



CHALMERS

Chalmers Publication Library

Genome Scale Metabolic Modeling of the Riboflavin Overproducer *Ashbya gossypii*

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

Biotechnology and Bioengineering (ISSN: 0006-3592)

Citation for the published paper:

Ledesma-Amaro, R. ; Kerkhoven, E. ; Revuelta, J. (2014) "Genome Scale Metabolic Modeling of the Riboflavin Overproducer *Ashbya gossypii*". *Biotechnology and Bioengineering*, vol. 111(6), pp. 1191-1199.

<http://dx.doi.org/10.1002/bit.25167>

Downloaded from: <http://publications.lib.chalmers.se/publication/200098>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Genome Scale Metabolic Modeling of the Riboflavin Overproducer *Ashbya gossypii*

Rodrigo Ledesma-Amaro,¹ Eduard J. Kerkhoven,² José Luis Revuelta,¹ Jens Nielsen²

¹Departamento de Microbiología y Genética, Metabolic Engineering Group, Universidad de Salamanca, Campus Miguel de Unamuno, Salamanca, Spain

²Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden; telephone: +46-31-772-38-04; fax: +46(0)31 772 3801; e-mail: nielsenj@chalmers.se

ABSTRACT: *Ashbya gossypii* is a filamentous fungus that naturally overproduces riboflavin, or vitamin B2. Advances in genetic and metabolic engineering of *A. gossypii* have permitted the switch from industrial chemical synthesis to the current biotechnological production of this vitamin. Additionally, *A. gossypii* is a model organism with one of the smallest eukaryote genomes being phylogenetically close to *Saccharomyces cerevisiae*. It has therefore been used to study evolutionary aspects of bakers' yeast. We here reconstructed the first genome scale metabolic model of *A. gossypii*, iRL766. The model was validated by biomass growth, riboflavin production and substrate utilization predictions. Gene essentiality analysis of the *A. gossypii* model in comparison with the *S. cerevisiae* model demonstrated how the whole-genome duplication event that separates the two species has led to an even spread of paralogs among all metabolic pathways. Additionally, iRL766 was used to integrate transcriptomics data from two different growth stages of *A. gossypii*, comparing exponential growth to riboflavin production stages. Both reporter metabolite analysis and in silico identification of transcriptionally regulated enzymes demonstrated the important involvement of beta-oxidation and the glyoxylate cycle in riboflavin production.

Biotechnol. Bioeng. 2013;9999: 1–9.

© 2013 Wiley Periodicals, Inc.

KEYWORDS: *Ashbya gossypii*; riboflavin; genome-scale metabolic model; *Eremothecium*; systems metabolic engineering; microbial biotechnology

Introduction

Ashbya gossypii (or *Eremothecium gossypii*) is a filamentous hemiascomycete with unique features which make it a widely studied organism in both basic research (Wendland and Walther, 2005) and industrial applications (Kato and Park, 2012).

Its genome, which has been sequenced and annotated (Dietrich et al., 2004; Hermida et al., 2005), is one of the smallest eukaryotic genomes with only 4,718 protein-coding genes. It is highly similar to the well-studied organism *Saccharomyces cerevisiae* and 95% of *A. gossypii* genes have homologs in yeast (Dietrich et al., 2004; Fig. 1A). In spite of this high homology, the *S. cerevisiae* genome encodes over 800 additional genes (5,570 in total), many of which have presumably originated from a whole-genome duplication event (Wolfe and Shields, 1997). While most components are shared between the two genomes, the life cycles of these two organisms show important differences, such as the probable lack of a sexual cycle and the multinucleated hyphae development in *A. gossypii*. It has therefore been considered a model organism to study the minimal eukaryotic free-living cell (Finlayson et al., 2011), the event of genome duplication (Wolfe and Shields, 1997), the pheromone signal cascade and sexual cycle (Wendland et al., 2011), fungal development and polarized hyphal growth (Schmitz and Philippsen, 2011) and to reinvestigate the *S. cerevisiae* genome (Brachat et al., 2003).

A. gossypii is a natural overproducer of riboflavin (vitamin B2), which animals and humans cannot synthesize. Riboflavin is a high value industrial compound used not only in pharmaceuticals, but also in animal feed additives, cosmetics and in the food industry (Shi et al., 2009). *A. gossypii* has been considered a paradigm of the industry-friendly white biotechnology and it is nowadays one of the world's top producers of riboflavin by microbial fermentation (Kato and Park, 2012; Ledesma-Amaro et al., 2013b; Stahmann et al., 2000). The availability of the genome sequence and genetic engineering tools (Wendland et al., 2000) have allowed the development of metabolic engineering approaches that have significantly increased the vitamin production titer (Jimenez

The authors declare that they have no competing interests.

Correspondence to: J. Nielsen

Contract grant sponsor: Spanish Ministerio de Economía y Competitividad

Contract grant number: BIO2011-23901

Received 26 August 2013; Revision received 5 November 2013; Accepted 26 November 2013

Accepted manuscript online xx Month 2013;

Article first published online in Wiley Online Library (wileyonlinelibrary.com).

DOI 10.1002/bit.25167

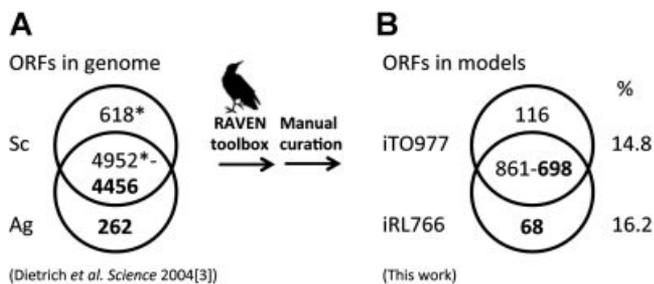


Figure 1. Comparison between the *A. gossypii* and *S. cerevisiae* genomes and the iTO977 and iRL766 models. (A) Comparison of the total ORFs in the genomes of *S. cerevisiae* (Sc) and *A. gossypii* (Ag). (B) Comparison of the genes in the yeast model iTO977 and the *Ashbya* model iRL766, after extensive manual curation. The overlapping regions of (A) and (B) represent the number of homologous genes. The numbers in bold are *A. gossypii* genes with homologs in *S. cerevisiae* while normal numbers are genes from *S. cerevisiae* with homologs in *A. gossypii*. Asterisks denote approximate value.

et al., 2005, 2008; Park et al., 2011). In addition to its well-established use in riboflavin production, the capabilities of *A. gossypii* to produce proteins (Ribeiro et al., 2010, 2013) and bioethanol from glycerol (Ribeiro et al., 2012) are being investigated.

Genome-scale metabolic models (GEMs) have become an important tool in the understanding of metabolic networks and they have been useful in both the study and the applications of biological systems (Garcia-Albornoz and Nielsen, 2013; Kim et al., 2012; Ledesma-Amaro et al., 2013a). GEMs can serve to describe novel metabolic networks, find essential genes and metabolites, predict substrate consumption capabilities and aid in re-annotating genomic information (Costanzo et al., 2010; Ibarra et al., 2002; Vongsangnak et al., 2008). Additionally, they can be used as a scaffold to integrate omics experiments and thus the physiological differences between different conditions can be analyzed in the context of the metabolic network (Osterlund et al., 2013). Thirdly, GEMs have been widely used in the biotechnology field for identification of metabolic bottlenecks and novel metabolic engineering targets (Agren et al., 2013b; Caspeta and Nielsen, 2013; Otero et al., 2013; Park et al., 2007b).

Here we present the first curated genome-scale metabolic model of the filamentous fungus *A. gossypii*, called iRL766. Along this work iRL766 was validated and its capabilities to predict biomass growth, riboflavin production and substrate utilization were shown. Gene essentiality was studied in the context of its own network and of its close relative *S. cerevisiae*, enlightening the event of the whole genome duplication. Due to the industrial interest in this organism for riboflavin production, the biosynthetic pathway of the vitamin within the model was studied in more detail. Finally the metabolic network was used to integrate transcriptomics data and transcriptionally controlled reactions in the transition from a trophic phase to a riboflavin productive phase, representing possible targets for further strain engineering.

Materials and Methods

Reconstruction of Genome-Scale Metabolic Model

The first step in reconstructing the iRL766 model was to create two draft models using the RAVEN toolbox (Agren et al., 2013a). The first draft model was based on protein orthology between the *S. cerevisiae* model iTO977 (Osterlund et al., 2013) and *A. gossypii*. For this, homology between the protein coding sequences of the *S. cerevisiae* and *A. gossypii* genomes were evaluated using bi-directional BLASTp, with the following cut-offs: E-value $< 1e-30$, identity $> 40\%$ and alignment length > 200 amino acids. All 1-1 orthologs were included, to capture potential paralogs while ensuring that only genes that map back to the original gene in the BLASTp in the opposite direction are included. For the second draft model the *A. gossypii* genome was queried using hidden Markov models (HMMs), which were constructed from Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2012) consensus sequences, as described in (Agren et al., 2013a). The protein sequence of *A. gossypii* strain ATCC10895 (<http://genolevures.org/>) was used for construction of both draft models. The two draft models were merged to generate a model representing reactions with and without homology in *S. cerevisiae*, using the naming convention and charge balancing from iTO977. Subsequently, reactions in the combined model were manually curated to assure correctness with respect to reaction stoichiometry, metabolites and co-factors involved, and to identify gaps in the metabolic network. Manual curation was performed using KEGG, MetaCyc (Caspeta et al., 2012), Saccharomyces Genome Database (Cherry et al., 2012), Ashbya Genome Database (Gattiker et al., 2007), and available literature. Where *A. gossypii* specific knowledge of gene-associations, metabolic reactions and localization was insufficient, data from the closely related *S. cerevisiae* was used.

The biomass equation and ATP for maintenance were taken from the iTO977 model, as insufficient information is available from *A. gossypii*. When oil was used as a carbon source, the biomass equation was modified in lipid composition according to previously reported data (Stahmann et al., 1994).

All simulations were performed using RAVEN and COBRA toolboxes in Matlab. iRL766 is available in the BioMet Toolbox (Cvijovic et al., 2010; <http://sysbio.se/BioMet/>—will be uploaded upon acceptance of paper).

Constraints-Based Flux Analysis and Simulations

Flux balance analysis was widely performed used in simulations with the reconstructed GEMs (Park et al., 2009). Loops were removed from the solution by minimizing the number of reactions carrying flux. To simulate the trophic phase, either glucose or oleic acid uptake rates were set, while biomass production was set as objective function. To simulate the riboflavin production phase, the consumption rate of lipid bodies was set, while riboflavin production was set as

objective function. ATP for maintenance was not considered in the productive phase since we assumed that this is produced from different molecular sources (as mutants unable to consume lipid bodies have the same growth pattern as wild type (Ledezma-Amaro R., personal communication). Experimental data were extracted from previous publications using EasyNData (<http://puwer.web.cern.ch/puwer/EasyNData/>).

Gene Essentiality Predictions

In silico evaluation of single gene knock-outs were performed using the RAVEN toolbox, which takes the presence of isoenzymes into account. Lethality was estimated according to a relative fitness value ($f = \text{mutant growth rate/wild type growth rate}$), when f was smaller than the cutoff value (0.9) the gene was considered essential (Osterlund et al., 2013). Less strict cutoff values did not drastically reduce the number of lethal genes (Supplementary File 7). Two different glucose-based media were used, minimal media (MM) and rich media (YPD). Minimal media consisted of the uptake of glucose, ammonium, oxygen, sulfate and phosphate while YPD media also allowed the uptake of amino acids and nucleotides.

In Silico Prediction of Carbon and Nitrogen Utilization

Assessment of growth on different carbon sources was performed by constraining the uptake of each carbon source individually to 6 C-mol/gCDW/h and maximizing for cell growth. In silico growth rates were qualitatively compared to experimental growth rates, and therefore categorized as: no growth, reduced growth ($0.001\text{--}0.090\text{ h}^{-1}$) and normal growth (0.090 h^{-1}). Experimental data were classified in the same three categories.

Integrative Data Analysis

Normalized microarray dataset E-MEXP-1945 (Gattiker et al., 2007) was obtained from ArrayExpress (www.ebi.ac.uk/arrayexpress/). Eighteen and 103 h samples were analyzed, corresponding to fast growing advanced mycelia and sporulating mycelia, respectively. The sporulating mycelia represent the riboflavin productive stage, as sporulation is a process directly correlated to vitamin production (Stahmann et al., 2001). These data were integrated in the model iRL766 in order to identify reporter metabolites and transcriptionally regulated key enzymes. Reporter metabolites were identified with R package Piano (Varemo et al., 2013). A random sampling algorithm was applied in order to identify transcriptionally regulated enzymes (Bordel et al., 2010). Normalized array data as described above were used as input. The upper and lower bounds of the exchange fluxes and biomass or riboflavin reactions in iRL766 were constrained according to experimentally measured fluxes for each condition (Supplementary file 5). A set of 500 flux distributions was

generated for each of the considered conditions (Osterlund et al., 2013). Z-scores were calculated from the two sets of flux distributions and transformed into probabilities of change by using the cumulative Gaussian distribution, as described in Bordel et al. (2010). The values of the moderated t -statistic from the transcriptomics data were used to calculate a probabilistic score P for each reaction representing the probability that the flux and the transcription are significantly changed in the same direction between the two conditions (cutoff $P > 0.9$).

Results and Discussion

Reconstruction and Comparative Analysis of the *A. gossypii* Metabolic Network

The high sequence similarity between *S. cerevisiae* and *A. gossypii* allowed us to create a first draft model using the yeast model iTO977 as template (Osterlund et al., 2013), using bi-directional BLASTp, containing 725 genes. No homologs could be found for 69 genes from iTO977, corresponding to 76 reactions (Supplementary file 3). Among these were reactions involved in biotin and myo-inositol synthesis, rendering *A. gossypii* auxotrophic for both nutrients (Demain, 1972). Additionally, 12 reactions related to galactose metabolism were absent in *A. gossypii*, explaining the incapability of this organism for growth using galactose as a sole carbon source (Kurtzman et al., 2011). The use of a *S. cerevisiae* model as a template has the disadvantage that the resulting draft model can only contain genes, and their associated reactions, that have high sequence identity in *A. gossypii*. To include the function of *A. gossypii* genes that were not represented in the *S. cerevisiae* genome, a second draft model was generated based on similarity with KEGG consensus sequences (Kanehisa et al., 2012), containing 570 genes. KEGG consensus sequences provided only 11 *A. gossypii* genes with no homologs in *S. cerevisiae*, responsible for 30 reactions. Eleven of these reactions were predicted to require riboflavin derivatives, such as flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD). These reactions were added to the first draft model, by transferring the reactions indicated in Supplementary file 3. Subsequent extensive manual curation of the merged model generated the model iRL766. iRL766 included 766 genes which were involved in 1,595 reactions, comprising 799 unique metabolites distributed in four different compartments (cytoplasm, mitochondria, peroxisome, and extracellular). After manual curation, the number of unique genes in each model increased to 116 and 69 for the iTO977 and iRL766 models, respectively. Although the model has less annotated genes than iTO977 (977 genes) they still constitute 16.2% of the total open reading frames of the genome, which is a higher coverage than for the yeast model (14.8%; Fig. 1B).

While some differences between the models were observed, it is unlikely that these are sufficient to explain

the unique properties of each strain but a combination of these and the unequal regulation at several biological levels.

Riboflavin Biosynthetic Pathway in iRL766

A. gossypii is used on a large-scale in the bioindustry for riboflavin production, often using glucose (Jimenez et al., 2008) or oils (Park et al., 2011) as carbon source. The oils used are typically rich in triacylglycerols, predominantly composed of oleic acid (triolein). To investigate the production of riboflavin from glucose and triolein, FBA was performed using either carbon source and optimized for riboflavin production (Supplementary file 4 and Fig. 2).

When glucose was used as sole carbon source, 86 reactions were involved in riboflavin production (Fig. 2A, Supplementary file 4). The most important pathways activated in the model were glycolysis, TCA cycle and oxidative phosphory-

lation, to provide the energy supply, followed by purine metabolism and pentose phosphate pathway to provide the precursors GTP and ribulose-5-phosphate respectively, and finally riboflavin metabolism to synthesize the active vitamin.

When oleic acid was used as carbon source a total of 134 reactions were involved in riboflavin synthesis (Fig. 2B, Supplementary file 4). The activities of glycolysis and the pentose phosphate pathway were reduced as fatty acid degradation was now involved in substrate utilization. Precursors and riboflavin biosynthesis reactions remained almost invariant during growth on both carbon sources.

These pathways described by iRL766 are in concordance with previously published data (Park et al., 2007a, 2011; Stahmann et al., 2000; Sugimoto et al., 2009). Flux variability analysis was performed to search for alternative solutions (Supplementary file 10), and these results adhere to the same pathways as shown in Figure 2.

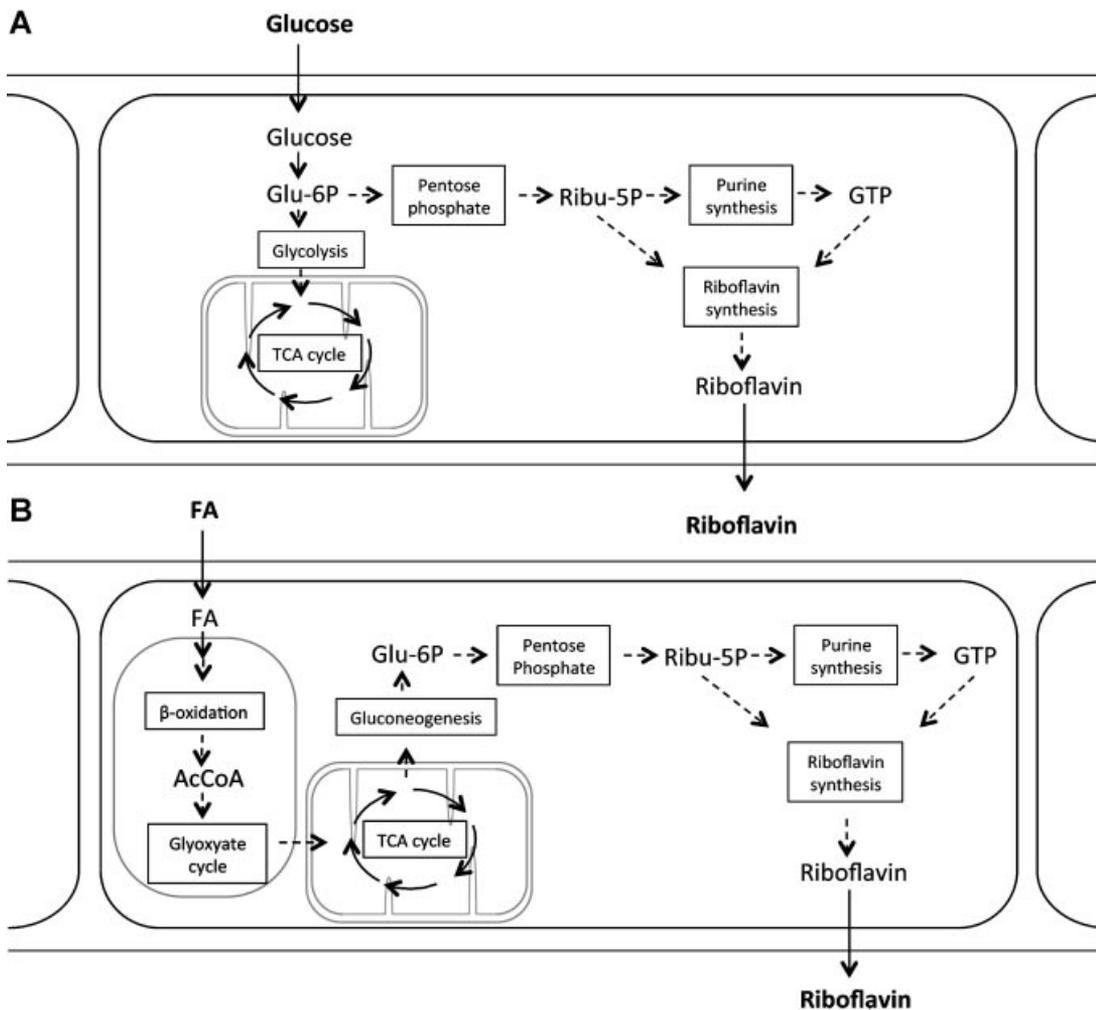


Figure 2. Scheme of pathways that carry flux when the riboflavin production phase is simulated in iRL766 when (A) glucose or (B) oleic acid is used as carbon source. FA, fatty acid; AcCoA, acetyl coA; TCA, tricarboxylic; Glu-6P, glucose-6-phosphate; Ribu-5P, ribulose-5-phosphate.

Growth Rate and Riboflavin Synthesis Predictions

A. gossypii has two growth stages, a first trophic phase, when cells grow exponentially while riboflavin production is minimal; and a second phase, or productive phase when cells stop growing and riboflavin is overproduced (Mateos et al., 2006). These two conditions were simulated in the model, to validate the growth rate and riboflavin production predictions with experimental data (Supplementary file 5). The growth predictions on both glucose and triolein were in good concordance with experimental data (Fig. 3).

There is limited comprehensive data available that can be used to validate the riboflavin productive phase. Firstly, the carbon source used for the vitamin production has not been entirely elucidated. In most experimental approaches, glucose and oils have been exhausted from the media at the end of the exponential phase. It has been hypothesized that fatty acids, accumulated as lipid bodies during growth, are remobilized in the productive phase permitting riboflavin synthesis, supported by microscopy and Nile red staining of *A. gossypii* lipid bodies (Lim et al., 2003; Stahmann et al., 1994, Ledesma-Amaro et al., 2013c). This hypothesis is further supported by an increase in riboflavin synthesis when oils are used for growth, as *A. gossypii* accumulates at least twice the amounts of lipids during the first growth phase in these conditions (Stahmann et al., 1994). Using experimental lipid bodies consumption rate to constrain the model and riboflavin production as objective function resulted in a vitamin production rate of 0.0156 mmol/gCDW/h, which was close to the experimental value of 0.0126 mmol/gCDW/h (Stahmann et al., 2001).

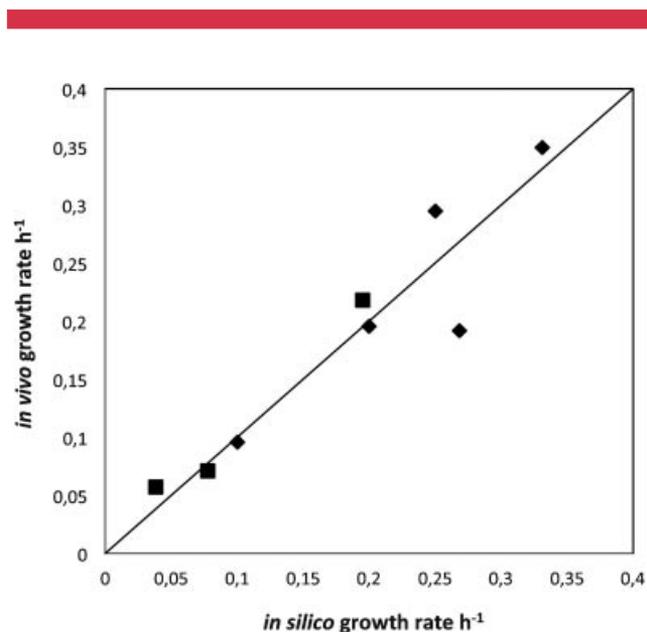


Figure 3. Comparison between the *in silico* prediction of specific growth rate with experimental data. Growth phenotypes were collected from literature and compared with simulated values when glucose (diamonds) and oils (squares) were used as carbon sources (Supplementary file 5). Black line represent perfect correlation.

In Silico Prediction of Carbon and Nitrogen Utilization

The model was further validated by comparison with experimental growth data on different carbon and nitrogen sources (Kurtzman et al., 2011; Ribeiro et al., 2012; Table I and Supplementary file 6). Of the 41 tested carbon and nitrogen sources, 39 showed comparable growth in both model and experiments, while only D-ribose and D-xylose were identified in the model as false positives. While *A. gossypii* is capable of growth with D-ribose as sole carbon source, it can do so only in concentrations below 0.3% due to possible toxicity, which cannot be predicted by the model (Revuelta, J.L. personal communication). Regarding D-xylose, it was recently found that while the machinery for D-xylose consumption is encoded in the *A. gossypii* genome, as demonstrated in iRL766, D-xylose can only be converted to xylitol and not metabolized further (Ribeiro et al., 2012). Even though the putative genes for xylose catabolism are present, they cannot be expressed to sufficient level to ensure growth, as it is the case in *S. cerevisiae* (Scalcinati et al., 2012; van Zyl et al., 1989).

In Silico Analysis of Reaction Essentiality

As *A. gossypii* has one of the smallest eukaryotic genomes (Dietrich et al., 2004), iRL766 represents an interesting platform to study gene essentiality and the minimal set of reactions to maintain a free-living eukaryotic cell. Additionally, *A. gossypii* shares high homology to the well-studied yeast *S. cerevisiae*, while baker's yeast encodes 852 more genes within its genome, due to the whole-genome duplication event that took place in baker's yeast but not in *A. gossypii* (Dietrich et al., 2004). We compared gene essentiality in iRL766 to iTO977 to investigate this WGD event.

In both models, iTO977 and iRL766, each gene was knocked out sequentially, and the effect on the specific growth rate was examined in minimal and rich media (MM and YPD). As anticipated, the fraction of essential genes in the model increased from 18.9% in yeast to 22.5% in *A. gossypii* in MM, and from 12.4% in yeast to 14.2% in *A. gossypii* in YPD (Fig. 4A). Many of the differences between the two species can be explained by reactions that are associated with a single gene in iRL766, whereas they are associated with two or three different genes in iTO977 (e.g., IMD3-IMD4). Essential genes for the two models were grouped according to their metabolic function (Fig. 4B, Supplementary file 7). Many essential genes were found in lipid metabolism and oxidative phosphorylation, followed by sugar and nucleotide metabolism. In MM also many genes in amino acid metabolism were found to be essential, while the presence of amino acids in YPD allowed auxotrophs to scavenge amino acids from the growth medium. There were no major differences in the spread of essential genes among the different metabolic pathways in the two models, indicating that paralogs in *S. cerevisiae* are evenly spread out over all metabolic functions.

Table 1. Carbon and nitrogen utilization.

	Predicted value	Experimental value	Refs.
Carbon source			
Acetate	o	o	Mickelson (1950)
Cellobiose	oo	o	Kurtzman et al. (2011)
Cellulose	—	—	Farries and Bell (1930); Ribeiro et al. (2012)
Citrate	o	o	Kurtzman et al. (2011)
D arabinose	—	—	Kurtzman et al. (2011)
D glucosamine	—	—	Kurtzman et al. (2011)
D mannitol	—	—	Kurtzman et al. (2011)
D ribose	oo	—	Kurtzman et al. (2011)
D xylose	oo	—	Kurtzman et al. (2011); Ribeiro et al. (2012)
D sorbitol	oo	o	Kurtzman et al. (2011)
D gluconate	o	o	Kurtzman et al. (2011)
DL lactate	—	—	Kurtzman et al. (2011)
Erythritol	—	—	Kurtzman et al. (2011)
Ethanol	oo	o	Kurtzman et al. (2011); Mickelson (1950)
Fructose	oo	oo	Farries and Bell (1930)
Galactitol	—	—	Kurtzman et al. (2011)
Galactose	—	—	Kurtzman et al. (2011)
Glucose	oo	oo	Kurtzman et al. (2011); Ribeiro et al. (2012)
Glycerol	oo	oo	Kurtzman et al. (2011); Ribeiro et al. (2012)
Hexadecane	—	—	Kurtzman et al. (2011)
Inulin	—	—	Kurtzman et al. (2011)
L arabinose	—	—	Kurtzman et al. (2011); Ribeiro et al. (2012)
L rhamnose	—	—	Kurtzman et al. (2011)
L sorbose	—	—	Kurtzman et al. (2011)
Lactose	—	—	Farries and Bell (1930); Kurtzman et al. (2011); Mickelson (1950)
Maltose	oo	o	Kurtzman et al. (2011)
Melezitose	—	—	Kurtzman et al. (2011)
Melibiose	—	—	Kurtzman et al. (2011)
Methanol	—	—	Kurtzman et al. (2011)
Methyl-alpha-D glucoside	—	—	Kurtzman et al. (2011)
Myoinositol	—	—	Kurtzman et al. (2011)
NADglucosamine	—	—	Kurtzman et al. (2011)
Pyruvate	o	o	Mickelson (1950)
Raffinose	o	oo	Kurtzman et al. (2011)
Ribitol	—	—	Kurtzman et al. (2011)
Starch	oo	oo	Ribeiro et al. (2012)
Succinate	o	o	Kurtzman et al. (2011)
Sucrose	oo	oo	Kurtzman et al. (2011)
Trehalose	oo	oo	Kurtzman et al. (2011)
Nitrogen source			
NH ₄	oo	oo	Ribeiro et al. (2012)
Nitrate	—	—	Kurtzman et al. (2011); Ribeiro et al. (2012)

Comparison of *in silico* (predicted) and *in vivo* (experimental) growth on different carbon and nitrogen sources. — represents no growth; o represents reduced growth; while oo represents unimpaired growth, compared to growth on glucose (Supplementary file 6).

Integrative Data Analysis

Another powerful use of GEMs is as a scaffold for the analysis of data from different biological levels. Focusing on riboflavin production, we analyzed previously published microarray data (Gattiker et al., 2007) from trophic stage and riboflavin productive stage. Even though these microarray data have extensively been investigated for cell development (Rischatsch, 2007), the information they carry about riboflavin production has remained unexplored.

Reporter Metabolites

Patterns in the transcriptional responses of the metabolic network can be revealed by using the network topology from

a GEM (Patil and Nielsen, 2005). We therefore used iRL766 to group genes into sets all coding for enzymes that catalyze reactions where a specific metabolite takes part. Enriched or significant gene sets (i.e., reporter metabolites) can be interpreted as metabolic hotspots or metabolites around which important transcriptional changes occur (Supplementary file 8). Among the most significant reporter metabolites were primarily amino acids, likely a consequence of the protein turnover that is observed between the two growth stages (Rischatsch, 2007). While riboflavin was among these top metabolites, its precursor cytosolic glycine appeared as the most reliable reporter metabolite. Glycine has been described as a key factor in riboflavin production and the increase of this, either by metabolic engineering approaches or media supplementation leads to riboflavin

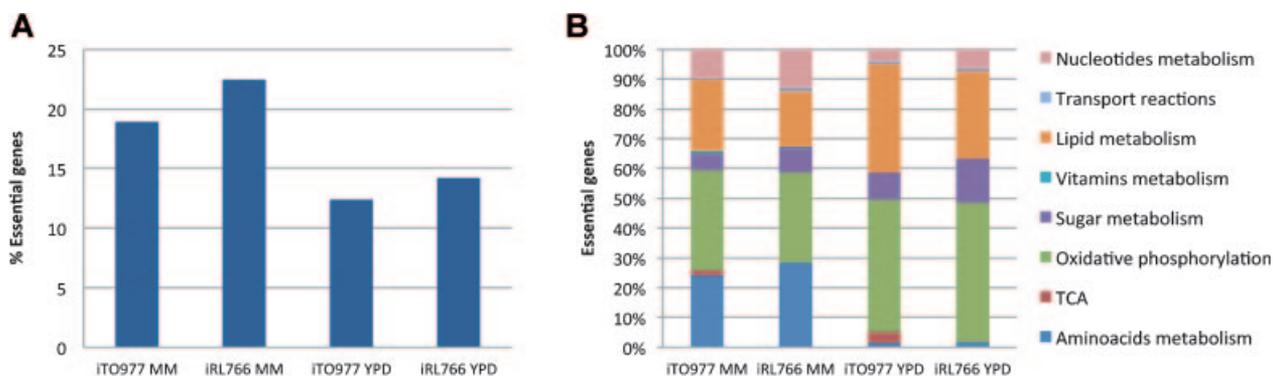


Figure 4. Gene essentiality was investigated in both models iRL766 and iTO977 by knocking out individual genes and assessing their growth rate. Whole models comparison where (A) the fraction of essential genes over all genes in the model, and (B) metabolic groups of these essential genes are shown. MM, minimal media; YPD, rich media.

overproduction (Monschau et al., 1998; Schlupen et al., 2003). Glyoxylate and the vast presence of fatty acids derivatives as reporter metabolites supported the proposed mechanism for riboflavin production through beta-oxidation of lipid bodies and the glyoxylate cycle. Also other intermediates in riboflavin biosynthesis appeared at high positions in the up regulated reporter metabolite list such as 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine, 6,7-dimethyl-8-(1-D-ribityl)lumazine or 3,4-dihydroxy-2-butanone 4-phosphate.

Identification of Transcriptionally Regulated Enzymes

Genetic engineering of *A. gossypii* has allowed to increase the production of riboflavin to an industrial level (Kato and Park, 2012). Enzymes that correlate in expression levels and metabolic fluxes are potential targets for overexpression, in an attempt to direct the metabolic flux towards a desired product. We identified what changes in metabolic fluxes in response to the riboflavin production phase are transcriptionally regulated. The transcriptomics dataset is described above.

For the metabolic fluxes, a space of feasible flux distributions was defined by constraining the model by a small set of experimental metabolic fluxes and appropriate objective functions for each condition. A collection of possible flux distributions was calculated using a random sampling algorithm (Bordel et al., 2010). Flux distributions from the two conditions were compared and probability scores were calculated to identify the significance of the observed changes. The probability score of the change in metabolic flux was compared to the probability score of a transcriptional change, obtained from the microarray dataset. The reactions with a probability score higher than 0.9, as calculated by the random sampling algorithm (Bordel et al., 2010), present both changes in the same direction. Statistics for both up- and down-regulated reactions were calculated (Supplementary file 9).

In total, 38 reactions demonstrated correlated increases in both transcript levels and metabolic fluxes; these reactions were transcriptionally up-regulated. Four of those reactions were part of the riboflavin synthetic pathway: RIB2, RIB3, RIB4, and RIB5 (Fig. 5). Other up-regulated reactions were involved in beta-oxidation (e.g., POX and FOX2) and gluconeogenesis (FBP1, PCK1), supporting the hypothesis of lipid bodies consumption in the riboflavin productive phase. Almost all other up-regulated reactions were involved in extracellular uptake of amino acids and nucleosides. *A. gossypii* can suffer autolysis in the latest phases of its life cycle: some cells can lyse in order to supply nutrients to the remaining cells which continue their riboflavin overproduction and sporulation (Stahmann et al., 2000). This explains the up-regulation of amino acid and nucleoside permeases in this growth phase. All the genes involved in the reactions identified as up-regulated are potential candidates for overexpression in order to increase riboflavin production. While some, such as the RIB genes, are more obvious, especially the genes of beta-oxidation and gluconeogenesis

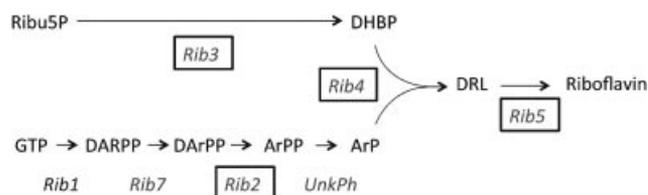


Figure 5. Transcriptionally controlled reactions (reactions where the change in flux correlate with change in expression of the involved gene) in riboflavin synthesis. Gene in squares are responsible for an up-regulated reaction identified by the random sampling algorithm when comparing trophic phase and riboflavin productive phase. UnkPh (unknown phosphatase), Rib5P (ribulose 5 phosphate), DHBP (3,4-dihydroxy-2-butanone 4-phosphate), DARPP (2,5-diamino-6-hydroxy-4-(5'-phosphoribosylamino)-pyrimidine), DArPP (2,5-diamino-6-ribitylamino-4(3H)-pyrimidinone 5'-phosphate), ArPP (5-amino-6-(5'-phosphoribitylamino)uracil), ArP (4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine), DRL (6,7-dimethyl-8-(1-D-ribityl)lumazine).

would require further investigation to elucidate whether the sole overexpression of these genes lead to some increase in the vitamin titer and experimental data must be done to confirm this.

A larger number of reactions, 127, were transcriptionally down-regulated and many of those are related to biomass formation. Interestingly, many reactions from purine biosynthesis were down-regulated, despite this pathway's involvement in the supply of GTP for riboflavin production. This has been previously described for some of the genes of purine biosynthesis, such as ADE4 (Mateos et al., 2006), indicating that either the nucleotides have been synthesized in excess during the trophic phase or the nucleotides/nucleosides from partially broken mycelia are taken up from the environment using the up-regulated permeases. Additionally, down-regulation also prevents riboflavin consumption by riboflavin kinase (FMN1) and glycine degradation by SHM2 and GCV. Indeed the disruption of SHM2 in *A. gossypii* leads to an accumulation of riboflavin (Schlupen et al., 2003) and both SHM2 and ADE4 have been identified as down-regulated by the transcription factor BAS1 (Mateos et al., 2006).

Conclusions

In this study, we reconstructed and validated the first manually curated genome-scale model of *A. gossypii*, iRL766, with particular focus on the riboflavin production by this fungus. The model is in good concordance with existing experimental data on the pathways involved in riboflavin production, growth rate and the use of different carbon sources. Subsequently, the curated GEM was used to analyze gene essentiality, particularly in comparison to the *S. cerevisiae* GEM iTO977. The increased number of predicted essential genes in iRL766 could be transcribed to the whole genome duplication event that has occurred in *S. cerevisiae* but not in *A. gossypii*. Categorizing the predicted essential genes in metabolic pathways indicated that this event has led to an even spread of paralogs among all metabolic pathways. This is in agreement with the proposed rationale of gene duplication, where duplicated genes do not belong to any particular dominated function and do not occur more frequently in essential genes, but often overlap functional roles (Kuepfer et al., 2005).

Subsequently, the model was used as a scaffold to analyze microarray data to investigate the transcriptional changes related to riboflavin production. Reporter metabolite analysis indicated glycine, glyoxylate, and fatty acids as important metabolic hot spots, together with the anticipated riboflavin. A random sampling algorithm was employed to identify transcriptionally regulated reactions in the GEM. Together with the anticipated RIB genes, also beta oxidation and gluconeogenesis were transcriptionally up-regulated. Additionally observed important reporter metabolites and transcriptionally regulated reactions were supported by previous publications. This underwrites the validity of the model and indicates how it can be used in future approaches not only for overproduction of riboflavin, but also other

vitamins, proteins, or lipids, as well as it can be used to improve our knowledge of this organism.

R. L.-A. was recipient of an FPU predoctoral fellowship from the *Ministerio de Educación, Cultura y Deporte* and an EMBO short-term fellowship.

References

- Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. 2013a. The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput Biol* 9(3):e1002980.
- Agren R, Otero JM, Nielsen J. 2013b. Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production. *J Ind Microbiol Biotechnol* 40(7):735–747.
- Bordel S, Agren R, Nielsen J. 2010. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput Biol* 6(7):e1000859.
- Brachat S, Dietrich FS, Voegeli S, Zhang Z, Stuart L, Lerch A, Gates K, Gaffney T, Philippsen P. 2003. Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*. *Genome Biol* 4(7):R45.
- Caspeta L, Nielsen J. 2013. Toward systems metabolic engineering of *Aspergillus* and *Pichia* species for the production of chemicals and biofuels. *Biotechnol J* 8(5):534–544.
- Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 40(Database issue):D742–D753.
- Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, et al. 2012. *Saccharomyces* genome database: The genomics resource of budding yeast. *Nucleic Acids Res* 40(Database issue):D700–D705.
- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, et al. 2010. The genetic landscape of a cell. *Science* 327(5964):425–431.
- Cvijovic M, Olivares-Hernandez R, Agren R, Dahr N, Vongsangnak W, Nookaew I, Patil KR, Nielsen J. 2010. BioMet Toolbox: Genome-wide analysis of metabolism. *Nucleic Acids Res* 38(Web Server issue):W144–W149.
- Demain AL. 1972. Riboflavin oversynthesis. *Annu Rev Microbiol* 26:369–388.
- Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, et al. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 304(5668):304–307.
- Farries EHM, Bell AF. 1930. On the metabolism of *Nematospira gossypii* and related fungi, with special reference to the source of nitrogen. *Ann Bot* 44:423–455.
- Finlayson MR, Helfer-Hungerbuehler AK, Philippsen P. 2011. Regulation of exit from mitosis in multinucleate *Ashbya gossypii* cells relies on a minimal network of genes. *Mol Biol Cell* 22(17):3081–3093.
- Garcia-Albornoz MA, Nielsen J. 2013. Application of genome-scale metabolic models in metabolic engineering *Indus Biotechnol* 9(4):203–214.
- Gattiker A, Rischatsch R, Demougin P, Voegeli S, Dietrich FS, Philippsen P, Primig M. 2007. *Ashbya* Genome Database 3.0: A cross-species genome and transcriptome browser for yeast biologists. *BMC Genom* 8:9.
- Hermida L, Brachat S, Voegeli S, Philippsen P, Primig M. 2005. The *Ashbya* Genome Database (AGD)—A tool for the yeast community and genome biologists. *Nucleic Acids Res* 33(Database issue):D348–D352.
- Ibarra RU, Edwards JS, Palsson BO. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420(6912):186–189.
- Jimenez A, Santos MA, Pompejus M, Revuelta JL. 2005. Metabolic engineering of the purine pathway for riboflavin production in *Ashbya gossypii*. *Appl Environ Microbiol* 71(10):5743–5751.

- Jimenez A, Santos MA, Revuelta JL. 2008. Phosphoribosyl pyrophosphate synthetase activity affects growth and riboflavin production in *Ashbya gossypii*. *BMC Biotechnol* 8:67.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(Database issue):D109–D114.
- Kato T, Park EY. 2012. Riboflavin production by *Ashbya gossypii*. *Biotechnol Lett* 34(4):611–618.
- Kim TY, Sohn SB, Kim YB, Kim WJ, Lee SY. 2012. Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr Opin Biotechnol* 23(4):617–623.
- Kuepfer L, Sauer U, Blank LM. 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res* 15(10):1421–1430.
- Kurtzman CP, Fell JW, Boekhout T. 2011. The yeast: A taxonomic study. 5th edn. Elsevier, Amsterdam Volume 2, Part IVb.
- Ledesma-Amaro R, Jimenez A, Santos MA, Revuelta JL. 2013a. Biotechnological production of feed nucleotides by microbial strain improvement. *Process Biochem* 48(9):1263–1270.
- Ledesma-Amaro R, Jimenez A, Santos MA, Revuelta JL. 2013b. Microbial production of vitamins. Book chapter in: *Microbial production of food ingredients, enzymes and nutraceuticals*, Sawston, Cambridge, UK: Woodhead Publishing Series in Food Science. Technology and Nutrition No. 246.
- Ledesma-Amaro R, Santos MA, Jiménez A, Revuelta JL. 2013c. Strain design of *Ashbya gossypii* for single-cell oil production. *Applied and Environmental Microbiology*, doi: 10.1128/aem.03560-13. *AEM.03560-13 published ahead of print 6 December 2013*.
- Lim SH, Ming H, Park EY, Choi JS. 2003. Improvement of riboflavin production using mineral support in the culture of *Ashbya gossypii*. *Food Technol Biotechnol* 41(2):137–144.
- Mateos L, Jimenez A, Revuelta JL, Santos MA. 2006. Purine biosynthesis, riboflavin production, and trophic-phase span are controlled by a Myb-related transcription factor in the fungus *Ashbya gossypii*. *Appl Environ Microbiol* 72(7):5052–5060.
- Mickelson MN. 1950. The metabolism of glucose by *Ashbya gossypii*. *J Bacteriol* 59(5):659–666.
- Monschau N, Sahn H, Stahmann K. 1998. Threonine aldolase overexpression plus threonine supplementation enhanced riboflavin production in *Ashbya gossypii*. *Appl Environ Microbiol* 64(11):4283–4290.
- Osterlund T, Nookaew I, Bordel S, Nielsen J. 2013. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol* 7:36.
- Otero JM, Cimini D, Patil KR, Poulsen SG, Olsson L, Nielsen J. 2013. Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. *PLoS ONE* 8(1):e51414.
- Park EY, Ito Y, Nariyama M, Sugimoto T, Lies D, Kato T. 2011. The improvement of riboflavin production in *Ashbya gossypii* via disparity mutagenesis and DNA microarray analysis. *Appl Microbiol Biotechnol* 91(5):1315–1326.
- Park EY, Zhang JH, Tajima S, Dwiarti L. 2007a. Isolation of *Ashbya gossypii* mutant for an improved riboflavin production targeting for biorefinery technology. *J Appl Microbiol* 103(2):468–476.
- Park JH, Lee KH, Kim TY, Lee SY. 2007b. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc Natl Acad Sci U S A* 104(19):7797–7802.
- Park JM, Kim TY, Lee SY. 2009. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnol Adv* 27(6):979–988.
- Patil KR, Nielsen J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* 102(8):2685–2689.
- Ribeiro O, Domingues L, Penttila M, Wiebe MG. 2012. Nutritional requirements and strain heterogeneity in *Ashbya gossypii*. *J Basic Microbiol* 52(5):582–589.
- Ribeiro O, Magalhaes F, Aguiar TQ, Wiebe MG, Penttila M, Domingues L. 2013. Random and direct mutagenesis to enhance protein secretion in *Ashbya gossypii*. *Bioengineered* 4(5):322–331.
- Ribeiro O, Wiebe M, Ilmen M, Domingues L, Penttila M. 2010. Expression of *Trichoderma reesei* cellulases CBHI and EGI in *Ashbya gossypii*. *Appl Microbiol Biotechnol* 87(4):1437–1446.
- Rischatsch R. 2007. Transcriptome profiling of the model organism *A. gossypii*: Comparison of life cycle stages and transcription factor deletions. PhD thesis.
- Scalinati G, Otero JM, Van Vleet JR, Jeffries TW, Olsson L, Nielsen J. 2012. Evolutionary engineering of *Saccharomyces cerevisiae* for efficient aerobic xylose consumption. *FEMS Yeast Res* 12(5):582–597.
- Schlupen C, Santos MA, Weber U, de Graaf A, Revuelta JL, Stahmann KP. 2003. Disruption of the SHM2 gene, encoding one of two serine hydroxymethyltransferase isoenzymes, reduces the flux from glycine to serine in *Ashbya gossypii*. *Biochem J* 369(Pt 2):263–273.
- Schmitz HP, Philippsen P. 2011. Evolution of multinucleated *Ashbya gossypii* hyphae from a budding yeast-like ancestor. *Fungal Biol* 115(6):557–568.
- Shi S, Chen T, Zhang Z, Chen X, Zhao X. 2009. Transcriptome analysis guided metabolic engineering of *Bacillus subtilis* for riboflavin production. *Metab Eng* 11(4–5):243–252.
- Stahmann KP, Arst HN, Jr., Althofer H, Revuelta JL, Monschau N, Schlupen C, Gatgens C, Wiesenburg A, Schlosser T. 2001. Riboflavin, overproduced during sporulation of *Ashbya gossypii*, protects its hyaline spores against ultraviolet light. *Environ Microbiol* 3(9):545–550.
- Stahmann KP, Kupp C, Feldmann SD, Sahn H. 1994. Formation and degradation of lipid bodies found in the riboflavin-producing fungus *Ashbya gossypii*. *Appl Microbiol Biotechnol* 42(1):121–127.
- Stahmann KP, Revuelta JL, Seulberger H. 2000. Three biotechnical processes using *Ashbya gossypii*, *Candida famata*, or *Bacillus subtilis* compete with chemical riboflavin production. *Appl Microbiol Biotechnol* 53(5):509–516.
- Sugimoto T, Kanamasa S, Kato T, Park EY. 2009. Importance of malate synthase in the glyoxylate cycle of *Ashbya gossypii* for the efficient production of riboflavin. *Appl Microbiol Biotechnol* 83(3):529–539.
- van Zyl C, Prior BA, Kilian SG, Kock JL. 1989. D-xylose utilization by *Saccharomyces cerevisiae*. *J Gen Microbiol* 135(11):2791–2798.
- Varemo L, Nielsen J, Nookaew I. 2013. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res* 41(8):4378–4391.
- Vongsangnak W, Olsen P, Hansen K, Krogsgaard S, Nielsen J. 2008. Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. *BMC Genom* 9:245.
- Wendland J, Ayad-Durieux Y, Knechtel P, Rebischung C, Philippsen P. 2000. PCR-based gene targeting in the filamentous fungus *Ashbya gossypii*. *Gene* 242(1–2):381–391.
- Wendland J, Dunkler A, Walther A. 2011. Characterization of alpha-factor pheromone and pheromone receptor genes of *Ashbya gossypii*. *FEMS Yeast Res* 11(5):418–429.
- Wendland J, Walther A. 2005. *Ashbya gossypii*: A model for fungal developmental biology. *Nature reviews. Microbiology* 3(5):421–429.
- Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387(6634):708–713.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.