

RESEARCH ARTICLE

Open Access

Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and *in silico* evaluation of their potentials

Luis Caspeta[†], Saeed Shoaie[†], Rasmus Agren, Intawat Nookaew and Jens Nielsen^{*}

Abstract

Background: *Pichia stipitis* and *Pichia pastoris* have long been investigated due to their native abilities to metabolize every sugar from lignocellulose and to modulate methanol consumption, respectively. The latter has been driving the production of several recombinant proteins. As a result, significant advances in their biochemical knowledge, as well as in genetic engineering and fermentation methods have been generated. The release of their genome sequences has allowed systems level research.

Results: In this work, genome-scale metabolic models (GEMs) of *P. stipitis* (iSS884) and *P. pastoris* (iLC915) were reconstructed. iSS884 includes 1332 reactions, 922 metabolites, and 4 compartments. iLC915 contains 1423 reactions, 899 metabolites, and 7 compartments. Compared with the previous GEMs of *P. pastoris*, PpaMBEL1254 and iPP668, iLC915 contains more genes and metabolic functions, as well as improved predictive capabilities. Simulations of physiological responses for the growth of both yeasts on selected carbon sources using iSS884 and iLC915 closely reproduced the experimental data. Additionally, the iSS884 model was used to predict ethanol production from xylose at different oxygen uptake rates. Simulations with iLC915 closely reproduced the effect of oxygen uptake rate on physiological states of *P. pastoris* expressing a recombinant protein. The potential of *P. stipitis* for the conversion of xylose and glucose into ethanol using reactors in series, and of *P. pastoris* to produce recombinant proteins using mixtures of methanol and glycerol or sorbitol are also discussed.

Conclusions: In conclusion the first GEM of *P. stipitis* (iSS884) was reconstructed and validated. The expanded version of the *P. pastoris* GEM, iLC915, is more complete and has improved capabilities over the existing models. Both GEMs are useful frameworks to explore the versatility of these yeasts and to capitalize on their biotechnological potentials.

Background

Pichia stipitis possesses the highest native ability of any yeast to metabolize xylose [1], and is therefore a key candidate for ethanol production from biomass, as well as for engineering xylose metabolism in *Saccharomyces cerevisiae* [2-4]. Xylose consumption requires two additional reactions that are catalyzed by xylose reductase (XYL1, Xyl1p) and xylitol dehydrogenase (XYL2, Xyl2p). Cofactor requirements for these reactions affect the oxygen demand of cells [4]. Xyl1p has a higher affinity for NADPH whereas Xyl2p prefers NAD⁺, hence the formed NADH

cannot be properly recycled at oxygen limited conditions [4]. Therefore, the efficient conversion of xylose to ethanol or biomass occurs under defined aerobic conditions. *P. stipitis* produces ethanol at yields close to the maximum at low oxygen transfer rates (~2 mM h⁻¹), but xylose uptake rate is only half of the maximum attained under fully respiratory conditions [5]. Attempts to increase the rate of xylose consumption under oxygen-limited conditions have been only partially successful, since the engineered cells were unable to use xylose alone or had reduced ethanol production [6,7]. Due to its advantages in the production of ethanol under anaerobic conditions, while resisting very high concentrations of ethanol, there has been much focus on converting *S. cerevisiae* into a xylose fermenter. However, the expression of target genes from *P. stipitis* into

* Correspondence: nielsenj@chalmers.se

† Contributed equally

Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

S. cerevisiae has not been quite successful due to the difficulties in regulating redox balances during xylose consumption and ethanol conversion [3,4]. Furthermore, the production of ethanol is differently regulated in the two yeasts. Whereas *S. cerevisiae* can produce ethanol via aerobic fermentation (the Crabtree effect) or anaerobically, especially in the presence of high concentrations of glucose [8], *P. stipitis* does it mainly in response to oxygen limitations (Pasteur effect) [5,8,9].

Pichia pastoris belongs to a small group of microorganisms capable of catabolizing methanol and fatty acids (e.g. oleic acid) as the sole carbon and energy source [10,11]. It can up-regulate the expression of crucial genes (e.g. alcohol oxidase, *AOX*; and the multifunctional beta-oxidation protein, *FOX2*), and multiply peroxisomes when it is growing on such carbon sources [11,12]. *P. pastoris* has therefore been used extensively for the production of recombinant proteins using the *AOX* gene promoter, as well as a model for studying peroxisome proliferation [11,13]. Consequently, a number of genetic tools and cultivation methods have been developed [13]. Cultivation techniques include fed-batch culture with glycerol, followed by induction with methanol alone, or in combination with glycerol or sorbitol which has shown to be useful for increasing the production of recombinant proteins [14-16]. The success of engineering the glycosylation pathway of *P. pastoris* to produce sialylated glycoproteins has increased the expectations for its use to produce pharmaceutically relevant proteins [17], as well as to transfer the glycosylation technology into *S. cerevisiae* [18].

With the availability of the complete genome sequences for *P. stipitis* and *P. pastoris* [19,20], there is the opportunity to study, at the system level, the native potentials of these yeasts. Genome-scale metabolic models (GEMs) can be used to support such a task since they can be used to assess metabolic capabilities of cells, to analyze metabolites connectivity and pathway redundancy, or for comparing metabolic capabilities between closely related species [21]. GEMs can also be used to predict genotypic-phenotypic relationships [22], and for the identification of metabolic engineering targets [23]. Moreover, by incorporating 'omics' data and *in silico* methods, GEMs can act as scaffolds for the design of optimal metabolic fluxes [24-26], or to evaluate the correlation between gene expression and metabolic changes in response to environmental perturbations [22,27,28].

Thus far, two GEMs of *P. pastoris* have been published [29,30]. However, neither one incorporates and evaluates methanol metabolism and protein production together, which are among the most important features of this yeast. We also show that our model better predicts the growth phenotype on methanol alone or in combination with glycerol. In this manuscript we report an extensive GEM for *P. pastoris* (iLC915), together with the first GEM

for *P. stipitis* (iSS884). Both GEMs correctly represented available experimental phenotypes. The iSS884 model, for example, can predict xylose consumption and distribution into biomass and ethanol using different oxygen up-take rates. The iLC915 model can closely simulate the physiological differences of *P. pastoris* growing at different carbon sources alone (including oleic acid and methanol), as well as methanol mixed with glycerol or sorbitol. iLC915 also includes reactions for the production of a model recombinant protein, for which the production under different oxygen uptake rates compared well with experimental results. The use of such networks on the design of fermentation technologies is also discussed.

Methods

Reconstruction of the models

Reconstruction of the GEMs was comprised by a semi-automatic approach enriched with functional analysis and extensive manual curation based on available literature. Figure 1 depicts an overview of the process. The RAVEN Toolbox was used to generate draft models (manuscript under review). This approach uses a set of reference GEMs of closely related species and infers reactions by means of protein homology. In parallel to this method, the RAVEN Toolbox also uses the KEGG database to infer reactions that may be missing or incorrect in the template models. Draft models were constructed using the genome sequences of *P. stipitis* (CBS 6054) [20], and *P. pastoris* (GS 115) [19]. The iIN800 GEM of *S. cerevisiae* was used as reference framework because of its accurate annotation of fatty acid metabolism, and the extensive information about metabolites and genes, which allowed for a convenient comparison with the KEGG database.

BLASTp (ncbi-blast software ver. 2.2.24) was used to identify homologous proteins among the three yeast species. Protein homologs were identified based on stringent cutoff values (E -values $< 10^{-40}$), and on the score to sequence length ratio according to David et al. (2008) [31]. KEGG Ontology (KO) identifiers were also used to additionally infer reactions which could not be found in *S. cerevisiae* from the genome sequences of the two *Pichia* species following the RAVEN Toolbox pipeline. Finally, the metabolic network of *S. cerevisiae* iIN800 was used to map genes from *P. pastoris* and *P. stipitis* having homologs in *S. cerevisiae*.

Subcellular compartmentalization of reactions was determined using the F-LocA (Fully-connected Localization Assignment), which is part of the RAVEN Toolbox. F-LocA incorporates subcellular localization predictors (CELLO and WoLFPSORT) [32], together with a constraint on network connectivity. Reactions without associated genes were compartmentalized according to biochemical evidence when available. It is important to note that these automated approaches were only used as

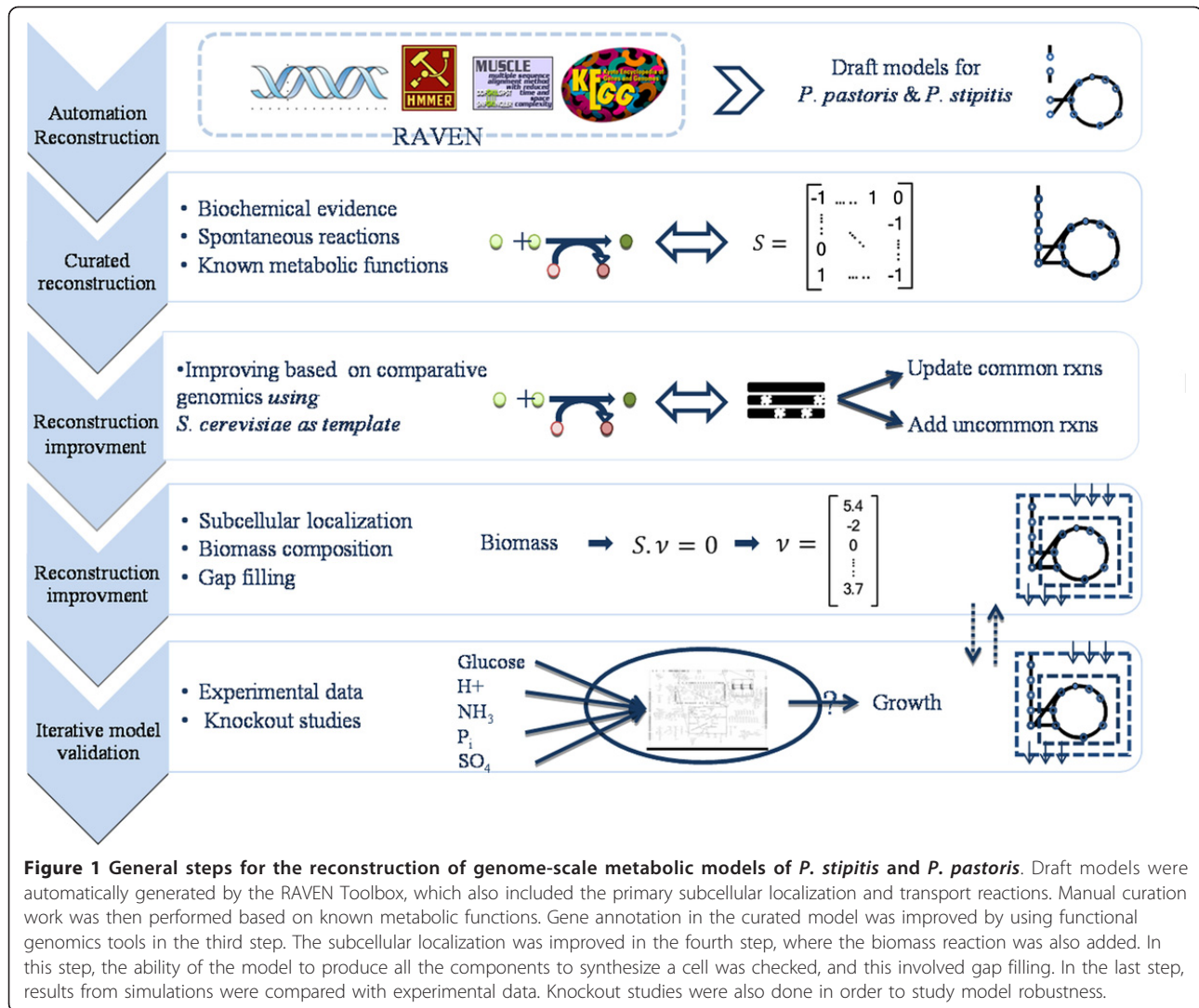


Figure 1 General steps for the reconstruction of genome-scale metabolic models of *P. stipitis* and *P. pastoris*. Draft models were automatically generated by the RAVEN Toolbox, which also included the primary subcellular localization and transport reactions. Manual curation work was then performed based on known metabolic functions. Gene annotation in the curated model was improved by using functional genomics tools in the third step. The subcellular localization was improved in the fourth step, where the biomass reaction was also added. In this step, the ability of the model to produce all the components to synthesize a cell was checked, and this involved gap filling. In the last step, results from simulations were compared with experimental data. Knockout studies were also done in order to study model robustness.

an aid in the reconstructions, and that biochemical and physiological evidence was always used to validate reaction localizations and gene associations. This was of particular importance in the peroxisomal metabolism where the predictive capability is lower due to the low quality of data from subcellular localization predictors (e.g. CELLO predicts that AOX is in the cytosol, but it is in the peroxisome). In cases where information about *P. pastoris* or *P. stipitis* was lacking, data from other closely related yeasts was used instead (e.g. *S. cerevisiae*, *Hansenula polymorpha*, *Candida tropicalis*, *C. shehateae*, and *C. boidinii* [33,34]). Both GEMs are available in the BioMet Toolbox [http://www.sysbio.se/BioMet/ - will be uploaded upon acceptance of paper].

Biomass reaction and energetic considerations

Biomass reactions were assembled from the macromolecular components (i.e. carbohydrates, proteins, lipids,

DNA and RNA). The contribution of each component to biomass (g gDW⁻¹), and the appropriate coefficients for every building block present in each macromolecule (mmol g⁻¹ of macromolecule) were calculated based on compositional analysis reports available for both yeast species [19,20,35-38]. These calculations are available in the Microsoft Excel model files.

Production of a model heterologous protein was only considered in the GEM of *P. pastoris* due to its use in recombinant protein production. This protein was the human monoclonal antibody 3H6 Fab fragment (FAB), since there are some experimental reports using it [39,40]. Accordingly, four reactions were included to polymerize nucleotides, ribonucleotides and amino acids separately, as well as for assembling them (Additional file 1). The amino acid composition of FAB was computed from the primary structure submitted to the National Center for Biotechnology Information (NCBI).

Since the nucleotide and ribonucleotide sequences have not been reported, these were obtained from the amino acid sequence using *P. pastoris* codon usage reported by De Shutter et al. (2009) [19].

There are two primary ways for a cell to use the generated ATP. The ATP required for biomass synthesis (i.e. precursor biosynthesis and polymerization), and that used for cell maintenance. Both are added to calculate the total amount of energy required for a cell to grow, as it is shown in the following equation:

$$r_{ATP} = Y_{XATP} \cdot \mu + m_{ATP} \quad (1)$$

In equation (1), r_{ATP} specifies the total amount of ATP being utilized, Y_{XATP} corresponds to the energy utilized for biomass synthesis, μ is the specific growth rate, and m_{ATP} is the ATP used for maintenance. The value of Y_{XATP} was calculated in two steps according to the procedure previously described [22,41]. First, the amount of ATP demanded for polymerization of cell macromolecules was computed according to Verduyn et al. (1991) [42]. The calculated value is then adjusted by incorporating a scaling factor to fit the predicted slope of the dilution rate vs. glucose-uptake-rate ($\mu/r_{glc} = Y_{sx}$ with experimental data [41]).

Constraints-based flux analysis and simulations

Flux balance analysis was used in the simulations with the reconstructed genome scale metabolic models [43,44]. This approach requires a pseudo-steady state to balance metabolite concentrations (X_i). The relation between metabolites' coefficients in the stoichiometric matrix (S_{ij}) and fluxes through reactions (v_j) is represented in equation (2).

$$S_{ij} \cdot V_j = \frac{dX_i}{dt} = 0 \quad (2)$$

After each simulation an additional flux variability analysis was performed to ensure that the reported fluxes represent the only optimal solution for the resulting objective function value.

In silico prediction of carbon utilization

A general overview of the physiological states through utilization of different carbon sources was performed by feeding 1 C-mol $gDW^{-1} h^{-1}$ of each carbon source (e.g. glucose, methanol, xylose, and oleic acid among others), and maximizing for cell growth. Experimental values for specific consumptions were fixed when it was required to predict experimental data. A minimal media was used for all simulations, i.e. reactions for consuming ammonia, sulfate and phosphate remained open during simulations. In simulations with dual carbon sources (e.g. methanol-glycerol or methanol-sorbitol), the upper boundaries of

their respective transport reactions were constrained to the experimental fluxes.

In silico production of a recombinant protein

FAB production was evaluated using glucose as carbon source and ammonia as nitrogen source. The objective of these simulations was to assess the effect of oxygen uptake rate and FAB production on the physiological state of cells. This state can depend on the availability of glucose and oxygen, as well as on the metabolic over-load exerted by the transcription, translation and post-translational processing of the recombinant protein. The amount of FAB that cells should synthesize was fixed, since the production of a recombinant protein is subjected to a number of posttranslational processing and regulatory functions, not included in the GEM. The fluxes through the reactions to produce biomass, ethanol and CO_2 , as well as the consumption of glucose, were then quantified and compared against experimental data. Another set of simulations were carried out to evaluate the maximum production of FAB using different carbon sources separately or in combination. In this case, the reaction exporting FAB was used as the objective function and the maximum production was attained.

In silico prediction of the effect of oxygen uptake rate on xylose and glucose fermentation

In silico evaluation of the effect of oxygen uptake rate on the production of ethanol and biomass using glucose or xylose was evaluated with the GEM of *P. stipitis*. The oxygen uptake rate (r_{O_2}) was used as the limiting parameter in simulated continuous cultures with xylose as sole carbon source (note that r_{O_2} was fixed as the GEM cannot predict that value). Experimental values of r_{O_2} reported by Skoog et al. (1986) [5] were used to fix the upper limits. Simulations were then carried out using the specific production rate of biomass as the objective function. With the results obtained from this evaluation, other simulations were also carried out using a mixture of glucose and xylose at concentrations found in hydrolysates of lignocellulose. These computations were performed in three steps representing the use of continuous reactors in series reported by Grootjen et al. (1991) [45]. In the first reactor a mixture of 40 g/L of glucose and 10 g/L of xylose was fed. Glucose uptake rate was calculated from experimental data. At this step, the production of biomass was favored by using a high r_{O_2} value. The stream from this reactor was used to feed the second one. Here, the glucose was completely depleted and xylose started to be consumed at a rate calculated from the experimental data. In this simulation the oxygen uptake rate was decreased to favor the production of ethanol. This strategy was also used in the third reactor where xylose was mainly converted to ethanol.

In silico analysis of reaction essentiality

Reactions deletion in GEMs can be applied by constraining the flux through a single reaction, or to a set of reactions associated with one gene, to zero. Flux balance analysis is then performed to predict the maximum growth rate of the *in silico* mutant strain. The preceding procedure was repeated for all reactions in the network. Percentages of the mutants growth (*Pmw*) with respect to the wild type were calculated using fluxes through the biomass reaction ($Pmw = v_{X-mutant}/v_{X-wild} \cdot 100$), and three different labels were assigned. Reactions were classified as essential (E), non-essential (NE), and partially essential (PE). Essential reactions are those where their deletion caused a non-growing phenotype in the *in silico* mutant, getting a *Pmw* of zero. Non-essential reactions have no effect on the growth rate upon deletion. Partially-essential reactions had *Pmw* values between of 0.01-to-0.99, hence a diminished growth rate was seen upon their deletion.

Results

Characteristics of the genome-scale metabolic models

GEMs for *P. pastoris*, named iLC915, and *P. stipitis*, named iSS884, are provided in Microsoft Excel, SBML and BioOpt formats and are available in the BioMet Toolbox <http://www.sysbio.se/BioMet>.

Table 1 gives a summary of the GEMs' main features and how they compare with other models, including the fully compartmentalized consensus model iMM904 of *S. cerevisiae* [46]. The iSS884 model includes 1332 reactions and 922 unique metabolites compartmentalized into four subcellular locations. The number of annotated genes (884) constitutes 15.2% of the total open reading frames (ORFs) found in the genome. The iLC915 contains 1423

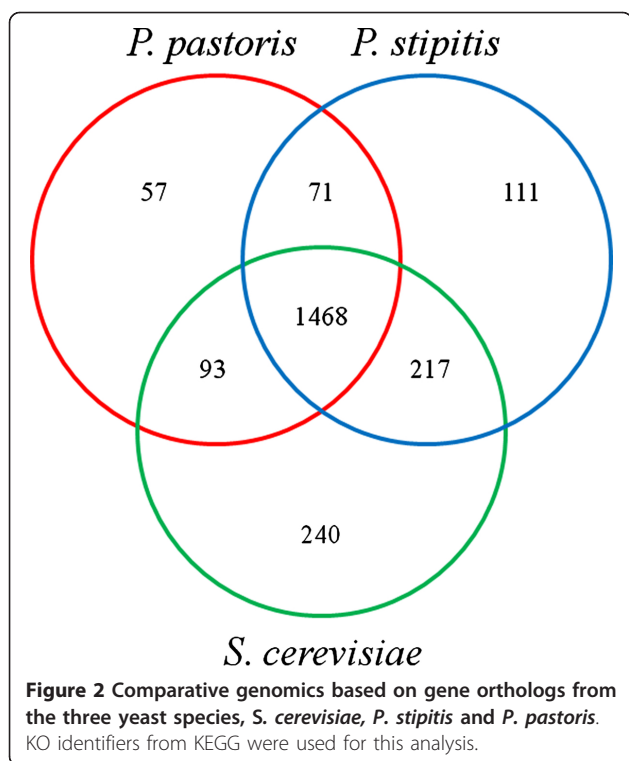
reactions and 899 unique metabolites compartmentalized into seven subcellular locations. In total, 915 ORFs were annotated. Comparing with the previous models (PpaMBEL1254 [47] and iPP668 [48]), iLC915 contains more annotated genes (915 compared with 540 and 668, respectively). The number of reactions distributed in the compartments is comparable with the previous GEMs of *P. pastoris* and the iMM904.

Gene orthology analysis using the KEGG orthology (KO) system was performed to improve GEMs reconstruction (Additional file 2). Such an analysis was also valuable to compare the three yeast species (*P. stipitis*, *P. pastoris*, and *S. cerevisiae*) by their GEMs. In total, 1468 genes orthologs are common (Figure 2), of which 522 were annotated in the GEMs (i.e. central, amino acids and nucleotides metabolic reactions). The remaining 946 are associated with cellular processes, as well as genetic expression and regulation, which are not included in the GEMs. In comparison to iIN800 and iSS884, the iLC915 model contains around one hundred more reactions mainly used for posttranslational protein processing and degradation (these are mainly localized into the vacuole, ER and Golgi). 71 genes orthologs are shared solely between *P. pastoris* and *P. stipitis*, of which 16 are associated to the NADH dehydrogenases family (Complex I, also called NADH:ubiquinone oxidoreductase) neither found in *S. cerevisiae* [49], nor in the previous GEMs of *P. pastoris* [47,48]. The substitute oxidases PAS_chr3_0408 and AOX_PICST from *P. pastoris* and *P. stipitis* were not detected during this analysis. These genes were also not included in the previous models for *P. pastoris* [47,48]. However, they have been shown to have an important role in the respiratory metabolism [50,51]. Most of genes orthologs only found for

Table 1 Characteristics of the reconstructed metabolic models iLC915 and iSS884 of *Pichia pastoris* and *stipitis*, in comparison with the *S. cerevisiae* model iIN800 and iMM904

Characteristics	<i>P. stipitis</i>		<i>P. pastoris</i>		<i>S. cerevisiae</i>	
	iSS884	iLC915	PpaMBEL1254	iPP668	iMM904	iIN800
Reactions	1332	1423	1202	1354	1312	1446
Cytoplasm	824	790	604	623	635	906
Mitochondria	207	205	155	163	180	161
Peroxisome	60	64	66	66	63	
Vacuole		12	6	3	3	
ER		34	7	15	17	
Golgi		4	8	4	2	
Nucleus			17	16	18	
Extracellular			11	12	7	
Transport	239	314	328	452	387	379
Metabolites	920	899	1147	1177	1168	1013
Gene coverage (%)	15.1	17.2	9.9	12.6	13.7	12.11
Coding genes	5841	5313	5313	5313	6607	6607

Characteristics of the previous GEMs of *P. pastoris* are also included to compare with iLC915



P. pastoris and *S. cerevisiae* were not included in the GEMs since they encode functions of cellular and genetic information processing. A higher number of shared genes orthologs were found among *P. stipitis* and *S. cerevisiae* (217). Genes, whose proteins are required to metabolize xylose and arabinose, were annotated in iSS884 and iIN800. There are also some hydrogenases, reductases and phosphatases without gene association in *P. pastoris* (Additional file 2). Genes encoding alcohol oxidases (AOX) and formaldehyde dehydrogenase (FALD), required for methanol metabolism, were found neither in *P. stipitis* nor in *S. cerevisiae*. However, these yeasts have genes that code for the remaining four genes of methanol catabolism (S-formylglutathione hydrolase, formate dehydrogenase, catalase, and dihydroxyacetone synthase). Additional file 2 is available for further information.

***In silico* prediction of reaction essentiality**

In silico prediction of reaction essentiality was performed. Results have been distributed in three categories: essential, partially-essential, and non-essential (see materials and methods section). In total, 1417, 1328, and 1361 unique reactions were deleted individually from iLC915, iSS884 and iIN800 GEMs, respectively. Out of these deleted reactions, 173 in iLC915, 155 in iSS884 and 193 in iIN800 were found to be essential for growth in minimal media with glucose.

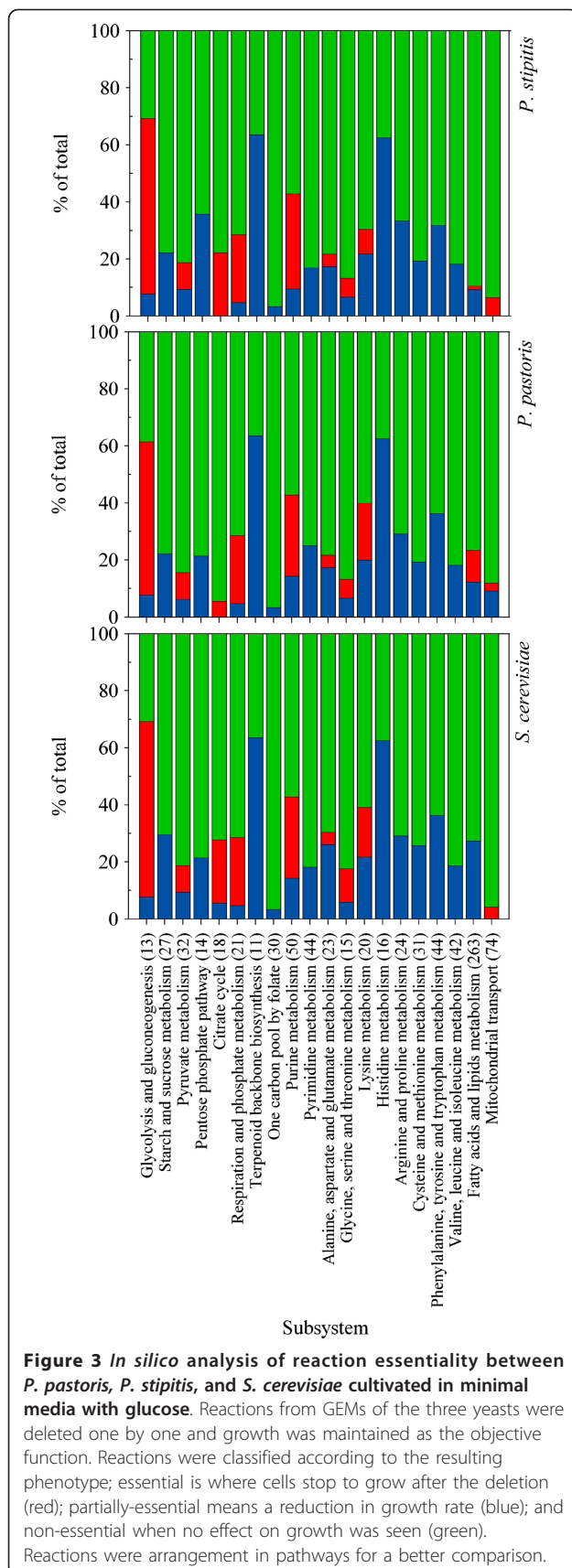
Reactions which only caused growth retardation (partially-essential) accounted for 85 in iLC915 and 58 in iSS884 and iIN800.

Deleted reactions for the three models were grouped into different metabolic pathways (Figure 3 and Additional file 3). Reactions deleted in amino acid synthesis grouped most of the essential genes, with the highest score for essential-reactions in the metabolism of histidine. Biosynthesis of the backbone for terpenoids accumulated 6 lethal reactions comprising 60% of the reactions included in this pathway. The highest amount of partially-essential reactions was distributed in the central carbon metabolism. The iIN800 model of *S. cerevisiae* accumulated a higher number of lethal phenotypes in reactions located in the lipids network (25%, compared to 10% in the other yeasts), which is likely due to the fact that the iIN800 comprises more details on lipids composition in the biomass equation [52].

***In silico* evaluation of carbon utilization**

In silico evaluation of carbon assimilation using the iLC915 and iSS884 GEMs was carried out in simulated chemostats with minimal media. With iSS884, glycerol was better assimilated (Figure 4), followed by glucose and mannose. Xylose, arabinose and cellobiose showed the lowest conversion yields. The maximum conversion of carbon to biomass in the *P. pastoris* model iLC915 was computed for glycerol and sorbitol, followed by glucose and trehalose. Lower values were obtained with oleic acid, alanine and methanol. Simulations also reproduced the higher oxygen requirements to oxidize methanol, alanine and oleic acid, which is the first step for their metabolism.

Energy requirements for biomass synthesis strongly depend on the nature of the carbon source [53]. The total amount of ATP required to produce one gram of biomass (Y_{xATP}) and the efficiency of aerobic respiration measured as P/O ratio were calculated for each carbon source. With glucose, glycerol, sorbitol, or trehalose, the computed values for Y_{xATP} using the iLC915 GEM were 70.5, 76.5, 73.1, and 74.2 mmolATP gDW⁻¹, respectively. Even when the biomass yield with alanine was low compared with the other carbon sources, a comparable result of the Y_{xATP} was calculated (72.3 mmolATP gDW⁻¹). The highest energy requirements for growth were estimated with methanol and oleic acid (152.5 and 95.8 mmolATP gDW⁻¹). Mixtures of glycerol with methanol at decreasing ratios (i.e. 4.33, 1.44, and 0.70 grams of glycerol per gram of methanol) showed increasing Y_{xATP} values (80, 87, and 94 mmolATP gDW⁻¹, respectively), and reducing biomass yields (Y_{sx}) accordingly (0.51, 0.48, and 0.45 gDW g⁻¹). With the iSS884 model of *P. stipitis*, ATP required for biomass formation was similar for some carbon sources (glycerol,

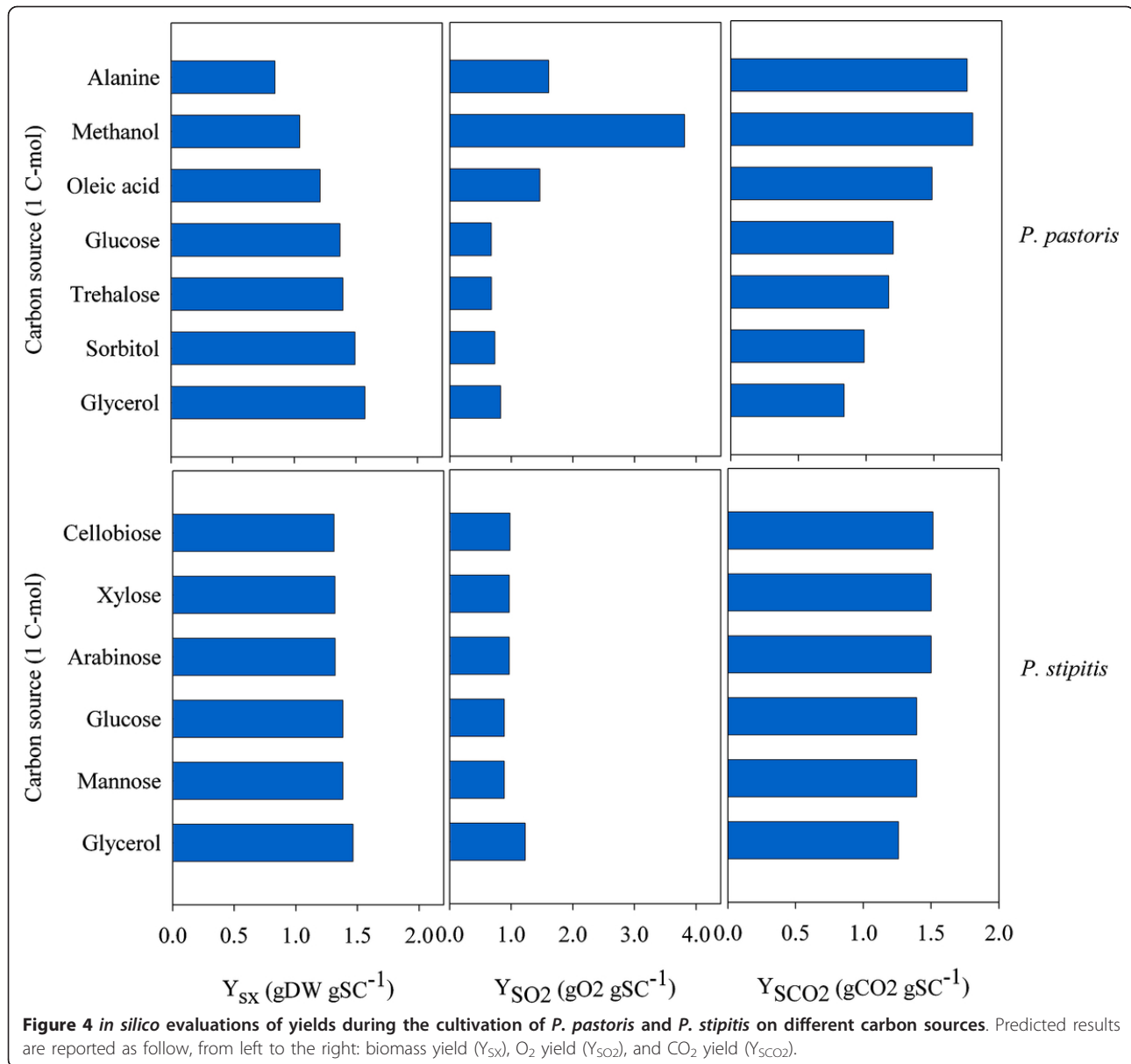


95.5 mmolATP gDW⁻¹; glucose and mannose, 82.2 mmolATP gDW⁻¹). Xylose, arabinose, and cellobiose assimilation needed 87.5, 87.4 and 86.0 mmolATP gDW⁻¹, respectively.

Comparison between *in silico* phenotypic predictions and experimental data

Only a few studies on continuous culturing of *P. stipitis* to investigate its physiology have been done [5,54]. Particularly, the oxygen transfer rate is a crucial factor for xylose metabolism that, at defined levels, can maximize the productivity and yield of ethanol. The role of oxygen in xylose fermentation can be explained by the fact that cells have to maintain redox balance, xylose transportation, cell growth or keep mitochondrial function [55]. The iSS884 model was used to evaluate the metabolism of xylose by constraining it to experimental oxygen transfer rates. Results from *in silico* predictions of biomass yields (Y_{SX}), dilution rates (μ), ethanol yields ($Y_{S\text{EtOH}}$), CO₂ specific productivities (r_{CO_2}) and ethanol specific productivities (r_{EtOH}) were compared with experimental data. Figure 5 shows that the GEM predicts the correlation of oxygen transfer rate with metabolism, which passed from fermentative to respiratory. These results are in agreement with experimental data [5]. Furthermore, *in silico* simulations predicted the inability of *P. stipitis* to grow in anaerobic conditions, as well as the small amount of ethanol that cells produce *in-vivo*. Xylitol production was not observed neither in the experimental results [5], nor in the *in silico* evaluations (indicating that the regeneration of NAD⁺ was not limiting).

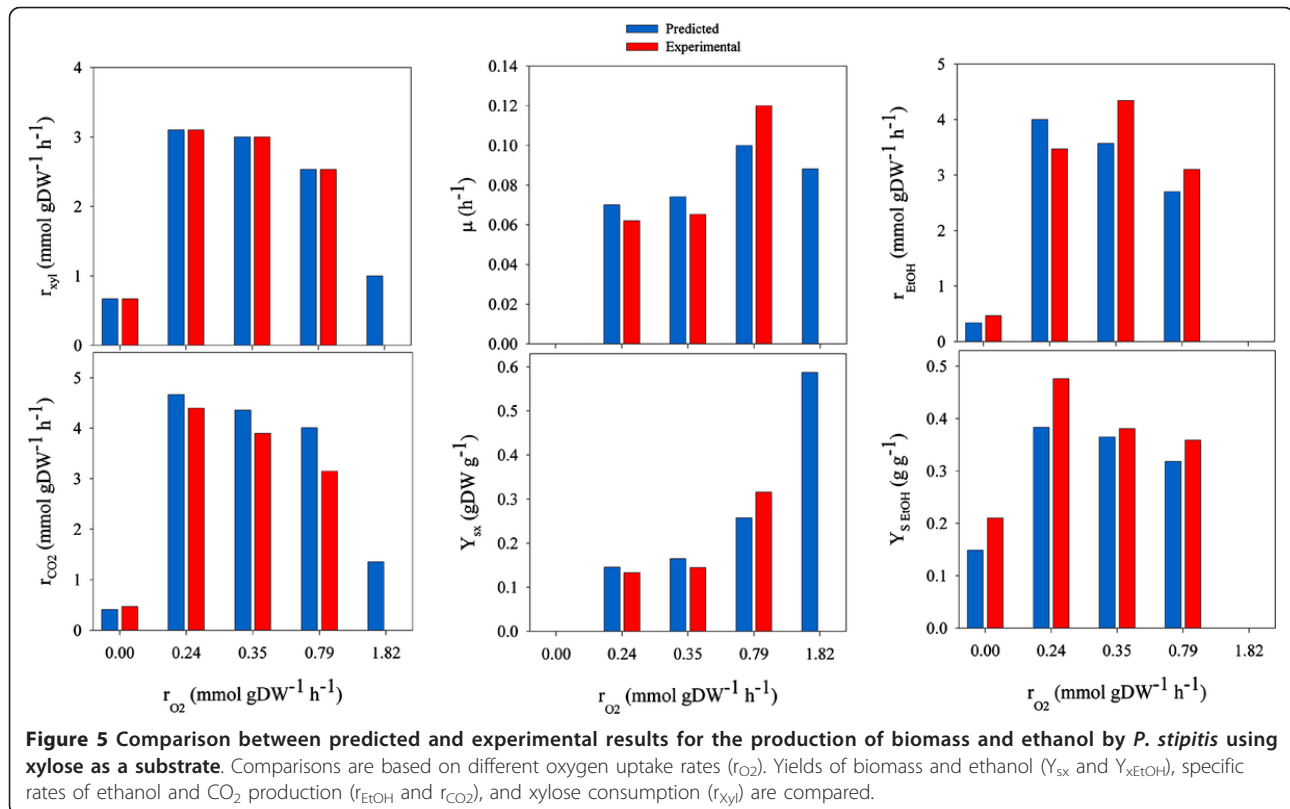
In silico physiological predictions with the iLC915 GEM of *P. pastoris* using glucose, methanol, glycerol and a mixture of methanol with glycerol at various dilution rates were also carried out. Predicted values for specific growth rates (μ), oxygen consumptions (r_{O_2}), and CO₂ productions (r_{CO_2}) are shown in the Figure 6. Comparing with PpaMBEL1254, the iLC915 had a similar capacity to predict *P. pastoris* phenotype on glucose. Such predictions were less accurate when using the iPP668 model, which computed higher μ values. This was also the case when the models were used to simulate glycerol metabolism. Results of methanol utilization in the three GEMs were different. While iPP668 predicted higher values of μ , the PpaMBEL1254 computed lower ones (0.0135 and 0.005 h⁻¹). This was reflected in the amount of CO₂ produced by both models (0.91 and 0.392 mmol/gDW⁻¹). Computations with iLC915 provided values of 0.0125 h⁻¹ and 0.48 mmol/gDW⁻¹ for μ and r_{CO_2} . Those values were closer to *P. pastoris* phenotype on this carbon source, as it is shown by comparing with experimental data (Figure 6). Furthermore, the better prediction of *P. pastoris*' physiology with iLC915 resulted in improved estimates when methanol and glycerol were used simultaneously.



***In silico* simulation of a recombinant protein production using the *P. pastoris* model iLC915**

To evaluate the ability of iLC915 for simulating recombinant protein production, we looked at the production of the human monoclonal antibody 3H6 Fab fragment (FAB), which has been experimentally assessed [39,40]. Simulations were carried out in simulated chemostats with glucose at different oxygen uptake rates (Table 2). It should be noted that simulations just included the metabolic adjustments of the biochemical responses according to the metabolic capabilities represented by the GEM constrained with experimental data. Cellular processes like molecular regulation of stress responses and protein quality control are not included.

Table 2 shows a comparison between predicted and experimental data. With constant values of Y_{xFAB} and r_{O_2} , model predictions are very close to the experimental values. The maximum biomass yield was obtained experimentally and *in silico* under fully aerobic conditions (case 1). By decreasing the percentage of oxygen in the inlet air to 11 and 8%, Baumann et al. (2010) [40] observed a 2-fold higher specific oxygen uptake rate. By setting these values to be constant, the simulations were unable to predict the specific rate of glucose consumption and ethanol production (data not shown), as a fully respiratory metabolism is preferred. However, after forcing glycolytic fluxes through the pyruvate node, as it was observed by Baumann et al. (2010) [40], the simulations were capable to predict the flux of glucose



necessary to complete carbon requirements for biomass and protein synthesis, as well as to cover energy consumption and ethanol production under such respiro-fermentative conditions.

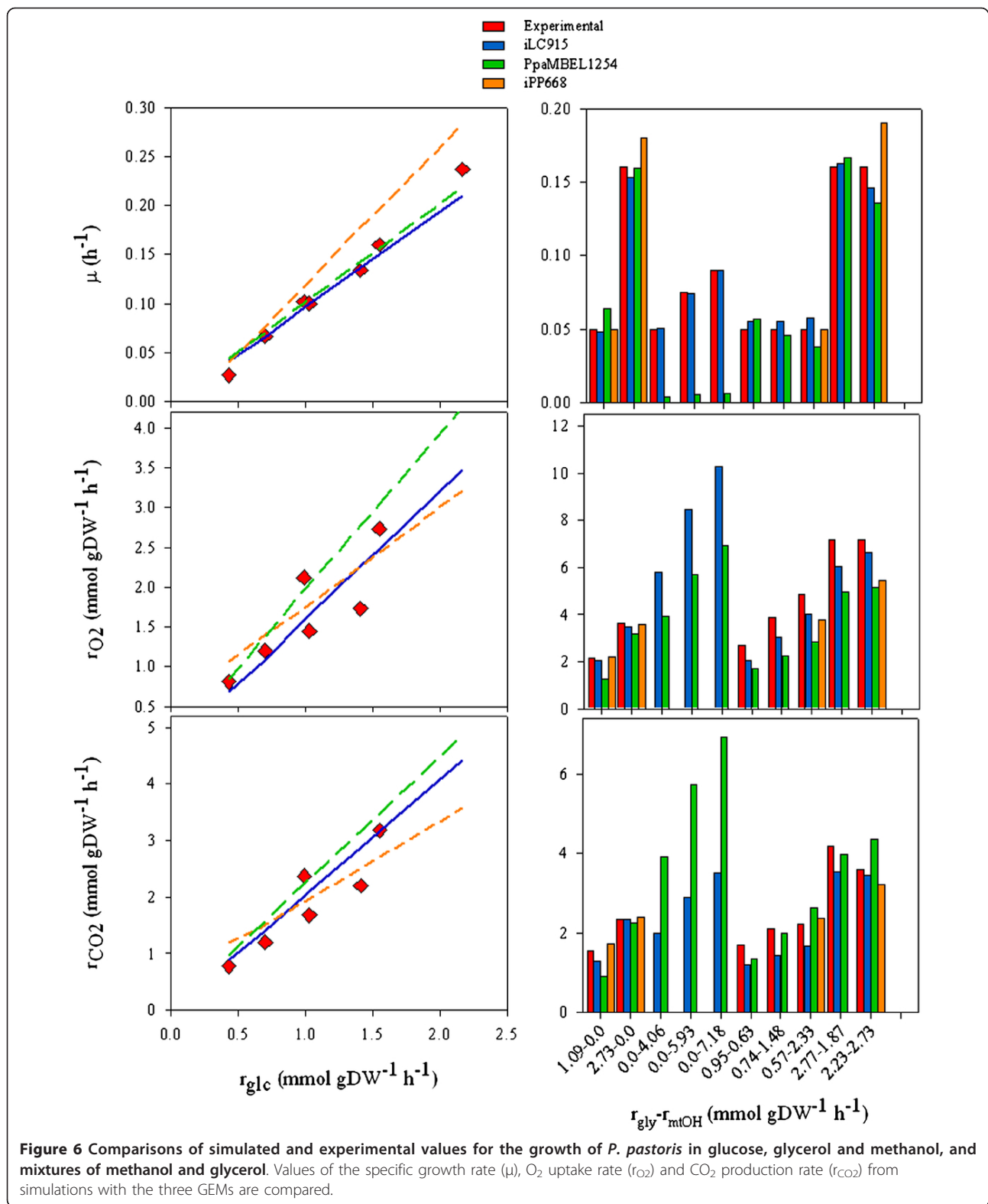
Discussion

The genome-scale metabolic models iLC915 and iSS884 of the methylotrophic yeast *P. pastoris* and the xylose fermenting *P. stipitis* were reconstructed and evaluated. Taken together, the results of constraint-based analysis used to evaluate cellular functions of these models were in close agreement with experimental results (Figures 6 and 5, and Table 2). Predicted physiological states of both yeasts in relation to biomass yields computed for different carbon sources of experimental relevance compared well to experimental values compiled by Verduyn (1991) [53]. ATP yields calculated by fitting the reconstructed model to experimentally-determined biomass yields on different carbon sources reproduced the dependence of Y_{xATP} on the nature of the carbon source previously observed [42,53]. Such calculations were extended to estimate the physiological benefits of the co-assimilation of glycerol with methanol using the iLC915 model of *P. pastoris* (Y_{xATP} decreased as the ratio of glycerol/methanol increased).

Several non-genetic approaches have been tested in order to optimize the production of recombinant

proteins using *P. pastoris* as host. One of them included the use of other well-assimilating carbon sources (e.g. glycerol and sorbitol) together with methanol, which is poorly assimilated and very toxic [14,15]. GEMs can be used to predict the capacity of *P. pastoris* for the synthesis of amino acids and recombinant proteins using methanol alone or in combination (Figure 7). Methanol is the worse carbon source for the synthesis of amino acids (since it is not well assimilated) and recombinant proteins. However, when it is used together with glycerol (e.g. 20-80%), the production increases to levels comparable to those obtained with glycerol alone. Predictions also suggested that some combinations of methanol-glycerol may be better to increase the production of FAB when compared to glycerol, sorbitol or glucose as the sole carbon source. Since simulations cannot predict the effect of such mixtures on the transcription of the AOX gene promoter, these findings must be complemented with experimental work.

P. stipitis has been used to produce ethanol from a synthetic media simulating the concentrations of glucose and xylose remaining after the hydrolysis of lignocellulose [45]. The metabolism of glucose exerts catabolic repression. Hence, xylose uptake starts after depletion of glucose. Grootjen et al. (1991) [45], therefore proposed the sequential fermentation of glucose-xylose mixtures using reactor in series. This strategy was reproduced *in silico* using the



iSS884 model (Figure 8). Simulations were performed by feeding the media with glucose and xylose in the first *in silico* bioreactor. Here, the glucose is rapidly metabolized

to produce biomass and ethanol. The output stream that contains xylose, remaining glucose, ethanol and biomass produced in the first reactor was the input for the second

Table 2 In-silico prediction of experimental parameters during the FAB production using three different oxygen uptake rates.

Parameter	Case 1 O2 inlet air = 8%		Case 2 O2 inlet air = 11%		Case 3 O2 inlet air = 21%	
	Pred	Exp	Pred	Exp	Pred	Exp
Y_{XFAB} (mg gDW ⁻¹)		0.54		0.38		0.22
r_{O_2} (mmol gDW ⁻¹ h ⁻¹)		4.14		2.26		2.23
Y_{SX} (gDW g ⁻¹)	0.31	0.25	0.43	0.45	0.48	0.47
r_{Glc} (mmol gDW ⁻¹ h ⁻¹)	1.81	1.85	1.28	1.27	1.15	0.98
r_{CO_2} (mmol gDW ⁻¹ h ⁻¹)	5.11	5.62	2.85	2.57	2.69	2.34
r_{EtOH} (mmol gDW ⁻¹ h ⁻¹)	0.71	1.12	0.11	0.17	0.00	0.00

Experimental data were taken from Baumann et al., (2008 and 2010) [39,40]

reactor. Herein glucose was completely depleted and almost half of the xylose was converted mainly into ethanol. However, the productivity of ethanol was lower than in the first reactor, since the consumption of glucose occurred faster than xylose. The stream coming out from the second reactor contained only xylose as carbon source. This was fed to the third reactor, where it was converted completely into ethanol. Results from this simulation are in close agreement with the data reported by Grootjen et al. (1991) [45], and shows the potential of the simulations with the iSS884 GEM for the design of strategies to produce ethanol from different combinations of glucose and xylose coming from the hydrolysis of lignocellulose.

By comparing the reconstructed GEMs of both *Pichia* species with *S. cerevisiae*, we found that the maximum capacities for the production of the 20 amino acids and the essentiality of reactions for the three yeasts to grow on glucose are similar (Additional file 4). This can be explained by the large number of shared metabolic reactions, in particular in the central, amino acid, lipid and nucleotide metabolism (Additional file 2). Llorente et al. (2000a and b) [56,57] suggested, based on a comparative analysis of chromosome maps and synteny with *S. cerevisiae*, that yeasts evolution may be driven by a balance between gene duplications and deletions, given change to the existence of orthologous and paralogous genes. Genes

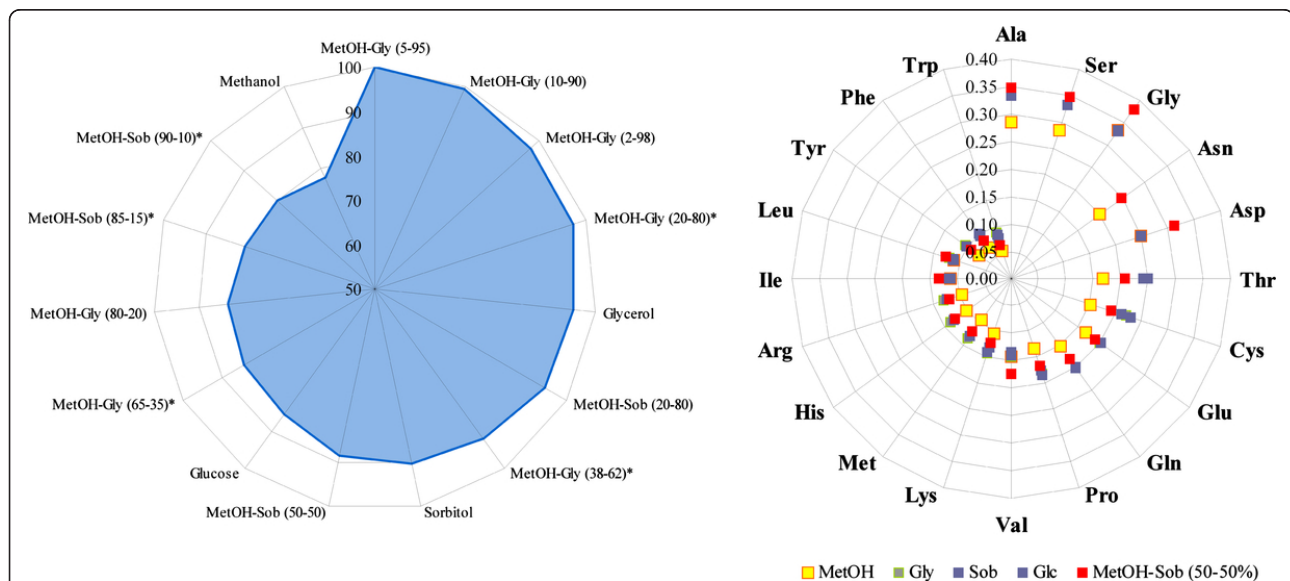
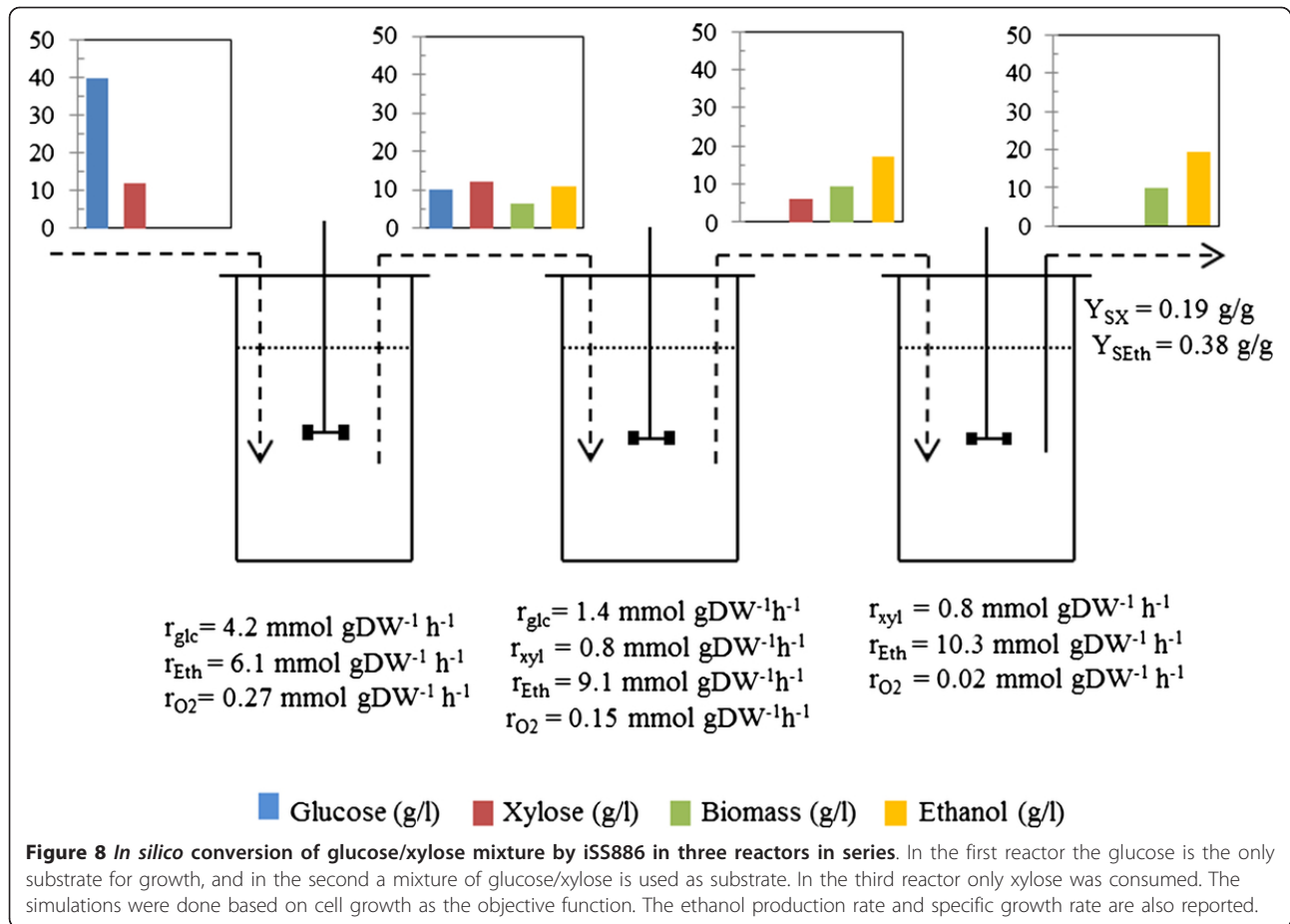


Figure 7 Comparison between the maximum capacities for amino acids and recombinant protein production with different carbon sources using the iLC915 model of *P. pastoris*. The production of amino acids was evaluated with glucose, glycerol and methanol, as well as with a mixture of methanol with sorbitol, data is expressed in mole of amino acid per carbon mole of substrate. The production of FAB was simulated with glucose, glycerol, sorbitol and methanol, and mixtures of methanol with glycerol or sorbitol. FAB production was used as the objective function. Data is expressed as the percentage of the maximum production, which was obtained with a mixture 5-95% of methanol and glycerol. Asterisks over some mixtures indicate that such combinations were already evaluated in experimental work, and therefore can be used in a production process.



coding for proteins involved in basic tasks are quite conserved, meanwhile unnecessary gene copies have been deleted rather than conserved, and hence unusual metabolic functions are almost entirely absent [58]. The latter applied for the particular capabilities of *P. pastoris* and *P. stipitis* to metabolize methanol and xylose as a sole carbon source. The metabolism of xylose is a singular case, since even when *S. cerevisiae* has coding genes to transport and reduce xylose, as well as for the dehydrogenation of xylitol, it cannot grow on xylose as a sole carbon source. This suggests that unusual metabolic traits may emerge not from the use of existing enzymes, but from a larger number of genes and/or regulatory factors found in the native organisms [59]. Metabolic modeling at genome scale is a useful tool to investigate these traits.

Conclusions

In summary, we presented two high quality GEMs that can be used to gain understanding on the metabolic capabilities of the two yeasts *P. stipitis* and *P. pastoris*. These models also provide a versatile tool for rationale strain improvement, scaffold for data integration and

interpretation and evolutionary analysis of yeasts metabolism.

Additional material

Additional file 1: Composition of the human monoclonal antibody 3H6 Fab fragment (FAB) used as model recombinant protein.

Additional file 2: Gene orthology analysis among the three yeast species: *Pichia stipitis*, *Pichia pastoris*, and *Saccharomyces cerevisiae*.

Additional file 3: Analysis of reaction essentiality predicted by simulations.

Additional file 4: Maximum capacities for amino acids production among the three yeast species: *Pichia stipitis*, *Pichia pastoris*, and *Saccharomyces cerevisiae*.

Acknowledgements

We thank Tobias Österlund for his thoughtful discussion and supporting us with his information during this project. This work was financially supported by the Knut and Alice Wallenberg Foundation and Chalmers Foundation.

Authors' contributions

LC and SS reconstructed the GEMs, designed and performed simulations, and wrote the manuscript. RA and IN assisted with the reconstruction and modeling and reviewed the manuscript. JN supervised the work and

assisted in manuscript preparation. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 22 December 2011 Accepted: 4 April 2012

Published: 4 April 2012

References

1. Van Dijken JP, Van Den Bosch E, Hermans JJ, De Miranda LR, Scheffers WA: **Alcoholic fermentation by 'non-fermentative' yeasts.** *Yeast* 1986, **2**:123-127.
2. Slininger PJ, Branstrator LE, Bothast RJ, Okos MR, Ladisch MR: **Growth, death, and oxygen uptake kinetics of *Pichia stipitis* on xylose.** *Biotechnol Bioeng* 1991, **37**:973-980.
3. Van Vleet JH, Jeffries TW: **Yeast metabolic engineering for hemicellulosic ethanol production.** *Curr Opin Biotechnol* 2009, **20**:300-306.
4. Thomas WJ: **Engineering yeasts for xylose metabolism.** *Curr Opin Biotechnol* 2006, **17**:320-326.
5. Skoog K, Hahn-Hagerdal B: **Effect of Oxygenation on Xylose Fermentation by *Pichia stipitis*.** *Appl Environ Microbiol* 1990, **56**:3389-3394.
6. Shi NQ, Jeffries TW: **Anaerobic growth and improved fermentation of *Pichia stipitis* bearing a URA1 gene from *Saccharomyces cerevisiae*.** *Appl Microbiol Biotechnol* 1998, **50**:339-345.
7. Dahn K, Davis B, Pittman P, Kenealy W, Jeffries T: **Increased xylose reductase activity in the xylose-fermenting yeast *Pichia stipitis* by overexpression of XYL1.** *Appl Biochem Biotechnol* 1996, **57-58**:267-276.
8. De Deken RH: **The Crabtree effect: a regulatory system in yeast.** *J Gen Microbiol* 1966, **44**:149-156.
9. Klinner U, Fluthgraf S, Freese S, Passoth V: **Aerobic induction of respiratory fermentative growth by decreasing oxygen tensions in the respiratory yeast *Pichia stipitis*.** *Appl Microbiol Biotechnol* 2005, **67**:247-253.
10. Wegner GH, Harder W: **Methylotrophic yeasts – 1986.** *Antonie Van Leeuwenhoek* 1987, **53**:29-36.
11. Wriessnegger T, Gubitz G, Leitner E, Ingolic E, Cregg J, de la Cruz BJ, Daum G: **Lipid composition of peroxisomes from the yeast *Pichia pastoris* grown on different carbon sources.** *Biochim Biophys Acta* 2007, **1771**:455-461.
12. Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR: **Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*.** *Mol Cell Biol* 1985, **5**:1111-1121.
13. Cregg JM, Cereghino JL, Shi J, Higgins DR: **Recombinant protein expression in *Pichia pastoris*.** *Mol Biotechnol* 2000, **16**:23-52.
14. Celik E, Calik P, Oliver SG: **Fed-batch methanol feeding strategy for recombinant protein production by *Pichia pastoris* in the presence of co-substrate sorbitol.** *Yeast* 2009, **26**:473-484.
15. Jungo C, Marison I, von Stockar U: **Mixed feeds of glycerol and methanol can improve the performance of *Pichia pastoris* cultures: a quantitative study based on concentration gradients in transient continuous cultures.** *J Biotechnol* 2007, **128**:824-837.
16. Inan M, Meagher MM: **Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of *Pichia pastoris*.** *J Biosci Bioeng* 2001, **92**:585-589.
17. Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi B-K, et al: **Humanization of yeast to produce complex terminally sialylated glycoproteins.** *Science* 2006, **313**:1441-1443.
18. Nielsen J, Jewett MC: **Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2008, **8**:122-131.
19. De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, Van de Peer Y, Callewaert N: **Genome sequence of the recombinant protein production host *Pichia pastoris*.** *Nat Biotechnol* 2009, **27**:561-566.
20. Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, Salamov A, Schmutz J, Lindquist E, Dehal P, Shapiro H, et al: **Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*.** *Nat Biotechnol* 2007, **25**:319-326.
21. Oberhardt MA, Palsson BO, Papin JA: **Applications of genome-scale metabolic reconstructions.** *Mol Syst Biol* 2009, **5**:320.
22. Famili I, Forster J, Nielsen J, Palsson BO: ***Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network.** *Proc Natl Acad Sci USA* 2003, **100**:13134-13139.
23. Patil KR, Akesson M, Nielsen J: **Use of genome-scale microbial models for metabolic engineering.** *Curr Opin Biotechnol* 2004, **15**:64-69.
24. Patil KR, Nielsen J: **Uncovering transcriptional regulation of metabolism by using metabolic network topology.** *Proc Natl Acad Sci USA* 2005, **102**:2685-2689.
25. Patil K, Rocha I, Forster J, Nielsen J: **Evolutionary programming as a platform for in silico metabolic engineering.** *BMC Bioinformatics* 2005, **6**:308.
26. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale metabolic network reconstructions.** *Genome Res* 2004, **14**:301-312.
27. Usaite R, Patil KR, Grotkjaer T, Nielsen J, Regenberg B: **Global transcriptional and physiological responses of *Saccharomyces cerevisiae* to ammonium, L-Alanine, or L-Glutamine Limitation.** *Appl Environ Microbiol* 2006, **72**:6194-6203.
28. Bordel S, Agren R, Nielsen J: **Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes.** *PLoS Comput Biol* 2010, **6**:e1000859.
29. Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee DY: **Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement.** *Microb Cell Fact* 2010, **9**:50.
30. Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY: **Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production.** *Biotechnol J* 2010, **5**:705-715.
31. David H, Ozcelik IS, Hofmann G, Nielsen J: **Analysis of *Aspergillus nidulans* metabolism at the genome-scale.** *BMC Genomics* 2008, **9**:163.
32. Yu C-S, Chen Y-C, Lu C-H, Hwang J-K: **Prediction of protein subcellular localization.** *Proteins: Structure, Function, and Bioinformatics* 2006, **64**:643-651.
33. Egli T, Van Dijken JP, Veenhuis M, Harder W, Fiechter A: **Methanol metabolism in yeasts: regulation of the synthesis of catabolic enzymes.** *Arch Microbiol* 1980, **124**:115-121.
34. Toivola A, Yarrow D, Van Den Bosch E, van Dijken JP, Scheffers WA: **Alcoholic Fermentation of D-Xylose by Yeasts.** *Appl Environ Microbiol* 1984, **47**:1221-1223.
35. Carnicer M, Baumann K, Topfitz I, Sanchez-Ferrando F, Mattanovich D, Ferrer P, Albiol J: **Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels.** *Microb Cell Fact* 2009, **8**:65.
36. Wriessnegger T, Leitner E, Belegatis MR, Ingolic E, Daum G: **Lipid analysis of mitochondrial membranes from the yeast *Pichia pastoris*.** *Biochim Biophys Acta* 2009, **1791**:166-172.
37. Rizzi M, Harwart K, Erlemann P, Bui-Thanh N-A, Dellweg H: **Purification and properties of the NAD⁺-xylytol-dehydrogenase from the yeast *Pichia stipitis*.** *J Ferment Bioeng* 1989, **67**:20-24.
38. Yuan T, Ren Y, Meng K, Feng Y, Yang P, Wang S, Shi P, Wang L, Xie D, Yao B: **RNA-Seq of the xylose-fermenting yeast *Scheffersomyces stipitis* cultivated in glucose or xylose.** *Appl Microbiol Biotechnol* 2011, **92**:1237-1249.
39. Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D: **Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins.** *Biotechnol Bioeng* 2008, **100**:177-183.
40. Baumann K, Carnicer M, Dragosits M, Graf A, Stadlmann J, Jouhten P, Maaheimo H, Gasser B, Albiol J, Mattanovich D, Ferrer P: **A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions.** *BMC Syst Biol* 2010, **4**:141.
41. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110.** *Appl Environ Microbiol* 1994, **60**:3724-3731.
42. Verduyn C, Stouthamer AH, Scheffers WA, Dijken JP: **A theoretical evaluation of growth yields of yeasts.** *Antonie Van Leeuwenhoek* 1991, **59**:49-63.
43. Price ND, Papin JA, Schilling CH, Palsson BO: **Genome-scale microbial in silico models: the constraints-based approach.** *Trends Biotechnol* 2003, **21**:162-169.
44. Park JM, Kim TY, Lee SY: **Constraints-based genome-scale metabolic simulation for systems metabolic engineering.** *Biotechnol Adv* 2009, **27**:979-988.

45. Grootjen DRJ, Jansen ML, van der Lans RGJM, Luyben KCAM: **Reactors in series for the complete conversion of glucose/xylose mixtures by *Pichia stipitis* and *Saccharomyces cerevisiae*.** *Enzyme Microb Technol* 1991, **13**:828-833.
46. Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Buthgen N, Borger S, Costenoble R, Heinemann M, et al: **A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology.** *Nat Biotech* 2008, **26**:1155-1160.
47. Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY: **Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production.** *Biotechnol J* 2010, **5**:705-715.
48. Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee DY: **Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement.** *Microb Cell Fact* 2010, **9**:50.
49. de Vries S, Marres CA: **The mitochondrial respiratory chain of yeast. Structure and biosynthesis and the role in cellular metabolism.** *Biochim Biophys Acta* 1987, **895**:205-239.
50. Veiga A, Arrabaça JD, Loureiro-Dias MC: **Cyanide-resistant respiration, a very frequent metabolic pathway in yeasts.** *FEMS Yeast Res* 2003, **3**:239-245.
51. Kern A, Hartner FS, Freigassner M, Spielhofer J, Rumpf C, Leitner L, Frohlich KU, Glieder A: ***Pichia pastoris* "just in time" alternative respiration.** *Microbiology* 2007, **153**:1250-1260.
52. Nookaew I, Jewett MC, Meechai A, Thammarongtham C, Laoteng K, Cheevadhanarak S, Nielsen J, Bhumiratana S: **The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism.** *BMC Syst Biol* 2008, **2**:71.
53. Verduyn C: **Physiology of yeasts in relation to biomass yields.** *Antonie Van Leeuwenhoek* 1991, **60**:325-353.
54. Fiaux J, Cakar ZP, Sonderegger M, Wuthrich K, Szyperski T, Sauer U: **Metabolic-flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*.** *Eukaryot Cell* 2003, **2**:170-180.
55. Skoog K, Jeppsson H, Hahn-Hägerdal B: **The effect of oxygenation on glucose fermentation with *pichia stipitis*.** *Appl Biochem Biotechnol* 1992, **34-35**:369-375.
56. Llorente B, Malpertuy A, Neuvéglise C, de Montigny J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, et al: **Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*.** *FEBS Lett* 2000, **487**:101-112.
57. Llorente B, Durrens P, Malpertuy A, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, et al: **Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*.** *FEBS Lett* 2000, **487**:122-133.
58. Dujon B: **Yeast evolutionary genomics.** *Nat Rev Genet* 2010, **11**:512-524.
59. Alper H, Stephanopoulos G: **Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential?** *Nat Rev Micro* 2009, **7**:715-723.

doi:10.1186/1752-0509-6-24

Cite this article as: Caspeta et al.: Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in silico evaluation of their potentials. *BMC Systems Biology* 2012 **6**:24.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

