cell growth increased with an increase in oxygen consumption (Figure 2B) when glucose was available. Under microaerobic conditions (i.e. $<2 \text{ mmol gDW}^{-1}\text{h}^{-1}$), *M. circinelloides* was still able to grow, which may be a unique feature of this dimorphic fungus that enables it to grow in a yeast-like form under oxygen limitation [30]. Interestingly, the inhibitory effect of a high GUR was attenuated in the presence of an oxygen supply, which corresponds with a previous report that the closely related species, *Mucor rouxii*, could grow at high glucose concentrations of up to 80 gL^{-1} [31]. Hence, the model is capable of predicting the optimal growth performance of *M. circinelloides* on the substrates, such as glucose and NH_3 , and under different levels of oxygenation using PhPP analysis. It is worth noting that the *M. circinelloides* model versatility is widely capable of predicting the optimal growth performance on multiple substrates (Table 2, Supplementary file 6) and different levels of oxygen supply.

3.3 Comparative analysis of three oleaginous genome-scale metabolic models

Because of the variation in lipid production by oleaginous yeast and oleaginous Zygomycete fungi, a comparison of oleaginous genome-scale metabolic models between oleaginous species was undertaken to unravel the cellular metabolic processes that may contribute to oleaginicinity. In this work, the genome-scale metabolic model of *M. circinelloides* was compared with the models of *M. alpina* and *Y. lipolytica* (Table 3). Using data from the KEGG database [32], their metabolisms were divided into seven categories as shown in Figure 3. We observed that the largest category of numbers of genes in the *M. circinelloides*

model was carbohydrate metabolism (422 genes), accounting for 34.8% of the total genes in the model, higher than that in the *M. alpina* (24.2%) and *Y. lipolytica* (21.8%) models. We also noted a number of genes encoding extracellular enzymes in *M. circinelloides* model, such as α -glucosidase, trehalase, glucoamylase, and stachyose synthetase, which make it capable of utilization of variable carbon sources for cell growth. In comparison to *M. alpina* model, the capability of utilizing different carbon sources for cell growth was also qualitatively predicted. Nonetheless, it could not achieve growth with four types of carbon sources, such as ethanol, xylose, maltose, and rhamnose [26]. This is possible because of missing enzymes and reactions involved in relevant metabolisms and cell growth on such carbon sources. As an example in rhamnose source, the lack of L-rhamnulose-1-phosphatelactaldehyde-lyase (EC: 4.1.2.19) showed that glycerone phosphate was not be generated, which could otherwise be used in the glycolysis. Reaction numbers associated with lipid metabolism were marginally higher in *M. alpina* (435 reactions) compared with *M. circinelloides* (318 reactions). The unique reactions associated with lipid metabolism in *M. alpina* belong to fatty acids and very-long-chain PUFAs biosynthetic pathways, such as dihomo- γ -linolenic acid (DGLA), arachidonic acid (ARA) and eicosapentaenoic acid (EPA) [26]. As expected, the *Y. lipolytica* model contained lower numbers of genes and reactions in carbohydrate, amino acid, and lipid metabolisms than the models of the Zygomycetes oleaginous fungi due to less number of annotated genes in a smaller genome size (Figure 3). Nonetheless, the *Y. lipolytica* model could predict the growing abilities on different substrates [27]. A comparison of *in silico* biomass growth rates on glucose as a substrate at different uptake rates among three oleaginous genome-scale

metabolic models is presented in Table 3. Among three oleaginous genome-scale metabolic models, a similar trend in terms of gene and reaction numbers for the remaining metabolism, such as cofactor metabolism is noticeable (Figure 3).

3.4 Models as scaffolds for guiding growth-related metabolic routes in oleaginous

Zygomycetes

Because the closely related fungi, *M. circinelloides* and *M. alpina*, share similar genotypes and metabolic phenotypes in terms of gene and reaction numbers in addition to their obese phenotypes, they were subjected to further analysis to identify metabolic routes relevant to cell growth and oleaginicity. The analysis of metabolic reaction activity of the two species was performed by FVA. For more precise identification of the uniqueness of an optimal solution of the models, FVA calculated the minimum and maximum flux values of each reaction. In this study, FVA was performed under GUR of 1.52 and 1.00 mmol gDW⁻¹h⁻¹ for *M. circinelloides* and *M. alpina*, respectively.

The range of results from FVA in which the flux can be changed without affecting the simulated maximal growth rate is shown in Supplementary File 7. Three types of active reaction set were observed based on the FVA results, including “always-active reactions”, “sometimes-active reactions”, and “never-active reactions”. The “always-active reactions” are the reactions whose minimum and maximum flux values are non-zero with the same sign. The so-called “sometimes-active reactions” refer to the reactions in which the ranges of possible fluxes contain zero. If minimum and maximum flux values of reactions are equal to zero during optimal growth, the reactions were considered as “never-active reactions”. In the

M. circinelloides model, we found 237 (18%) always-active reactions, 316 (24%) sometimes-active reactions, and 773 (58%) never-active reactions. For *M. alpina*, we found 348 (19%) always-active reactions, 623 (33%) sometimes-active reactions, and 883 (48%) never-active reactions. The results of three types of active reaction set observed based on FVA analysis of the two oleaginous models are shown in Figure 4 and their full details are provided in Supplementary Files 8 and 9.

We observed an overlap between the identified always-active reactions in the two oleaginous Zygomycetes species, including 105 reactions with the same Enzyme Commission (EC) numbers and enzyme functions. Interestingly, 12 precursor metabolites for biosynthesis towards 15 common always-active reactions were identified, including glucose-6-phosphate, fructose-6-phosphate, glyceraldehyde3-phosphate, 3-phosphoglycerate, phosphoenolpyruvate, pyruvate, ribose-5-phosphate, erythrose-4-phosphate, acetyl-CoA, α -ketoglutarate, oxaloacetate and succinyl-CoA. These biosynthetic precursor metabolites are involved in glycolysis/gluconeogenesis, the pentose-phosphate pathway, the TCA cycle, lanosterol biosynthesis, pyruvate metabolism, and amino acid metabolisms (e.g. lysine, phenylalanine, tyrosine, and tryptophan biosynthesis) as shown in Table 4. This suggests that the core active reactions and key enzymes for growth are similar between the two oleaginous Zygomycete fungi. For example, glycolysis breaks down glucose into pyruvate via the main glycolytic enzymes, i.e., from hexokinase (EC: 2.7.1.1) through pyruvate kinase (EC: 2.7.1.40) (Figure 5). A full list of the common always-active reactions and corresponding enzymes identified in both *M. circinelloides* and *M. alpina* are presented in Supplementary File 10.

Interestingly, we observed that several of the sometimes-active reactions were associated with fatty acid synthesis and lipid metabolism (Figure 5), including ATP-citrate lyase (EC: 2.3.3.8) and acetyl-CoA synthetase (EC: 6.2.1.1), which are responsible for supplying acetyl-CoA for fatty acid synthesis. NADP-malic enzyme (EC: 1.1.1.40), which provides NADPH for fatty acid synthesis, and isocitrate dehydrogenase (EC: 1.1.1.42), which plays an indirect role in lipid metabolism via the TCA cycle, were also identified. In addition, key enzymes involved in β -oxidation, such as 3-hydroxyacyl-CoA dehydrogenase (EC: 1.1.1.35), enoyl-CoA hydratase (EC: 4.2.1.17) and long-chain fatty acyl-CoA ligase (EC: 6.2.1.3) were found. For enzymes involved in other related metabolisms, such as leucine degradation enzyme, hydroxymethylglutaryl-CoA lyase (EC: 4.1.3.4), glycine degradation enzyme, and D-amino acid oxidase (EC: 1.4.3.3), were also noticed.

Unique sometimes-active reactions in individual oleaginous species were investigated to understand the basis of lipid accumulation. In *M. alpina*, unique sometimes-active reactions and corresponding enzymes were associated with glycerolipid synthesis (Figure 5). These included diacylglycerol O-acyltransferase (EC: 2.3.1.20), which is directly involved in triacylglycerol formation, as well as carnitine O-acetyltransferase (EC: 2.3.1.7), which promotes long-chain fatty acid transport into mitochondria for β -oxidation, and triacylglycerol lipase (EC: 3.1.1.3), which is required for triacylglycerol hydrolysis. These pathways are consistent with mechanisms of lipid synthesis and regulation in *M. alpina* reported by Chen et al. [33]. In addition, acetyl-CoA acetyltransferase (EC: 2.3.1.9) involved in fatty acid degradation, glycosyl transferase (EC: 2.4.1.-) and ribose-phosphate pyrophosphokinase II (EC: 2.7.6.1), involved in carbohydrate metabolism, and

3-deoxy-7-phosphoheptulonate synthase (EC: 2.5.1.54), associated with nitrogen metabolism, were also identified.

Unique sometimes-active reactions and corresponding enzymes associated with glycerophospholipid synthesis were also identified for *M. circinelloides* (Figure 5). These included glycerol-3-phosphate 1-O-acyltransferase (EC: 2.3.1.15), 1-acylglycerol-3-phosphate O-acyltransferase (EC: 2.3.1.51), and phosphatidate phosphatase (EC: 3.1.3.4). In addition, acetyl-CoA carboxylase (EC: 6.4.1.2) and fatty acid synthase (EC: 2.3.1.86), involved in fatty acid synthesis, and key acetyl-CoA and NADPH-generating enzymes, such as pyruvate dehydrogenase complex (EC: 1.2.4.1; 1.8.4.1; 2.3.1.12), 6-phosphogluconate dehydrogenase (EC: 1.1.1.44) and glucose-6-phosphate dehydrogenase (EC: 1.1.1.49), were also found. The always- and sometimes-active reactions identified in *M. circinelloides* and/or *M. alpina* that are associated with central carbon metabolism via lipid synthesis are summarized in Figure 5.

4. Discussion

The availability of sophisticated bioinformatics and systems biology tools has resulted in a paradigm shift in our understanding of cellular metabolism of oleaginous microorganisms. The annotated genome sequences of oleaginous microorganisms have led to construction of whole-cell metabolic models; these provide powerful analytical scaffolds from which the behavior of complex metabolic networks in relation to their obese phenotypes can be queried. Few published genome-scale metabolic models of oleaginous species, such as *Y. lipolytica* and *M. alpina*, currently exist. Herein, we reconstructed the *iWV1213* model of *M.*

circinelloides with improved detailed descriptions of metabolic reactions for *in silico* cell simulation. Furthermore, we demonstrated the validity of the *iWV1213* model for predicting growth performance under a range of cultivation conditions to determine the quantitative relationship between glucose uptake rate, nitrogen uptake rate, oxygen uptake rate, and cell growth. Interestingly, the cell growth of *M. circinelloides* under oxygen limitation, in which a yeast-like cell was developed [30], could be demonstrated by the genome-scale metabolic model. The predicted results were consistent with previous reports that several factors, such as carbon and nitrogen content, oxygen supply, and specific growth rate affected cell growth, which could impact lipid production under particular conditions [34]. Thus, the model offers an efficient tool for further investigation into quantitative genotype-phenotype relationships underlying cellular metabolism, and for prediction of feasible and optimal metabolic flux distribution for lipid production in oleaginous fungi.

The comparative analysis of three genome-scale metabolic models of oleaginous yeast and fungal species indicated that *M. circinelloides* displayed a higher numbers of genes involved in carbohydrate, amino acid, and lipid metabolisms. In particular, those coding for hydrolytic enzymes in carbohydrate metabolism, which may explain its ability to utilize a range of carbon sources, such as maltose, sucrose, lactose, raffinose, and starch. Sucrose, lactose, and starch have been reported to be very poor carbon sources for cell growth of *M. alpina* [35], whereas *Y. lipolytica* preferentially uses glucose and sucrose as carbon sources rather than other carbohydrates [36]. Recently, the existence of genes encoding thirteen endoglucanases, three β -D-glucosidases, and several other glycoside hydrolases in *M. circinelloides* has been

reported [7], which supports a potential capacity to utilize a wide range of substrates from the surrounding environment as nutrients for cell growth.

However, although a large number of genes, enzymes, and reactions were present in these oleaginous species, FVA indicated that not all associated metabolic fluxes were active. *M. circinelloides* and *M. alpina* shared a set of common active metabolic routes, which belong predominantly to central metabolic pathways and amino acid metabolisms (Supplementary File 10). The 12 precursor metabolites associated with 15 common always-active reactions found in oleaginous Zygomycetes are in agreement with a previous report showing a shift in metabolism during lipid accumulation in *Y. lipolytica* through the observed amino acid, glycolytic, and TCA cycle metabolites [37]. Interestingly, reactions contributing to oleaginicacy through carbohydrate metabolism (e.g. ATP-citrate lyase, acetyl-CoA synthetase, NADP-malic enzyme, and hydroxy methylglutaryl-CoA lyase), lipid metabolism (e.g. 3-hydroxyacyl-CoA dehydrogenase, enoyl-CoA hydratase and long-chain fatty acyl-CoA ligase), and amino acid metabolism (e.g. hydroxy methylglutaryl-CoA lyase) were identified as common sometimes-active reactions, consistent with previous reports [38]. These findings suggest that reactions leading to oleaginicacy or lipogenesis might be dependent on regulatory mode(s), growth phase, and surrounding conditions. However, a number of unique sometimes-active reactions found in the individual oleaginous fungi might also help to discriminate their phenotypes. A large set of unique sometimes-active reactions was found in *M. alpina* (444 reactions, see Supplementary File 11) that were distributed across several metabolisms including carbohydrate, lipid, amino acid, and nucleotide metabolisms. In addition, the unique sometimes-active reactions found in *M. alpina* carbohydrate metabolism

(i.e. glycosyl transferase, group 1, and ribose-phosphate pyrophosphokinase II), lipid metabolism (i.e. acetyl-CoA acetyltransferase and diacylglycerol O-acyltransferase), and nitrogen metabolism (i.e. 3-deoxy-7-phosphoheptulonate synthase) corresponded to the unique always-active reactions present in *M. circinelloides*, and vice versa. We propose that these sometimes-active reactions are targets for bioengineering to maximize lipid accumulation in individual fungi.

Accordingly, we demonstrate that FVA can be applied to genome-scale models to identify a set of gene/enzyme candidates under the given conditions not only for biomass production, but also for overproduction of lipid or lipid-based products in oleaginous Zygomycetes.

However, further experimental validation of metabolic fluxes is required. A combinatorial approach should be applied to address oleaginicinity at a systems level to enable the rational design of efficient cell factories for the bioproduction of microbial lipids.

In summary, this study provides a novel genome-scale metabolic model of *M. circinelloides* for studying phenotypic behavior. We demonstrate its capabilities in constraint-based *in silico* analysis to compute growth phenotypes of this oleaginous fungi. A combination of this whole-cell model with multi-level omics data could be used to understand the relationship between environmental and genetic factor and phenotype, and guide the development of strains optimized for biomass and/or lipid production.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) (grant no. 31200989), The Thailand Research Fund (TRG5880245), Preproposal Research Fund (PRF4/2558), and Department of Zoology, Faculty of Science, Kasetsart University. The authors would like to thank Nam Ninh Nguyen for useful discussion on improved metabolic annotation. We also thank Dr. Jittisak Senachak for data arrangement. The genome sequence data of *M. circinelloides* has been produced by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) in collaboration with the user community.

References

- [1] Salimon, J., Salih, N., Yousif, E., Industrial development and applications of plant oils and their biobased oleochemicals. *Arabian J Chem* 2012, 5, 2.
- [2] Liang, M. H., Jiang, J. G., Advancing oleaginous microorganisms to produce lipid via metabolic engineering technology. *Prog Lipid Res* 2013, 52, 395-408.
- [3] Ratledge, C., Wynn, J. P., The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. *Adv Appl Microbiol* 2002, 51, 1-51.
- [4] Ratledge, C., Yeasts, moulds, algae and bacteria as sources of lipids In: Kamel BS, Kakuda Y, editors., *Technological Advances in Improved and Alternative Sources of Lipids*, London: Blackie Academic and Professional 1994, pp. 235-291.
- [5] Beopoulos, A., Cescut, J., Haddouche, R., Uribe Larrea, J. L., et al., *Yarrowia lipolytica* as a model for bio-oil production. *Prog Lipid Res* 2009, 48, 375-387.
- [6] Huang, C., Chen, X.-f., Xiong, L., Chen, X.-d., et al., Single cell oil production from low-cost substrates: The possibility and potential of its industrialization. *Biotechnol Advances* 2013, 31, 129–139.
- [7] Wei, H., Wang, W., Yarbrough, J. M., Baker, J. O., et al., Genomic, proteomic, and biochemical analyses of oleaginous *Mucor circinelloides*: evaluating its capability in utilizing cellulolytic substrates for lipid production. *PLoS One* 2013, 8, e71068.
- [8] Xia, C., Zhang, J., Zhang, W., Hu, B., A new cultivation method for microbial oil production: cell pelletization and lipid accumulation by *Mucor circinelloides*. *Biotechnol Biofuels* 2011, 4, 15.
- [9] Vongsangnak, W., Ruenwai, R., Tang, X., Hu, X., et al., Genome-scale analysis of the metabolic networks of oleaginous Zygomycete fungi. *Gene* 2013, 521, 180-190.
- [10] Vongsangnak, W., Figueiredo, L. F., Forster, J., Weber, T., et al., Genome-scale metabolic representation of *Amycolatopsis balhimycina*. *Biotechnol Bioeng* 2012, 109, 1798-1807.

-
- 442 [11] Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., et al., Data, information, knowledge
443 and principle: back to metabolism in KEGG. *Nucleic Acids Res* 2014, 42, D199-205.
- 444 [12] Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., et al., InterProScan: protein
445 domains identifier. *Nucleic Acids Res* 2005, 33, W116-120.
- 446 [13] Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., et al., A comprehensive
447 evolutionary classification of proteins encoded in complete eukaryotic genomes.
448 *Genome Biol* 2004, 5, R7.
- 449 [14] Finn, R. D., Bateman, A., Clements, J., Coggill, P., et al., Pfam: the protein families
450 database. *Nucleic Acids Res* 2014, 42, D222-230.
- 451 [15] Marchler-Bauer, A., Panchenko, A. R., Shoemaker, B. A., Thiessen, P. A., et al., CDD: a
452 database of conserved domain alignments with links to domain three-dimensional
453 structure. *Nucleic Acids Res* 2002, 30, 281-283.
- 454 [16] Yu, C. S., Chen, Y. C., Lu, C. H., Hwang, J. K., Prediction of protein subcellular
455 localization. *Proteins* 2006, 64, 643-651.
- 456 [17] Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., et al., The EcoCyc and MetaCyc
457 databases. *Nucleic Acids Res* 2000, 28, 56-59.
- 458 [18] Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., et al., Quantitative prediction of
459 cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat*
460 *Protoc* 2011, 6, 1290-1307.
- 461 [19] Funtikova, N. S., Mysiakina, I. S., Konova, I. V., Synthesis of biologically active lipids
462 by fungus *Mucor lusitanicus* 306D grown on media with various composition. *Prikl*
463 *Biokhim Mikrobiol* 2002, 38, 644-648.
- 464 [20] Anderson, T. H., Assessment of DNA contents of soil fungi. *Landbauforschung - vTI*
465 *Agriculture and Forestry Research* 2008, 58, 19-28.
- 466 [21] Nielsen, J., Physiological engineering aspects of *Penicillium chrysogenum*, World
467 Scientific Pub Co Inc 1996.

-
- [22] Bredenkamp, A., Velankar, H., van Zyl, W. H., Gorgens, J. F., Effect of dimorphic regulation on heterologous glucose oxidase production by *Mucor circinelloides*. *Yeast* 2010, 27, 849-860.
- [23] McIntyre, M., Breum, J., Arnau, J., Nielsen, J., Growth physiology and dimorphism of *Mucor circinelloides* (syn. *racemosus*) during submerged batch cultivation. *Appl Microbiol Biotechnol* 2002, 58, 495-502.
- [24] Lübbehüsen, T. L., Nielsen, J., McIntyre, M., Aerobic and anaerobic ethanol production by *Mucor circinelloides* during submerged growth. *Appl Microbiol Biotechnol* 2004, 63, 543-548.
- [25] Edwards, J. S., Ramakrishna, R., Palsson, B. O., Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng* 2002, 77, 27-36.
- [26] Ye, C., Xu, N., Chen, H., Chen, Y. Q., et al., Reconstruction and analysis of a genome-scale metabolic model of the oleaginous fungus *Mortierella alpina*. *BMC Syst Biol* 2015, 9, 1.
- [27] Pan, P., Hua, Q., Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, *Yarrowia lipolytica*. *PLoS One* 2012, 7, e51535.
- [28] Mahadevan, R., Schilling, C. H., The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 2003, 5, 264-276.
- [29] Gudmundsson, S., Thiele, I., Computationally efficient flux variability analysis. *BMC Bioinformatics* 2010, 11, 489.
- [30] Lübbehüsen, T. L., Nielsen, J., McIntyre, M., Morphology and physiology of the dimorphic fungus *Mucor circinelloides* (syn. *M. racemosus*) during anaerobic growth. *Mycol Res* 2003, 107(Pt 2), 223-230.
- [31] Ruenwai, R., Cheevadhanarak, S., Rachdawong, S., Tanticharoen, M., Laoteng, K., Oxygen-induced expression of $\Delta 6$ -, $\Delta 9$ - and $\Delta 12$ -desaturase genes modulates fatty acid composition in *Mucor rouxii*. *Appl Microbiol Biotechnol* 2010, 86, 327-334.

-
- [32] Kanehisa, M., Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000, 28, 27-30.
- [33] Chen, H., Hao, G., Wang, L., Wang, H., et al., Identification of a critical determinant that enables efficient fatty acid synthesis in oleaginous fungi. *Sci Rep* 2015, 10, 11247.
- [34] Laoteng, K., Čertík, M., Cheevadhanark, S., Mechanisms controlling lipid accumulation and polyunsaturated fatty acid synthesis in oleaginous fungi. *Chem Pap* 2011, 65, 97-103.
- [35] Bajpai, P. K., Bajpai, P., Ward, O. P., Arachidonic acid production by fungi. *Appl Environ Microbiol* 1991, 57, 1255-1258.
- [36] Soccol, C. R., Vandenberghe, L. P. S., Pandey, A., New perspectives for citric acid production and application. *Food Technol Biotechnol* 2006, 44, 141–149.
- [37] Pomraning, K. R., Wei, S., Karagiosis, S. A., Kim, Y. M., et al., Comprehensive metabolomic, lipidomic and microscopic profiling of *Yarrowia lipolytica* during lipid accumulation identifies targets for increased lipogenesis. *PLoS One* 2015, 10, e0123188.
- [38] Vorapreeda, T., Thammarongtham, C., Cheevadhanarak, S., Laoteng, K., Alternative routes of acetyl-CoA synthesis identified by comparative genomic analysis. *Microbiology* 2012, 158, 217-228.
- [39] Vicente, G., Bautista, L. F., Rodríguez, R., Gutiérrez, F. J., et al., Biodiesel production from biomass of an oleaginous fungus. *Biochem Eng J* 2009, 48, 22-27.
- [40] Jackson, F. M., Fraser, T. C., Smith, M. A., Lazarus, C., et al., Biosynthesis of C18 polyunsaturated fatty acids in microsomal membrane preparations from the filamentous fungus *Mucor circinelloides*. *Eur J Biochem* 1998, 252, 513-519.
- [41] Pedersen, H., Carlsen, M., Nielsen, J., Identification of enzymes and quantification of metabolic fluxes in the wild type and in a recombinant *Aspergillus oryzae* strain. *Appl Environ Microbiol* 1999, 65, 11-19.

520 [42] Peter, M. G., Chitin and Chitosan in Fungi, In Biopolymers Online, Wiley-VCH Verlag
521 GmbH & Co. KGaA 2005.

522 [43] Orlowski, M., *Mucor* dimorphism. Microbiol Rev 1991, 55, 234-258.

523

524

525

526

527

528

Figure legends

Figure 1. Comparison of *in silico* and *in vivo* biomass growth rates of *M. circinelloides* on glucose and galactose substrates at different uptake rates.

Figure 2. PhPP plots of metabolic phenotypic behaviors of *M. circinelloides*. (A) Effect of glucose and nitrogen uptake rates on biomass growth rate (B) Effect of glucose and oxygen uptake rates on biomass growth rates.

Figure 3. Comparison of three oleaginous genome-scale metabolic models. (A) Reaction number distribution in different metabolisms. (B) Gene number distribution in different metabolisms.

Figure 4. Analysis of the distribution of three different reactions (always-active reactions, sometimes-active reactions, and never-active reactions) between two oleaginous Zygomycete fungi by FVA.

Figure 5. Diagram highlighting crucial common active routes related to growth, and unique routes related to fatty acid and lipid metabolism between two oleaginous Zygomycete fungi.

Note: Bold and underlined EC numbers highlight routes related to oleagenicity.

Table 1. Biomass composition in the genome-scale metabolic model of *M. circinelloides*.

Biomass components	Molecular weight (g/mol)	Actual composition (% w/w)	Normalized composition (% w/w)	Biomass stoichiometry	References
Protein	135.89	36.50	42.44	3.1233	[19]
Nucleotide	-	1.70	1.98	-	[19]
- DNA	333.2	0.33	0.38	0.0115	[20]
- RNA	340.6	1.37	1.59	0.0468	
Lipid	-	14.80	17.21	-	[19, 39, 40]
Carbohydrate	-	33.00	38.37	-	[19]
- Glycogen	666.6	0.09	0.11	0.0017	[41]
- Chitin	34500	7.65	8.9	0.0026	[42]
- Chitosan	1200000.0	6.28	7.3	0.0001	[42]
- Others	-	18.97	22.06	-	[43]
Total	-	86	100	-	

Table 2. List of substrates that can be utilized in the *M. circinelloides* genome-scale metabolic model.

Substrates	<i>In silico</i> data	<i>In vivo</i> data
Carbon sources[*]		
Glucose	+	+
Galactose	+	+
Xylose	+	+
Maltose	+	+
Lactose	+	+
Glycerol	+	+
Sucrose	+	+
Cellulose	+	+
Cellobiose	+	+
Starch	+	+
Nitrogen sources[#]		
NH ₄ Cl	+	+
NH ₄ NO ₃	+	+
(NH ₄) ₂ SO ₄	+	+
KNO ₃	+	+
Urea	+	+

^{*}Each carbon source was simulated when NH₃ was used as nitrogen source. [#]Each nitrogen source was simulated when glucose was used as carbon source. Each substance was set uptake rate of 1.0 mmol gDW⁻¹h⁻¹.

Table 3. Comparison of the characteristics of the *M. circinelloides* model (iWV1213) with other oleaginous genome-scale metabolic models.

Oleaginous species	<i>M. circinelloides</i>	<i>M. circinelloides</i> ¹	<i>M. alpina</i> ²	<i>Y. lipolytica</i> ³
Model name	iWV1213	-	iCY1106	iYL619_PCP
Network characteristics	This study	Earlier work		
Genes	1,213	1,208	1,106	619
EC numbers	626	581	626	428
Metabolites	1,413	1,426	1,732	843
Compartments	5	5	4	3
Total metabolic reactions	1,326	1,111	1,854	1,142
Transport and exchange reactions	178	0	463	361
Modeling and simulation	Glucose (GUR=1.52)	-	Glucose (GUR=2.00)	Glucose (GUR=20.00)
Growth rate (h ⁻¹)	0.1102	-	0.1546	0.0439

¹Metabolic network of *M. circinelloides* [9], ²SBML model of *M. alpina* [26], ³SBML model of *Y. lipolytica* [27]. Note: GUR stands for glucose uptake rate (mmol gDW⁻¹h⁻¹).

Table 4. List of common always-active reactions associated with the biosynthesis of 12 precursor metabolites for growth that are shared in *M. circinelloides* and *M. alpina*.

EC numbers	Common always-active reactions	Subsystems
2.7.1.1	GLC + ATP -> ADP + G6P	Glycolysis/gluconeogenesis
2.7.2.3	ADP + 13PDG <=> ATP + 3PG	Glycolysis/gluconeogenesis
4.2.1.11	2PG <=> PEP + H2O	Glycolysis/gluconeogenesis
2.7.1.40	ADP + PEP -> ATP + PYR	Glycolysis/gluconeogenesis
4.1.2.13	FDP <=> T3P2 + T3P1	Glycolysis/gluconeogenesis
5.3.1.9	G6P <=> F6P	Glycolysis/gluconeogenesis
2.3.3.10	COA + H3MCOA <=> ACCOA + AACCOA	Biosynthesis of lanosterol
2.2.1.1	XUL5P + R5P <=> T3P1 + S7P	Pentose-phosphate pathway
2.2.1.1	XUL5P + E4P <=> F6P + T3P1	Pentose-phosphate pathway
2.2.1.2	T3P1 + S7P <=> F6P + E4P	Pentose-phosphate pathway
5.3.1.6	RL5P <=> R5P	Pentose-phosphate pathway
6.4.1.1	ATP + PYR + HCO3 -> ADP + PI + OA	Pyruvate metabolism
1.2.4.2: 2.3.1.61: 1.8.1.4	NAD + COA + AKG -> NADH + CO2 + SUCCOA	Tricarboxylic acid cycle
1.5.1.7	NAD + SACP <=> NADH + AKG + LYS	Lysine metabolism
2.6.1.57	GLU + PHPYR <=> AKG + PHE	Phenylalanine, tyrosine and tryptophan biosynthesis

Note: The 12 precursors are glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), glyceraldehyde 3-phosphate (T3P1), 3-phosphoglycerate (3PG), phosphoenolpyruvate (PEP), pyruvate (PYR), ribose-5-phosphate (R5P), erythrose-4-phosphate (E4P), acetyl-CoA (ACCOA), alpha-ketoglutarate (AKG), oxaloacetate (OA), and succinyl-CoA (SUCCOA).

Figures

Figure 1

Figure 2

Figure 3

(A)

(B)

Figure 4

Figure 5