Comprehensive reconstruction and evaluation of *Pichia pastoris* genome-scale metabolic model that accounts for 1243 ORFs

Rui Ye, Mingzhi Huang*, Hongzhong Lu, Jiangchao Qian, Weilu Lin, Ju Chu, Yingping Zhuang and Siliang Zhang

**Abstract**

**Background:** *Pichia pastoris* is one of the most important cell factories for production of industrial enzymes and heterogenous proteins. The genome-scale metabolic model of high quality is crucial for comprehensive understanding of the *P. pastoris* metabolism.

**Methods:** In this paper, we upgraded *P. pastoris* genome-scale metabolic model based on the combination of latest genome annotations and literatures. Then the performance of the new model was evaluated using the Cobra Toolbox v2.0.

**Results:** Compared with the recently published model iMT1026, the reaction number in the new model iRY1243 was increased from 2035 to 2407 and the metabolite number was increased from 1018 to 1094. Accordingly, the unique ORF number was increased from 1026 to 1243. To improve the metabolic functions of *P. pastoris* genome-scale metabolic model, the biosynthesis pathways of vitamins and cofactors were carefully added. iRY1243 showed good performances when predicting the growth capability on most of the reported carbon and nitrogen sources, the metabolic flux distribution with glucose as a sole carbon source, the essential and partially essential genes, and the effects of gene deletion or overexpression on cell growth and S-adenosyl-L-methionine production.

**Conclusion:** iRY1243 is an upgraded *P. pastoris* genome-scale metabolic model with significant improvements in the metabolic coverage and prediction ability, and thus it will be a potential platform for further systematic investigation of *P. pastoris* metabolism.

**Keywords:** Genome-scale metabolic model, *Pichia pastoris*, Multi-omics, β-Galactosidase

---

*Correspondence: huangmz@ecust.edu.cn
State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, No.130, Meilong Road, Shanghai 200237, China

© The Author(s) 2017. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
predicting engineering target for improving the titer of industrial cell factories.

The earliest metabolic network model of \textit{P. pastoris} is the small model of central carbon metabolism network model (Maaheimo et al. 2001), which is used to calculate the $^{13}$C metabolic flux distribution (Sola et al. 2004). The gene sequencing of \textit{P. pastoris} has laid the foundation for GSMM construction (Mattanovich et al. 2009; Kuberl et al. 2011). There are different \textit{P. pastoris} GSMM models. Sohn et al. (2010), Chung et al. (2010), and Caspeta et al. (2012) established the PpaMBEL1254, iPP668, and iLC915 models, respectively. The model iMT1026 was reconstructed by merging PpaMBEL1254, iPP668, and iLC915 models, and adding the synthesis and decomposition pathway of fatty acid, sphingolipid biosynthesis pathway, oxidative phosphorylation, and glycosylation pathway (Tomas-Gamisans et al. 2016). Compared with the latest \textit{Saccharomyces cerevisiae} GSMMs, the quality of these \textit{P. pastoris} models still needs to be improved at least in three aspects: the balance of each reaction in mass and electrical charge, the model metabolic coverage, and the correction in GPRs. The biomass components of \textit{P. pastoris} (Carnicer et al. 2009) and the utilization of energy (Chung et al. 2010) have been reported. With the availability of new gene annotations and literatures about \textit{P. pastoris}, it is necessary to update the \textit{P. pastoris} GSMM models to improve the model performance.

In this paper, based on the latest gene annotations and the newly published literatures, we reconstructed a new genome-scale metabolic network model of \textit{P. pastoris}. The new model iRY1243 contained 1243 annotated genes, 2407 reactions, 1740 metabolites, and 9 reaction locations. At the same time, we updated the cell components of \textit{P. pastoris} based on the latest literatures. The model performance was evaluated by the data from RNA-Seq, chemostat experiments, and $^{13}$C labeling experiments. Model iRY1243 was then used to predict the secretion of products and essential genes on synthetic medium.

**Methods**

**Procedures for model reconstruction**

The procedures for reconstruction of \textit{P. pastoris} GSMM could be divided into two phases (Fig. 1). Firstly, by collating and evaluating the existing models, we upgraded the GSMM based on the existing iPP668, PpaMBEL1254, iLC915, and iMT1026 models and re-annotated the information of all metabolites by referring to the databases of KEGG and BiGG, to ensure that each metabolite had correct structure and charge in pH 7.2. Then, each reaction in models was manually checked and corrected, thus making it comparable with the published GSMM model.

Subsequently, to update the \textit{P. pastoris} GSMM systematically, the genome annotation information from mainly three databases (KEGG, UniProtKB, IMG) was gathered and merged (Table 1). To establish the new gene–protein–reaction associations (GPRs), the genes were collected from each database. With the aid of KEGG and Enzyme databases, the relationship between proteins and reactions were further checked to ensure the high quality of GPRs in the new \textit{P. pastoris} GSMM. Next, using the gapFind function of the Cobra Toolbox v2.0 to identify dead-end metabolites, essential reactions from KEGG and other databases were added to reduce the number of dead-end metabolites to improve the function of the metabolic network.

Lastly, the cell components and the energy parameter were updated. Compared with the primordial model, some new components of cell, like the cofactors including folate and thiamin, were added to ensure the function of the corresponding metabolic pathways. With the fitting of the data from chemostat cultivations, the non-growth-associated maintenance (NGAM) in the model was adjusted from 3.7 to 2.52 mmol ATP/gDCW h.

**Experimental procedures**

**Strains**

The \textit{Pichia pastoris} strains GS115 and G1HL were used to evaluate the model performance.

**Medium and culture conditions**

In this study, 50 mL YPG medium as the seed medium was taken in 500-mL flask (yeast extract 10%, peptone 20%, glycerol 20%). The composition of the medium used in chemostat cultivations was as follows: 10 g/L glucose, 9.1 g/L K2SO4, 0.46 g/L CaSO4, 7.5 g/L MgSO4·7H2O, 2.06 g/L KOH, 13.4 mL/L H3PO4, and 6 mL/L PTM1. The components of PTM1 were described by Baumann et al. (2008).

The culture conditions for seed culture and chemostat culture were described by Nie et al. (2014). Chemostat cultivation was carried out in a 5-L bioreactor (National Engineering Research Center for Biotechnology, Shanghai, China) with a working volume of 3 L. With glucose as the sole carbon source, the initial aeration rate was set to 1vvm, controlled by a mass flow controller (Sevenstar, Beijing, China). The dissolved oxygen (DO) was maintained above 30% by adjusting the agitation and the aeration to ensure fully aerobic conditions. The O2 and CO2 concentrations in the off-gas were measured online using a process mass spectrometer (MAX300-LG, Extrel, America). Pressure, pH, and temperature were controlled at 0.04 MPa, pH 5.5 (with adding NH3·H2O), and 30 °C, respectively. The specific growth rate of the cells was controlled by the dilution ratio (D).
**Determination of the concentration of cell, glucose, and other extracellular metabolites**

The cell concentration was measured from the optical density at 600 nm. For dry cell weight (DCW) measurement, 10 mL broth was filtered using a pre-dried and weighted microporous membrane (Shanghai Diqing Filtration Technology CO., LTD Shanghai, China). After being washed 3 times, the filter paper was dried at 105 °C.
for 3 h. The glucose kits (Shanghai Kexin Biotechnology Institute, China) were used for residual glucose concentration analysis. Extracellular metabolite measurements (such as ethanol, acetic acid, propionic acid, formic acid, pyruvic acid, and acetaldehyde) were performed as described in Nie et al. (2014).

**Sensitivity analysis**
During the sensitivity analysis of the GSMM, the specific glucose consumption rate was changed from 0 to 2.0 mmol/gDCW h. In each simulation, only one of the following model parameters was evaluated, including the lipid composition (6.2–16.2%), protein composition (37–47%), RNA composition (6.6–16.6%), DNA composition (0.1–1.0%), carbohydrate composition (26.9–36.9%), and NGAM value (1.26–3.78 mmol ATP/gDCW h). The specific biomass growth rate and oxygen uptake rate were simulated to evaluate the effects of the changes of parameters on the model prediction ability.

**Calculation of energy parameters**
The ATP required for cell survival can be calculated by chemostat experiments. The generated ATP can be divided into two forms: non-growth-associated ATP maintenance (NGAM) and growth-associated ATP maintenance (GAM). GAM required for biomass synthesis (i.e., precursor biosynthesis and polymerization) can be obtained from literature. NGAM, used for cell maintenance, is an independent reaction in the model, which can be calculated from chemostat data.

NGAM and GAM are added to calculate the total amount of ATP required for cell growth, as shown in the following equation:

\[
 r_{ATP} = Y_{xATP} \mu + m_{ATP},
\]
where \( Y_{xATP} \) and \( m_{ATP} \) represent the GAM and NGAM, respectively, \( \mu \) is the specific growth rate, and \( r_{ATP} \) represents the total amount of ATP consumed by *P. pastoris*.

**Biomass composition**
As the biomass composition has obvious effects on the model validation and the strain improvement, we updated the cell composition of *P. pastoris*. The biomass is composed of macromolecular substances (i.e., protein, lipid, DNA, RNA, carbohydrates, small-molecule pool). The details on the biomass components of *P. pastoris* were referred from Verduyn et al. (1991), Carnicer et al. (2009), De Schutter et al. (2009) and Tomas-Gamisans et al. (2016). The detailed information of biomass composition is described in the supporting information (see Additional file 2).

**Constraint-based FBA**
Constraint-based flux-balanced analysis (FBA) (Bodbar et al. 2014) was widely used in the reconstruction of GSMM model, analysis of strain capabilities under different environmental and genetic perturbations, and prediction of strain phenotype (Orth et al. 2010). Generally, FBA predicts metabolic fluxes based on maximization of objective functions (typically the specific growth rate or the target metabolite production rate) (Feist and Palsson 2010; Sanchez et al. 2012), and the basic framework of FBA is composed of variables, objectives, and constraints. The constraints used in FBA include the balance of intracellular metabolites (Eq. 2), the reversibility and the demarcation line of reactions, as well as the extracellular metabolite exchange rates (Eq. 3). In this study, all FBA was conducted with the cell growth rate as the objective function (Eq. 1). Each reaction had an upper and a lower bound on the flux it can carry. For reversible reaction, the upper and lower bounds were set at −1000 and 1000, respectively, while for irreversible reaction the lower bound was set at 0. FBA was conducted

| Table 1 | Information of databases used in this study |
|-------|------------------------------------------|
| Abbreviation | Full name | Purpose | Link |
| KEGG  | Kyoto encyclopedia of genes and genomes  | Obtain genome annotation, reaction information chemical compounds annotation | http://www.genome.jp/dbget-bin/www_bfind?compound |
| CheBI | Chemical entities of biological interest | Obtain chemical compounds annotation | https://www.ebi.ac.uk/chebi/ |
| BiGG  | Biochemical genetic and genomic knowledgebase | Obtain reaction information and chemical compounds annotation | http://bigg.ucsd.edu/bigg/main.pl |
| Ensembl fungi | Ensembl fungi | Obtain genome annotations | http://fungi.ensembl.org/index.html |
| UniProtKB | The universal protein resource knowledgebase | Obtain genome annotations | http://www.uniprot.org/ |
| IMG   | The integrated microbial genomes system | Obtain genome annotations | http://img.jgi.doe.gov/ |
| ENZYME | Enzyme nomenclature database | Obtain relations between enzyme and reactions | http://enzyme.expasy.org/ |
using Cobra Toolbox v2.0 (Schellenberger et al. 2011) and Gurobi 6 solver based on Matlab.

Objective: \[ \text{max/min } Z = C^T \ast \nu \]  
(1)

Constraints: \[ S \ast \nu = 0 \]  
(2)

\[ lb \leq \nu \leq ub, \]  
(3)

where \( S \) is a matrix of \( m \ast n \) and \( m \) and \( n \) are the numbers of metabolites and reactions, respectively. In Eq. 1, \( C^T \) represents the coefficient of metabolites in the objective function. In Eq. 3, \( \nu \) is the rate of all reactions. \( lb \) and \( ub \) defined the lower and upper bounds of the flux for each reaction, respectively.

During the simulation with iRY1243, NGAM and GAM were set at 2.52 mmol ATP/gDCW h and 20.4 mmol ATP/gDCW, respectively (see Additional file 3). The lower bound of exchange reactions except for glucose, ammonia, oxygen, sulfur, phosphorus, and Fe\({}^{2+}\) was set at 0. To predict the specific growth rate (\( \mu \)), the exchange rates of glucose were set at \(-0.478, -0.692, -0.942, -1.294, \) and \(-1.846 \) mmol/gDCW h, respectively. The exchange rates of other substrates and by-products were set at the corresponding values at each dilution rate.

Prediction of growth-supporting carbon and nitrogen sources

To investigate the prediction capability of iRY1243, 30 carbon and 21 nitrogen sources for \( P. \) pastoris were collected. FBA was used to analyze the growth capability on each carbon or nitrogen source. For prediction of carbon utilization, only NH\(_4\)\(^+\) was set as the nitrogen source, while the phosphorus and sulfur sources were maintained sufficient. At the same time, the metabolites of the other exchange reactions containing carbons were set as zero. For prediction of nitrogen utilization, glucose was used as the only carbon source. The target carbon or nitrogen source was considered growth supporting if the predicted specific growth rate was above zero.

Prediction of essential genes

The prediction of essential genes and partially essential genes was conducted using the singleGeneDeletion function based on Cobra Toolbox. During the essential gene analysis, according to the matrix of GPRs and the Boolean rule definition, the fluxes of reactions containing the relating genes were set at zero, while the bounds of other reactions were maintained constant. Based on the simulated specific growth rate, the genes were classified into essential genes (the calculated specific growth rate was 0), partially essential genes (the calculated specific growth rate of \( P. \) pastoris was in the range from 0 to the maximum specific growth rate), and non-essential genes (the calculated specific growth rate was the same as the maximum specific growth rate). The synthetic medium, which was made up of glucose, oxygen, ammonia, sulfur, and phosphorus, was used to predict the essential genes for \( P. \) pastoris growth.

Transcriptome analysis

The strain \( P. \) pastoris G1HL was used. The medium and procedures for seed culture and chemostat cultivations were described in Nie et al. (2014) with glucose as the only carbon source.

During the stable phase of \( P. \) pastoris G1HL culture, about 20 mL broth was taken into three 50-mL centrifuge tubes. Samples were centrifuged at 4000 rpm for 10 min at 4 °C and supernatant was discarded. The harvested cells were washed three times with 0.9% NaCl and dried with a filter paper. The samples were stored in a −80 °C refrigerator. The RNA extraction and sequence, read mapping, quantification, and data analysis were accomplished by KangChen Bio-tech Inc. (Shanghai, China).

Results and discussion

Reconstruction of the \( P. \) pastoris GSMM iRY1243

The update of the \( P. \) pastoris GSMM was based on the iPP668, PpaMBEL1254, iLC915, and iMT1026 models. The procedure for reconstruction of \( P. \) pastoris is shown in Fig. 1. First, the elements and charge in all reactions were balanced based on the metabolites’ annotation information. Gap analysis was also conducted and gaps were reduced by adding essential reactions from KEGG and other databases. To overcome the deficiencies of these models, \( P. \) pastoris gene annotation information from three common databases (KEGG, UniProtKB, and IMG) was sorted and compared. Among them, KEGG was the largest metabolic network database.

Comparison of iRY1243 with the other models

As shown in Table 2, iRY1243 contains 2407 reactions and 1740 metabolites, which is more than that in the model of iMT1026 (Tomas-Gamisans et al. 2016), indicating that this upgrade enlarged the scale of the \( P. \) pastoris GSMM. The ORF number in iRY1243 is 1243, which is also more than that in iMT1026. Based on the latest gene annotations, the model iRY1243 contains more genes compared to the previous model iMT1026, such as the genes relating to the metabolism of porphyrin and chlorophyll, biotin, the membrane transport reactions, etc. The biomass contains the products of these pathways, which are crucial for the prediction of essential genes. iRY1243 is an upgraded \( P. \) pastoris GSMM with significant improvements in the metabolic coverage. In conclusion, the \( P. \) pastoris GSMM was comprehensively improved based on gene annotations from different databases and literatures (Additional file 1).
Accurate cell composition and NGAM value were important for improving the metabolic functions of GSMM. By referring to the latest GSMM of *P. pastoris* and *Saccharomyces cerevisiae*, the essential cell trace components were added, including the small-molecule pool (see Additional file 2). The NGAM was calculated by chemostat data (see Additional file 3).

**Validation of GSMM**

**Model verification by RNA-Seq**

The gene in the newly updated *P. pastoris* GSMM iRY1243 was firstly verified with the latest RNA sequence data. The RNA-Seq results showed that the expression of 2393 genes could be determined in the sampling condition. The iRY1243 contains 1243 genes and the expression of most genes (895 genes, 73.42%) could be verified according to the results of RNA-seq (Fig. 2a) when glucose was used as the sole carbon source. Removing the exchange reactions, among the remaining 1773 reactions, the expression of genes from 14.33% (254 reactions) of these reactions was not measured (Fig. 2b). After removing the exchange reactions and reactions without the corresponding annotated genes, there are 1042 and 471 reactions associated with single and multiple genes, respectively. Based on the transcriptome analysis, about 79.37% of the single-gene reactions and 91.93% of the multi-gene reactions were verified (Fig. 2c, d), indicating that the existence of most reactions in the model iRY1243 was reliable.

**Sensitivity analysis of iRY1243**

During sensitivity analysis, the influence of different biomass components (lipid, protein, RNA, DNA, carbohydrate) and the energy parameters (NGAM) on the prediction accuracy of the strain phenotype was comprehensively investigated. As shown in Fig. 3, the specific growth rate (μ) and the specific oxygen uptake rate (qO2) were hardly affected by the changes of the protein, RNA, DNA, and carbohydrate contents. However, as the content of lipid increased, μ decreased. The reason may be that the synthesis of fatty acids requires more ATP and NADPH. Different from protein, RNA, DNA, carbohydrate, and the predicted μ and qO2 were more obviously affected by NGAM changes. As NGAM increased, the predicted μ decreased rapidly, while the qO2 increased accordingly as shown in Fig. 3f.

**Model verification by growth capabilities on different carbon and nitrogen sources**

To comprehensively evaluate the prediction ability of the model iRY1243, *P. pastoris* phenotype data reported from 1998 to 2017 were collected (Menendez et al. 1998; Hsieh et al. 2010; Klompmaker et al. 2010; Sohn et al. 2010; Zhang et al. 2015), together with physiological data from our lab. *P. pastoris* could grow on a series of carbon sources and nitrogen sources. The growth-supporting 30 carbon and 21 nitrogen sources can be found in Tables 3 and 4. The in silico growth capabilities of *P. pastoris* on these carbon and nitrogen sources were checked using iRY1243. As shown in Tables 3 and 4, growth on 30 carbon and 21 nitrogen sources could be predicted using iRY1243. Compared to model iMT1026 (Chung et al. 2010; Sohn et al. 2010; Caspeta et al. 2012; Nocon et al. 2014; Irani et al. 2016; Tomas-Gamisans et al. 2016), the prediction ability of iRY1243 was better.

**Model verification by physiological growth parameters**

The maximization of cell growth rate usually served as the objective function in FBA. Thus, a comparison between the measured and predicted μ values could show

---

**Table 2 Comparison among different GSMMs of *P. pastoris***

|          | iPP668 | PpaMBEL1254 | iLC915 | iMT1026 | iRY1243 |
|----------|--------|-------------|--------|---------|---------|
| Genes    | 668    | 540         | 915    | 1026    | 1243    |
| Metabolites | 1177  | 1058        | 1302   | 1689    | 1740    |
| Reactions | 1354   | 1254        | 1423   | 2035    | 2407    |
| Cytosolic | 623    | 604         | 790    | 1059    | 1316    |
| Mitochondrial | 163   | 155         | 205    | 268     | 313     |
| Peroxisomal | 66    | 66          | 64     | 102     | 102     |
| Extracellular | 12    | 11          | 0      | 16      | 5       |
| Endoplasmic reticulum | 15    | 7           | 34     | 41      | 43      |
| Golgi apparatus | 4     | 8           | 4      | 13      | 14      |
| Vacuolar  | 3      | 6           | 12     | 9       | 10      |
| Nuclear   | 16     | 17          | 0      | 17      | 17      |
| Intercompartmental/transport | 452  | 328         | 314    | 510     | 540     |
the quality of iRY1243. The reported NGAM of \textit{P. pastoris} was 2.26 mmolATP/gDCW h (Chung et al. 2010), while, according to our calculation based on $^{13}$C labeled experiment, the NGAM was 9.48 mmolATP/gDCW h for \textit{P. pastoris} G1HL (Nie et al. 2014). Due to the limitations of the central carbon metabolic network model used to calculate the fluxes of pathways, the model cannot contain all the ATP reactions. As the NGAM had an obvious effect on the specific growth rate, the NGAM in the iRY1243 was firstly calculated using \textit{P. pastoris} G1HL chemostat cultivation data without by-product formation. We further compared the predicted and measured values with a series of chemostat cultivations using \textit{P. pastoris} G1HL chemostat cultivation data without by-product formation. We further compared the predicted and measured values with a series of chemostat cultivations using \textit{P. pastoris} G1HL producing beta-galactosidase with the new NGAM value (2.52 mmol ATP/gDCW h). As shown in Fig. 4, iRY1243 could predict $\mu$, $q_{o2}$, $q_{co2}$, and RQ under a range of the specific glucose consumption rate inputs.

To compare the function of each model, the chemostat data published by Carnicer et al. (2012) were used to verify the predictive function of the models (see Additional file 4). FBA was used to compare the predictive ability of each model with the maximization of cell growth rate as the objective function. The results showed that iRY1243 and iMT1026 could predict \textit{P. pastoris} specific growth rate and CO2 specific consumption rate. However, the model iMT1026 predicted some unreasonable by-products (citric acid, etc.).

**Model verification by in vivo $^{13}$C fluxes**

To further evaluate the prediction performance of iRY1243, the simulated fluxes using pFBA were compared with the calculated metabolic flux from $^{13}$C labeled experiments (Nie et al. 2014). As shown in Fig. 5, the correlation coefficient between the simulated fluxes by GSMM and the calculated $^{13}$C fluxes for G1HL was...
0.88, initially showing the good performance of iRY1243. However, there are still significant differences at some points, which may be caused due to the following reasons. Firstly, the algorithms used to calculate the fluxes were totally different between FBA and $^{13}$C central flux analysis. Secondly, there is a great difference in the size of the two models for the $^{13}$C flux model only containing the central carbon metabolism pathways. Generally, the $^{13}$C fluxes were more accurate as there are more constraints from $^{13}$C labeled metabolite information. On the whole, the high consistency between the simulated fluxes and the calculated $^{13}$C fluxes validated the good prediction performance of iRY1243. The PPP pathway, EMP pathway, and TCA cycle flux distribution can be well predicted (see Additional file 5).

Prediction of essential genes

Based on the relationship of genes, proteins, and reactions in the genome-scale metabolic network model, the relationship between genotype and phenotype can be predicted (Thiele and Palsson 2010). The singleGeneDeletion function in Cobra Toolbox v2.0 can be used to...
predict the essential genes, partially essential genes, and non-essential genes (Schellenberger et al. 2011). On the synthetic medium, 123 essential genes were found, which were related to cofactor metabolism, TCA cycle, lipids biosynthesis, etc. (Fig. 6). At the same time, 169 partially essential genes were obtained (Additional file 6), which may be vital for subsequent in silico strain design (Pan and Hua 2012), considering the fact that the yield of target metabolite production may increase with the slight slowdown of growth.

To further evaluate the potential of the new *P. pastoris* GSMM in rational strain design, data of some engineered *P. pastoris* (the specific growth rate and *S*-adenosyl-<i>L</i>-methionine production) were collected from our lab and previously published studies. Then iRY1243 was used to simulate the effects of the corresponding gene insertion, deletion, or overexpression. The high consistency between the experimental and simulated results (see Additional file 7) indicated that iRY1243 can be used to design a more efficient *P. pastoris* for production of *S*-adenosyl-<i>L</i>-methionine or other heterogenous proteins.

### Table 3 Evaluation of growth capabilities of *P. pastoris* on different carbon sources (+ for growth and − for non-growth)

| Carbon sources | In vivo | iMT1026 | iRY1243 |
|----------------|---------|---------|---------|
| Xylose         | +       | +       | +       |
| Glucose        | +       | +       | +       |
| Glutamate      | +       | +       | +       |
| Fumarate       | +       | +       | +       |
| Glycogen       | +       | +       | +       |
| Mannose        | +       | +       | +       |
| Citrate        | +       | +       | +       |
| α-xylene       | +       | +       | +       |
| Acetate        | +       | +       | +       |
| Alanine        | +       | +       | +       |
| Mannitol       | +       | +       | +       |
| Sorbitol       | +       | +       | +       |
| Maltose        | +       | −       | +       |
| Glycerol       | +       | +       | +       |
| Methanol       | +       | +       | +       |
| Fructose       | +       | +       | +       |
| Maltotriose    | +       | +       | +       |
| Trehalose      | +       | +       | +       |
| Ethanol        | +       | +       | +       |
| L-Aspartate    | +       | +       | +       |
| α-Galactose    | +       | −       | +       |
| L-arabinose    | +       | −       | +       |
| Starch         | +       | −       | +       |
| L-rhamnose     | +       | +       | +       |
| Phenylalanine  | +       | +       | +       |
| Pyruvate       | +       | +       | +       |
| Ribose         | +       | +       | +       |
| Succinate      | +       | +       | +       |
| Xyitol         | +       | +       | +       |
| α-Ketoglutarate| +       | +       | +       |

### Table 4 Evaluation of growth capabilities of *P. pastoris* on different nitrogen sources (+ for growth and − for non-growth)

| Nitrogen sources | In vivo | iMT1026 | iRY1243 |
|------------------|---------|---------|---------|
| Alanine          | +       | +       | +       |
| Ammonia          | +       | +       | +       |
| Phenylalanine    | +       | −       | +       |
| Leucine          | +       | −       | +       |
| Arginine         | +       | +       | +       |
| Aspartate        | +       | +       | +       |
| Glutamate        | +       | +       | +       |
| Asparagine       | +       | +       | +       |
| Citrulline       | +       | −       | +       |
| 4-Aminobutanoate | +       | +       | +       |
| Glutamine        | +       | +       | +       |
| Glycine          | +       | +       | +       |
| Ornithine        | +       | +       | +       |
| Proline          | +       | +       | +       |
| Serine           | +       | +       | +       |
| Threonine        | +       | +       | +       |
| Tryptophan       | +       | +       | +       |
| Tyrosine         | +       | +       | +       |
| Urea             | +       | +       | +       |
| Xanthin          | +       | +       | +       |
| Putrescine       | +       | +       | +       |
Conclusion

*Pichia pastoris* GSMM was systematically upgraded based on the latest gene annotations and literatures. The resulting *P. pastoris* GSMM, iRY1243, accounts for 1243 ORFs and contains 2407 reactions and 1094 metabolites. The RNA-Seq data confirmed that the existence of most reactions in iRY1243 was reliable. iRY1243 showed good performance when predicting the cell growth physiological parameters, the growth capability on the reported carbon and nitrogen sources, the metabolic flux distribution with labeled glucose as the sole carbon source, the essential and partially essential genes, and the effects of gene deletion or overexpression on cell growth and S-adenosyl-l-methionine production. The significant improvement in the metabolic coverage and prediction ability make iRY1243 a potential platform for further systematic investigation of *P. pastoris* metabolism.

**Fig. 4** Predicted and measured $\mu$, $q_{o2}$, $q_{co2}$, and RQ for chemostat cultivation of *P. pastoris* G1HL producing $\beta$-galactosidase. Blue represents the GSMM simulated results, and red represents the experimental data.

**Fig. 5** Consistent changes in fluxes can be predicted using iRY1243 compared with $^{13}$C metabolic fluxes. The growth physiological parameters used for pFBA calculation, as well as the calculated $^{13}$C fluxes of *P. pastoris* G1HL, were described in previous research (Nie et al. 2014).
Fig. 6 Distribution of essential genes in metabolic subsystems of the *P. pastoris* GSMM

**Additional files**

- **Additional file 1.** Detailed information of iRY1243.
- **Additional file 2.** Biomass composition.
- **Additional file 3.** Non-growth-associated ATP maintenance (NGAM) requirement.
- **Additional file 4.** Detailed comparison of *Pichia pastoris* GSMMs prediction performance.
- **Additional file 5.** Comparison between $^{13}$C fluxes and simulated fluxes by iRY1078.
- **Additional file 6.** Essential genes and partially essential genes predicted by iRY1078.
- **Additional file 7.** Measured and simulated influences of gene deletion or overexpression on the growth and S-adenosyl-L-methionine production by *P. pastoris*.

**Abbreviations**

*P. pastoris*; GSMM: genome-scale metabolic network model; GPRs: gene–protein–reaction associations; NGAM: non-growth-associated maintenance; GAM: growth-associated ATP maintenance; D: dilution ratio; DO: dissolved oxygen; DCW: dry cell weight; FBA: constraint-based flux-balanced analysis; $\mu$: specific growth rate.

**Authors’ contributions**

RY conducted the experiments and wrote the manuscript. MH provided advice on the experiment design and data analysis. All authors read and approved the final manuscript.

**Acknowledgements**

This work was financially supported by the National Basic Research Program of China (No. 2013CB733600), the National Natural Science Foundation of China (No. 201276081), and the National Key Technology Support Program of China (No. 2011BAF02B05).

**Competing interests**

The authors declare that they have no competing interests.

**Availability of data and materials**

The dataset supporting the conclusions of this article is available.

**Funding**

National Basic Research Program of China (No. 2013CB733600). National Natural Science Foundation of China (No. 201276081). National Key Technology Support Program of China (No. 2011BAF02B05).

**Publisher’s Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Received: 28 March 2017 Revised: 17 April 2017 Accepted: 2 May 2017 Published online: 09 May 2017

References
Ahmad M, Hirz M, Pichler H, Schwab H (2014) Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol 98(12):5301–5317
Angione C, Conway M, Lio P (2016) Multiplex methods provide effective integration of multi-omic data in genome-scale models. BMC Bioinform 17:83
Baumann K, Maurer M, Dragosits M, Cos Q, Ferring P, Mattanovich D (2008) Hypoxia fed-batch cultivation of Pichia pastoris increases specific and volumetric productivity of recombinant proteins. Biotechnol Bioeng 100(1):177–183
Bhataay A, Schmidt-Dannert C, Lee PC (2009) Metabolic engineering of Pichia pastoris X-33 for lycopene production. Process Biochem 44(10):1095–1102
Bordbar A, Monk JM, King ZA, Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet 15(2):107–120
Camicer M, Baumann K, Toplitz I, Sanchez-Ferrando F, Mattanovich D, Ferring P, Albilj B (2009) Macromolecular and elemental composition analysis and extracellular metabolite balances of Pichia pastoris growing at different oxygen levels. Microb Cell Fact 8:65
Camicer M, ten Pierick A, van Dam J, Heijnen JJ, Albiol J, van Gulik W, Ferrer P (2012) Quantitative metabolomics analysis of amino acid metabolism in recombinant Pichia pastoris under different oxygen availability conditions. Microb Cell Fact 11:83
Caspetta L, Shoaie S, Agren R, Nookaew I, Nielsen J (2012) Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. BMC Syst Biol 6:24
Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast Pichia pastoris. FEMS Microbiol Rev 24(1):45–66
Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee H, Lee DY (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microb Cell Fact 9:50
De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, de Peer YV, Callewaert N (2009) Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 27(6):551–566
Famili I, Forster J, Nielsen J, Palsson BO (2003) Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci USA 100(23):13134–13139
Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using as optimization function constraints. Biotechnol J 5(7):705–715
Fumagalli GS115 using glucose as sole carbon source via C-13 metabolic flux analysis. J Biotechnol 187:124–134
Kubler A, Schneider J, Thallinger GG, Andler I, Wibberg D, Hajek T, Jaenicke S, Brinkkorf K, Goessmann A, Szczepanski R, Puhlcr A, Schwab H, Glieder A, Pichler H (2011) High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 154(4):312–320
Maahimo H, Fiaux J, Kakar ZR, Bailey JE, Sauer U, Szyperski T (2001) Central carbon metabolism of Saccharomyces cerevisiae explored by biosynthetic fractional C-13 labeling of common amino acids. Eur J Biochem 268(8):2464–2479
Menendez J, Delgado J, Gancedo C (1998) Isolation of the Pichia pastoris PYC1 gene encoding pyruvate carboxylase and identification of a suppressor of the pyc phenotype. Yeast 14(7):647–654
Morozkina EV, Vavilova EA, Zatsepin SS, Klyachko EV, Yagudin TA, Chulkina AM, Dudich IV, Semenkova LN, Churilova IV, Benevolensky SV (2016) Engineeering of a system for the production of mutant human alpha-fetoprotein in the methylotrophic yeast Pichia pastoris. Appl Biochem Microbiol 52(2):170–175
Nie YS, Huang MZ, Lu J, Qian JC, Lin WL, Chu J, Zhuang YP, Zhang SL (2014) Impacts of high beta-galactosidase expression on central metabolism of recombinant Pichia pastoris G515S using glucose as sole carbon source via C-13 metabolic flux analysis. J Biotechnol 187:124–134
Nicoon J, Steiger MG, Pfeffer M, Sohn SB, Kim TY, Maurer M, Ruismayer H, Pfurfl S, Ask M, Haberhauer-Troyer C, Olmertor K, Hans S, Koellensperger G, Gassser B, Lee SY, Mattanovich D (2014) Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metab Eng 24:129–138
Orth JD, Theile I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol 28(3):245–248
Pan PC, Hua Q (2012) Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, Yarrowia lipolytica. PLoS ONE 7(12):e51355
Saha R, Chowdhury A, Maranas CD (2014) Recent advances in the reconstruction of metabolic models and integration of omics data. Curr Opin Biotechnol 29:39–45
Sanchez CEG, Garcia CAV, Saez RGT (2012) Predictive potential of flux balance analysis of Saccharomyces cerevisiae using as optimization function combinations of cell compartmental objectives. PLoS ONE 7(8):e43006
Schellenberger J, Que R, Fleming RMT, Theile I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduek DR, Palsson BO (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6(9):1290–1307
Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferring P, Mattanovich D, Lee SY (2010) Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. Biotechnol J 5(7):705–715
Sola A, Maahimo H, Ylonen K, Ferrer P, Szyperski T (2004) Amino acid bio-synthesis and metabolic flux profiling of Pichia pastoris. Eur J Biochem 271(12):2462–2470
Theile I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 5(1):93–121
Tomas-Gamisans M, Ferrer P, Albiol J (2016) Integration and validation of the genome-scale metabolic models of Pichia pastoris: a comprehensive update of protein glycosylation pathways, lipid and energy metabolism. PLoS ONE 11(1):e0148031
Verduyn C, Stouthamer AH, Scheffers WA, van Dijken JP (1991) A theoretical evaluation of growth yields of yeasts. Antonie van Leeuwenhoek 59(1):49–63
Zhang X, Liu J, Yu X, Yang F, Li Y, Li Z, Liu Y, Ma L (2015) High-level expression of human arginase I in Pichia pastoris and its immobilization on chitosan to produce L-ornithine. BMC Biotechnol 15:66