FEMS Yeast Research, 17, 2017, fox050

doi: 10.1093/femsyr/fox050
Advance Access Publication Date: 10 July 2017
Minireview

MINIREVIEW

Genome-scale modeling of yeast: chronology, applications and critical perspectives

Helder Lopes[†] and Isabel Rocha[*]

CEB - Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal

[*]**Corresponding author:** CEB - Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal. Tel: +351 253 604 414 (Office);
E-mail: irocha@deb.uminho.pt

One sentence summary: Genome-scale metabolic models reconstructed over time for several yeasts and their application in the development of cell factories for diverse biotechnological products.
Editor: Zongbao Zhao

[†]Helder Lopes, http://orcid.org/0000-0001-9563-3844

ABSTRACT

Over the last 15 years, several genome-scale metabolic models (GSMMs) were developed for different yeast species, aiding both the elucidation of new biological processes and the shift toward a bio-based economy, through the design of *in silico* inspired cell factories. Here, an historical perspective of the GSMMs built over time for several yeast species is presented and the main inheritance patterns among the metabolic reconstructions are highlighted. We additionally provide a critical perspective on the overall genome-scale modeling procedure, underlining incomplete model validation and evaluation approaches and the quest for the integration of regulatory and kinetic information into yeast GSMMs. A summary of experimentally validated model-based metabolic engineering applications of yeast species is further emphasized, while the main challenges and future perspectives for the field are finally addressed.

Keywords: yeast; genome-scale metabolic model; metabolism; constraint-based modeling; metabolic engineering; cell factory

INTRODUCTION

The availability of low-cost whole-genome sequencing techniques led to an explosion of data for several organisms. This, alongside the advent of organism-specific omics data, advanced bioinformatics tools, and an increasing computational performance has paved the way to the reconstruction of metabolic networks at the genome scale.

Genome-wide reconstructions of the cell metabolism can be converted into predictive constraint-based models, establishing a complex network of biochemical reactions with information on stoichiometry, compartmentalization, biomass composition, thermodynamics and genes responsible for each reaction (Covert *et al.* 2001; Thiele and Palsson 2010). When combined with constraint-based algorithms, genome-scale metabolic models (GSMMs, also known as GEMs) offer an excellent opportunity for studying metabolism and genotype–phenotype relationships (O'Brien, Monk and Palsson 2015).

Hence, GSMMs have become a key framework in the systems biology field, in particular, for systems metabolic engineering (ME) approaches. After the first GSMM was published nearly 20 years ago (Edwards and Palsson 1999), many others have followed. Hitherto, there are GSMMs published and accessible for download in several websites for more than 100 organisms (e.g. www.optflux.org/models and http://systemsbiology.ucsd.edu/InSilicoOrganisms/OtherOrganisms), and the number keeps rising.

The yeast *Saccharomyces cerevisiae* was the first eukaryotic organism to be fully genome sequenced (Goffeau *et al.* 1996; Cherry *et al.* 1997), and it has been one of the workhorses in cell factory engineering for biotechnological production of several

Received: 20 April 2017; **Accepted:** 7 July 2017
© FEMS 2017. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

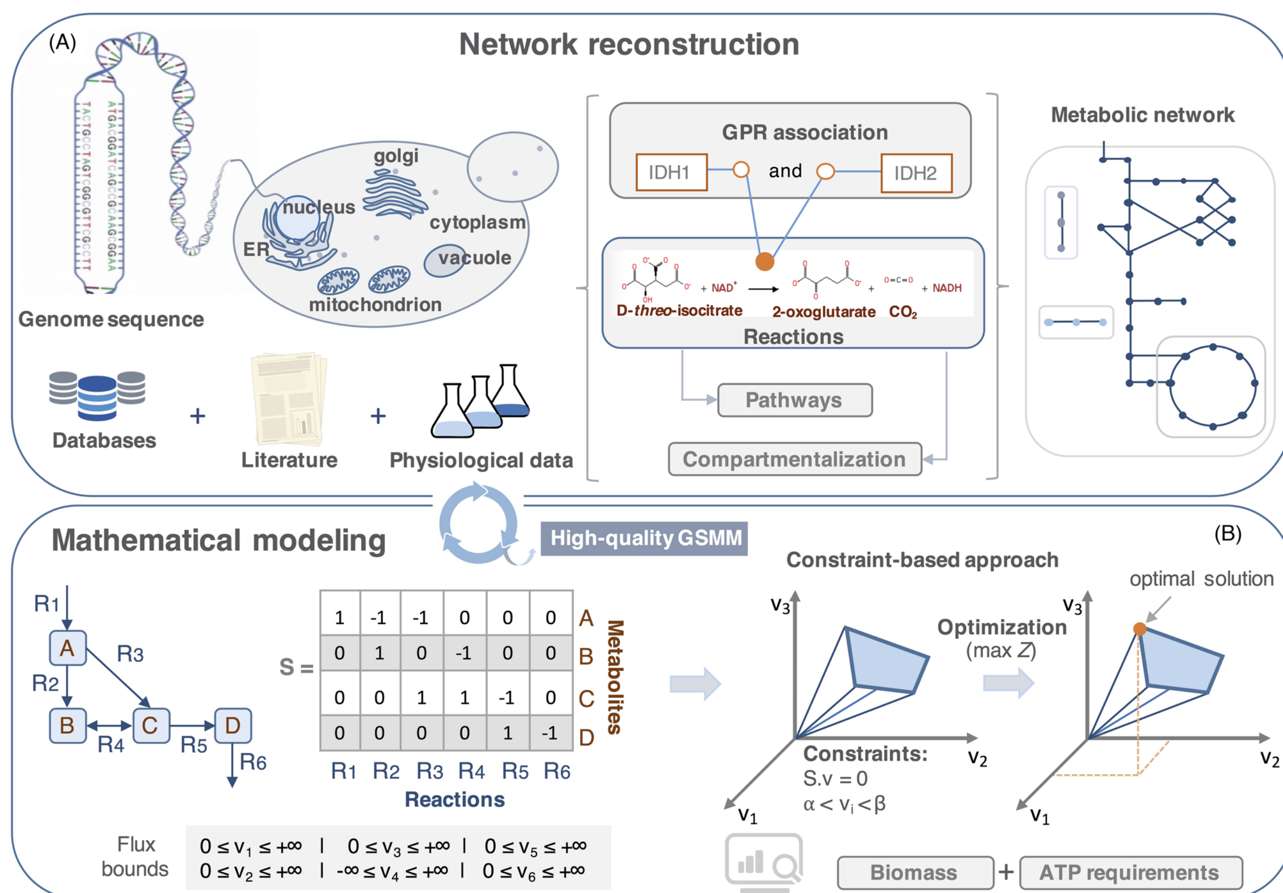

Figure 1. Metabolic network reconstruction and mathematical modeling of genome-scale networks. (A) A draft metabolic network can be generated using genomic, biochemical and physiological information available in primary literature or proper databases. All annotated metabolic genes are first matched to enzymes and then to the reactions—composed by different metabolites and cofactors—to obtain GPR associations. Reactions are assembled into pathways which together constitute the metabolite network. Localization has also to be considered since chemically identical metabolites may be present in different cellular compartments. (B) The reconstructed genome-scale metabolic network is then transformed into a constraint-based model, by first converting it to a mathematical representation using a stoichiometric matrix (S) of the metabolite coefficients in each reaction, and further assuming pseudo-steady state and constraining the reaction flux (v) bounds. The system of linear equations defines the admissible flux space of solutions (known as flux cone) and using an objective function defining an optimization problem it is possible to find optimal solutions for a desired output. To simulate model growth and obtain meaningful flux distributions, information on biomass composition and ATP requirements of the cell must also be available. The generation of high-end genome-scale metabolic models often requires several cycles of testing and refinement based on the comparative results of *in silico* simulations and experimental data.

compounds with widespread applications in food (Brochado *et al.* 2010; Li *et al.* 2013a), chemical (Hong and Nielsen 2012; Nielsen *et al.* 2013) and pharmaceutical industries (Paddon *et al.* 2013). Given the similarity and high number of features conserved with the human functions, it is also a role model for diseases, drug screening and fundamental biology studies (Sturgeon *et al.* 2006; Petranovic and Nielsen 2008). So, it is not a surprise that *S. cerevisiae* has been the first eukaryotic organism to have a GSMM (Förster *et al.* 2003), and is top-ranked if we count the number of available GSMMs per single organism. However, other yeast species constitute important human pathogens or have also proven to be suitable platforms for biotechnological applications and several models have been therefore reconstructed for different yeasts.

Here, we review the genome-scale modeling process in yeast, presenting an historical perspective of the GSMMs built along the time for different yeast species beyond the well-characterized *S. cerevisiae* through the representation of a chronological network containing the inherited features of several yeast models. We then present a critical perspective on the overall genome-scale modeling procedure in yeast, from incomplete model validation and evaluation approaches to the increasing pursuit for the integration of regulatory and kinetic information into metabolic networks. A summary of the main applications of yeasts' GSMMs in cell factory development is further addressed. Lastly, the future perspectives in the field are discussed.

METABOLIC NETWORK RECONSTRUCTION AND MATHEMATICAL MODELING

The reconstruction and mathematical modeling processes of genome-scale metabolic networks have been extensively described and reviewed elsewhere (Feist *et al.* 2009; Oberhardt, Palsson and Papin 2009; Thiele and Palsson 2010; O'Brien, Monk and Palsson 2015). Here, we briefly recapitulate the main steps of this systematic process through a schematic representation presented in Fig. 1, as a prelude to contextualize the topics discussed in the further sections.

The first requisite to start the reconstruction of a metabolic network is to have the genome sequence of the organism of

interest. If once this could be an issue, nowadays with the emergence of next-generation sequencing techniques, it is possible to obtain an organism's genome sequence overnight. Even for poorly studied organisms, we can easily generate draft models from a genome sequence of the organism using homology searching algorithms and semi-automated reconstruction tools, such as Model SEED, RAVEN or *merlin* (Henry *et al.* 2010; Agren *et al.* 2013; Dias *et al.* 2015). For example, over 2600 draft GSMMs for more than 1500 organisms across different phylogenetic domains were automatically generated through the Path2models project and, more recently, 773 human gut bacteria genome-scale reconstructions were generated through AGORA using metagenomics data (Büchel *et al.* 2013; Magnúsdóttir *et al.* 2016). However, the implementation of robust and high-quality GSMMs able to predict cellular phenotypes with reasonable accuracy requires additional time and efforts, since proficient manual curation and several iteration cycles of testing and refinement are necessary. The reconstruction process of a robust GSMM usually depends on having a very well-annotated genome and, consequently, reliable gene-protein-reaction (GPR) relationships, as well as information about the stoichiometric coefficients of substrates and products present in each reaction, cofactor information, reaction directionality and compartmentalization. Furthermore, to simulate microbial growth and obtain meaningful flux distributions, one must have experimental evidences on the biomass composition and an estimation of growth and non-growth-associated ATP requirements. All this information is usually collected from different sources ranging from primary literature to high-throughput data and organism-specific databases, if available.

To perform *in silico* simulations, the reconstructed metabolic network has first to be converted into a mathematical model which follows a matrix representation of the stoichiometric coefficients of each reaction. Assuming the steady-state behavior of internal metabolism, i.e. ensuring, for each metabolite, that the total rates of consumption and production are equal, and applying some flux constraints, such as reaction flux bounds to narrow down the space of feasible computational solutions, it is possible to evaluate the biological capabilities of an organism. Furthermore, one can use the well-known flux balance analysis (FBA) method, which consists in setting an objective function for maximizing or minimizing a subset of fluxes and finding optimal solutions by solving the resulting linear programming problems. A typical example is the maximization of the biomass objective function to simulate growth-focused cell behavior (Savinell and Palsson 1992; Orth, Thiele and Palsson 2010). This type of modeling procedure is known as constraint-based modeling (CBM), which has been frequently applied in ME projects (Long, Ong and Reed 2015). Besides FBA, several other variants have been developed within the CBM community, such as minimization of metabolic adjustment (MOMA), regulatory on/off minimization (ROOM) and RELATCH (Segrè, Vitkup and Church 2002; Shlomi, Berkman and Ruppin 2005; Kim and Reed 2012). Since extensive reviews of CBM methods have been provided, we will not describe or detail all the available methods here (see Park, Kim and Lee 2009; Senger *et al.* 2015; Maia, Rocha and Rocha 2016).

Before being published, GSMMs must be validated against experimental phenotypical evidences. There are several metrics commonly used to evaluate the predictive accuracy of these metabolic networks, including growth metrics and gene deletion metrics (single and double knockout analysis). Recently, Sánchez and Nielsen (2015) have thoroughly described which GSMMs of *S. cerevisiae* included these type of metrics in their publications. When the metabolic network is poorly connected or the simulation results differ from the experimental ones, the model has to be fine-tuned and continuously improved using, for example, gap-filling methods or by integrating omics data (Green and Karp 2004; Satish Kumar, Dasika and Maranas 2007; Sánchez and Nielsen 2015).

YEAST METABOLIC MODELS: CHRONOLOGICAL OVERVIEW

The yeast *Saccharomyces cerevisiae* was the pioneer organism on constraint-based genome wide modeling of eukaryotes. In this review, we update the history of yeast genome-scale modeling by revisiting the main published genome-scale models of *S. cerevisiae* and beyond. To support this discussion, Fig. 2 portraits a historical timeline highlighting the inherited features of new reconstructions over time, while Fig. 3 provides a quantitative assessment of the available GSMMs.

Nearly 15 years ago, Förster *et al.* (2003) built the first genome-scale model of *S. cerevisiae* named iFF708, a model containing 619 metabolic genes and 1172 reactions compartmentalized between cytosol, mitochondria and the extracellular space. Afterwards, three GSMMs of the same organism were derived from iFF708, namely iND750, iLL672 and iIN800. iND750 was the second genome-wide model of *S. cerevisiae* to be published introducing five additional cellular compartments, GPR associations and comprehensive proton balance (Duarte, Herrgård and Palsson 2004); iLL672 emerged afterwards with an improved connectivity of the network by deleting many dead-end reactions (Kuepfer, Sauer and Blank 2005); and iIN800 has then included tRNA synthesis, transport processes and a more detailed lipid metabolism (Nookaew *et al.* 2008). Later, iMM904 arose as an improved version of the iND750 model, introducing a new nomenclature for metabolites and reactions and integrating exometabolomic data, which consequently led to enhanced essentiality predictions, according to their authors (Mo, Palsson and Herrgård 2009). This GSMM was subsequently revised by another group giving rise to iAZ900 (Zomorrodi and Maranas 2010). The authors of this study suggested 120 corrections to the iMM904 model, including changes in the GPR associations, reversibility of reactions and biomass composition, as well as adding/removing reactions, compounds or genes, through the use of an automated procedure for restoring consistency with single gene deletion and synthetic lethality data, which has led to improvements in terms of the model specificity. The first consensus genome-scale metabolic network (Yeast 1) was reconstructed through the collaboration of several research groups, as an attempt to build a consolidated metabolic network using standardized identifiers, although not capable of performing computational simulations (Herrgård *et al.* 2008). This fact was circumvented with Yeast 4, which also included increased metabolite transport, a better description of the lipid metabolism (based on iIN800) and improved pathway connectivity (Dobson *et al.* 2010). Thereafter, the consensus yeast model has experienced frequent updates by the community either to incorporate short (Yeast 5: Heavner *et al.* 2012) or major expansions of the lipid metabolism (Yeast 7: Aung, Henry and Walker 2013) or to improve the quantitative predictions of the model, particularly phenotypes related to essential and auxotroph-inducing genes (Yeast 6: Heavner *et al.* 2013). One of the last published *S. cerevisiae* GSMMs is the iTO977 that, similarly to yeast 4, resulted from the merge of the consensus network yeast 1 and the iIN800 model (Österlund *et al.* 2013). This model integrated transcriptomic data, thus

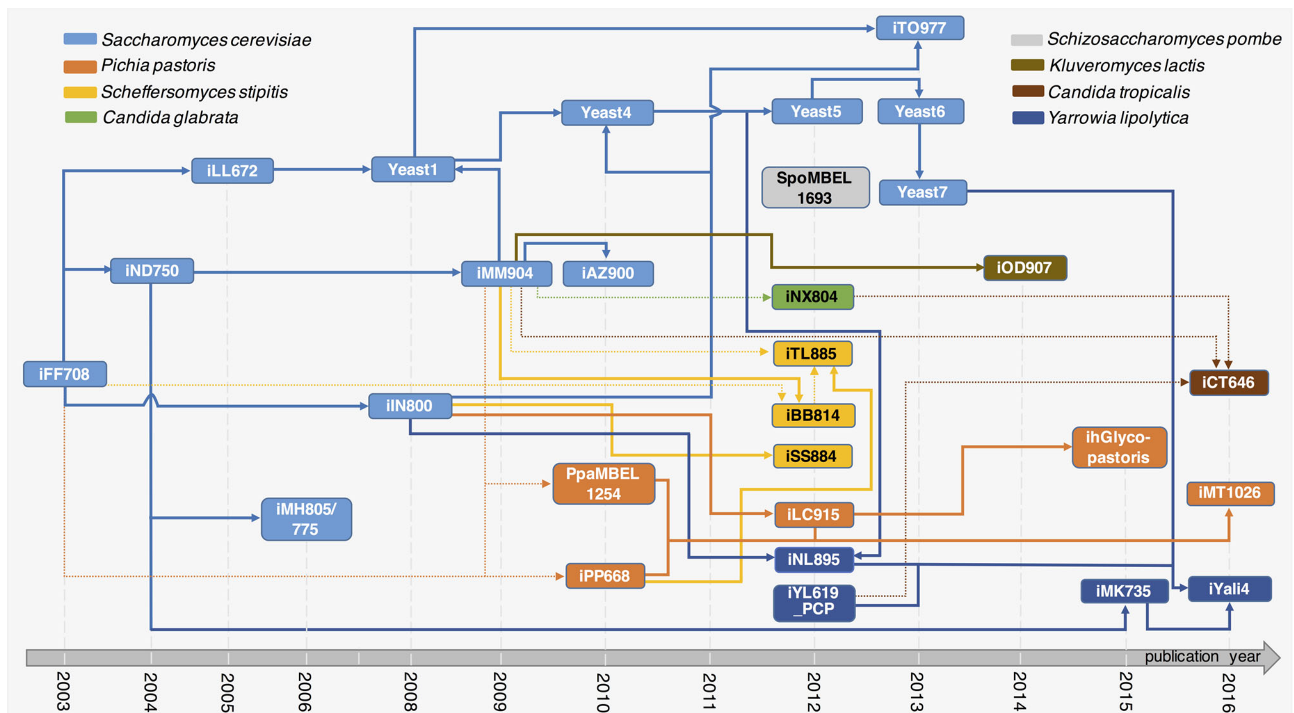

Figure 2. Evolutionary timeline of yeast GSMMs and their reconstruction inheritances. Each box contains the name of the metabolic model and is colored according to the respective yeast species color caption. Several GSMMs were reconstructed using previously available large-scale models as templates, from the same or different yeast species, which is represented in the figure through bold arrows connecting the respective boxes. The light-dashed colored lines represent the networks' relationship regarding the models that, although did not serve as structural scaffolds, have been used in the comparative/validation process of the subsequent GSMM.

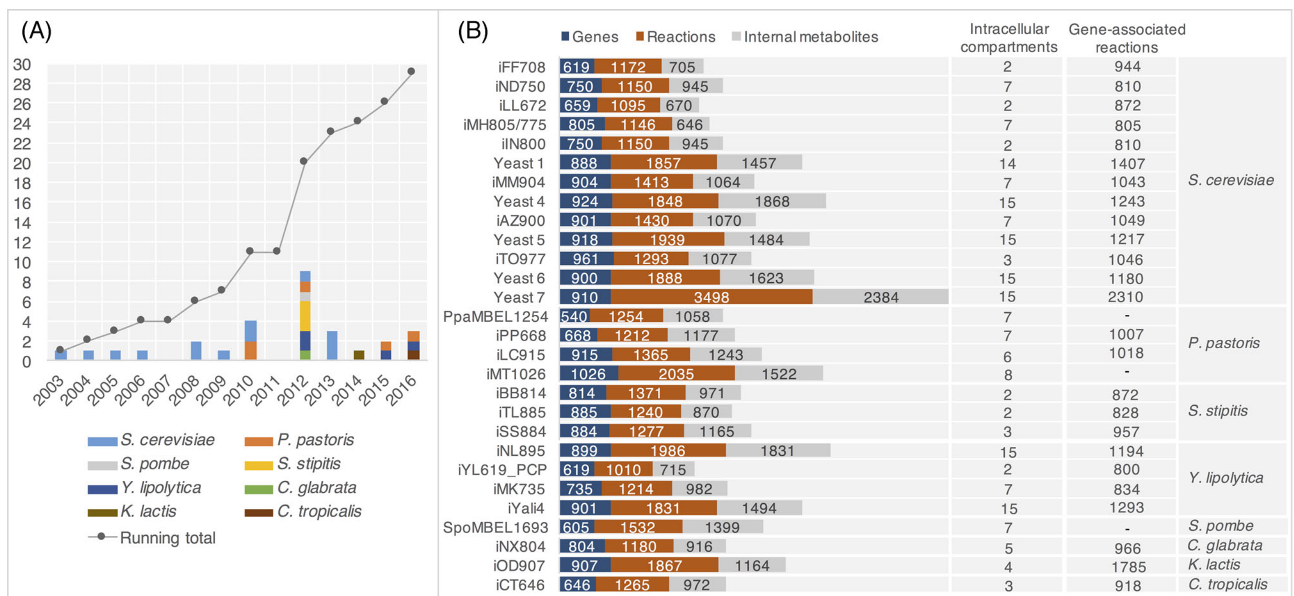

Figure 3. Genome-scale models of yeast in numbers. (**A**) Number of published GSMMs of yeast species over time. (**B**) Number of total genes, reactions (drains excluded), internal metabolites, intracellular compartments and reactions associated with genes of each GSMM. Inside each species categorization and if there is more than one GSMM for the same yeast, models are organized by date of publication (from top to down).

being able to identify transcriptionally controlled reactions and it is the one with highest gene coverage (Fig. 3B).

A big handicap of GSMMs is the absence of regulatory information in the GPR associations that could fully describe the physiological behavior of the cell in specific conditions. This absence is often used to justify inconsistencies observed in the simulations. Herrgård *et al.* (2006) added transcriptional regulatory constraints to an iND750-based GSMM to predict changes in the gene expression levels of some transcription factor deletion strains, which resulted in the iMH805/775 model, a model containing 82 nutrient signals and 55 transcription factors regulating 750 metabolic genes assembled from the primary literature. The increasing quest for adding a regulatory layer into GSMMs will be further discussed.

Saccharomyces cerevisiae is uniquely positioned among eukaryotic organisms to work as a robust, well-established and scalable cell factory. However, other yeast species have native traits and features that make them equally or even more adequate to produce certain products. Moreover, many yeast species represent important human pathogens. In that sense, lately, several efforts have been conducted to develop GSMMs for other yeast species. Figure 2 shows that *S. cerevisiae* GSMMs—particularly iMM904 and iIN800—have been frequently used as scaffolds for building or comparing new metabolic models reconstructed for other yeast species. Up until now, beyond the *S. cerevisiae* above-mentioned ones, 16 well-annotated GSMMs of other yeast species have been published (Fig. 3A). The yeast *Pichia pastoris* is considered one of the preferred host organism when it comes to the production of recombinant proteins (Werten *et al.* 1999; Damasceno *et al.* 2004). Hence, since 2010, it has been also a target organism in the metabolic modeling field for several research groups. The first two *P. pastoris* GSMMs corresponding to iPP668 and PpaMBEL1254 were almost simultaneously published, once genomic data were available for this yeast, being validated against physiological data from cultivations on different carbon sources (Chung *et al.* 2010; Sohn *et al.* 2010). Two years later, iLC915 was developed based on another sequenced genome of the same yeast (Caspeta *et al.* 2012). The model contains a broader genomic and metabolic coverage when compared to the previous ones, as well as a better agreement with experimental data with regard to the growth in different carbon sources, including methanol and glycerol. The same study has also resulted in one of the first GSMMs of the naturally occurring xylose-fermenting yeast *Scheffersomyces stipitis,* formerly known as *Pichia stipitis*, the iSS884 (Caspeta *et al.* 2012). Due to its native features, *S. stipitis* is a suitable candidate for xylose and pentose phosphate pathway metabolic studies, as well as for ethanol production from biomass. Both GSMMs used iIN800 as reference framework. Likewise, also in 2012, two additional genome-scale models of *S. stipitis* were published, namely iTL885 and iBB814 (Balagurunathan *et al.* 2012; Liu *et al.* 2012). iBB814 was reconstructed following a protocol for generating a high-quality GSMMs (Thiele and Palsson 2010) and compared with the first unicellular eukaryotic model (Förster *et al.* 2003) after a semi-automatic validation process, also including the experimental determination of the biomass macromolecular composition, while iTL885 used GSMMs of *S. cerevisiae* (iMM904) and *P. pastoris* (iPP668) as template frameworks to map the assigned genes to the list of original reactions (Liu *et al.* 2012), focusing on predictions related to xylose metabolism and xylose-derived ethanol production. No significant differences are observable among the three existing *S. stipitis* GSMMs in terms of model size (Fig. 3B). Back to *P. pastoris*, the iPP668 model was further merged with PpaMBEL1693 and iLC915 resulting in iMT1026, the most recent and comprehensive GSMM of this yeast, validated against a wider range of physiological data than the preexisting models (Tomàs-Gamisans, Ferrer and Albiol 2016). The iMT1026 model presents the highest genome coverage among all the GSMMs available for different yeast species (Fig. 3B) There is another *P. pastoris* GSMM available since 2015—ihGlycopastoris—which is an extension of the iLC915 model with native and humanized N-glycosylation, thus capable to simulate humanized glycosylation as well as estimate N-glycosylation of yeast native proteins, being considered the first functional GSMM of *P. pastoris,* with enhanced predictive capabilities in terms of protein yield (Irani *et al.* 2016).

The year 2012 was indeed a prolific year in terms of publications of different yeast GSMMs. Besides the already mentioned models for *S. cerevisiae*, *S. stipitis* and *P. pastoris* species, the—up to now—unique GSMMs of either *Schizosaccharomyces pombe* and *Candida glabrata* also became available that year, being named SpoMBEL1693 and iNX804, respectively (Sohn *et al.* 2012; Xu *et al.* 2013). *Schizosaccharomyces pombe*, also known as fission yeast, has been widely used as model system for studying the mammalian cell cycle control (Lee and Nurse 1988) and, like *C. glabrata*—known to be an important platform for pyruvate production—has been increasingly explored as a cell factory platform in biotechnological applications (Drăgan *et al.* 2011; Li *et al.* 2013b; Chen *et al.* 2015). Since *C. glabrata* is an opportunist human pathogen, the GSMM available for this yeast has also been used to predict potential drug targets for antimicrobial therapies. In turn, *Yarrowia lipolytica* is an oleaginous yeast that can accumulate large amounts of specialty lipids, making it of interest for biofuels and other chemicals production. The first two GSMMs of *Y. lipolytica* were also published in 2012: iNL895 emerged as the first well-annotated metabolic model of an oleaginous yeast, although derived from *S. cerevisiae* models (Loira *et al.* 2012), followed by iYL619_PCP which was built directly from knowledge bases with more specific information on the organism of interest (Pan and Hua 2012). More recently, another GSMM of *Y. lipolytica*, iMK735, has been published as an adapted version of the iND750 *S. cerevisiae* model (Kavšček *et al.* 2015). Information contained on the first two models of the oleaginous yeast was further used to build the most recent and comprehensive model of *Y. lipolytica*, iYali4, which used yeast 7.11 consensus network as template model to integrate multilevel omics data, and helped to demonstrate that lipid accumulation in this yeast is associated with regulation of amino acid biosynthesis and does not involve transcriptional regulation of lipid metabolism (Kerkhoven *et al.* 2016).

The yeast *Kluveromyces lactis* has also been used as host for the production of recombinant proteins, while *C. tropicalis* presents an interesting capacity for producing α,ω-dicarboxylic acids. Hence, these organisms have been attracting the attention of systems biologists, being now among the yeast species with a curated GSMM available. The first and unique GSMM of the milk yeast *K. lactis* (iOD907) was published in 2014, building upon the iMM904 *S. cerevisiae* model and fundamental literature for this organism (Dias *et al.* 2014), claiming reasonable predictive performance with regard to quantitative simulations of chemostat experiments and gene knockout phenotypes. iCT646 was reconstructed two years later through the assembly of genomic and biochemical information from databases and primary literature, thus allowing system wide analysis of *C. tropicalis* metabolic studies (Mishra *et al.* 2016).

Since today draft models can be easily derived by automatic application of reconstruction algorithms, we can find additional GSMMs available for other yeast species. However, those will not be discussed in this review due to their low level of curation and lack of validation.

CRITICAL PERSPECTIVE ON THE MODEL EVALUATION AND VALIDATION APPROACHES

Some reviews have already thoroughly covered the evaluation metrics most commonly applied each time, a new GSMM is published (Österlund, Nookaew and Nielsen 2012; Sánchez and Nielsen 2015). Here, we will forego the details of those metrics, instead providing a critical perspective concerning the use—or lack—of adequate validation approaches.

Despite the evolution of the GSMMs available for different organisms along the past two decades, uninfringeable evaluation criteria to assess the quality and completeness of GSMMs are still lacking. Typically, newer models of the same organism contain a broader metabolic coverage (Fig. 3B) and claim more consistent and improved predictive capabilities, particularly in terms of genotype–phenotype relationships, although the latter is always more subjective and questionable (Damiani *et al*. 2015; Heavner and Price 2015a). The scope of metabolic reconstructions in terms of the number of genes, reactions and metabolites has indeed been one of the highlighted evaluation criteria in the manuscripts of the published models. Due to the higher complexity of eukaryotic systems, including the presence of intracellular organelles and associated transport across cellular membranes, yeast genome-scale models have introduced another layer of characterization regarding the number of compartments represented in the metabolic network (Fig. 3B). However, from our point of view, it is important not to sacrifice quality over quantity in the metabolic reconstruction process, i.e. model update should not only be focused on model size improvements but also and foremost on the connectivity of the metabolic network and extent of manual curation, which will consequently influence the accuracy level of the resulting model.

Recently, despite the challenges, Heavner and Price (2015a) were able to evaluate the advances in *Saccharomyces cerevisiae* metabolic networks, through the direct comparison of 12 yeast genome-scale models. They have concluded that, in general, the iterative reconstruction of *S. cerevisiae* GSMMs has improved over time, particularly in terms of genomic coverage, number of reactions and single gene essentiality predictions, although some trade-offs between network size and model predictive performance were detected, meaning precisely that not always the expansion of the model scope has resulted in better predictive capabilities of gene essentiality. Interestingly, they were also able to cluster the different models according to their metabolite annotations reflecting their inheritances and chronological development, showing that model predictive ability usually reflects the iterative process of model curation. The same study noticed that the number of reactions that cannot carry any flux due to network structural constraints (known as blocked reactions) present in GSMMs of *S. cerevisiae* is over 20% for all of them across the different tested conditions, reaching nearly 40% in some cases (Heavner and Price 2015a). These reactions are often unconnected from the network, meaning that they are excluded from the computable metabolic space in strain optimization tasks, for example. However, if, on one hand, blocked reactions might reflect incorrect annotations or lack of manual curation, they often point the existence of gaps in biological knowledge, thus constituting an opportunity window for future research that should be harnessed to generate new knowledge and, consequently, enhance the connectivity of the models. In turn, Damiani *et al*. (2015) developed a system identification-based framework to compare the predictions of two GSMMs of the yeast *S. stipitis*. While iSS884 performed better in validations with physiological data, such as the prediction of growth rate or product excretion, iBB814 showed better qualitative agreements, such as predicting the effect on cell growth upon the inhibition of electron transport chain complexes. The developed validation framework corroborated that iBB814 has a better agreement with existing knowledge on that organism, while iSS884 presents some significant errors, despite good quantitative agreements.

There have been some appeals by experts of the metabolic modeling field to define standard quality criteria when constructing or assessing a new metabolic network—an effort we fully support—either stressing the need of collaborative research or clearer annotation standards, (Monk, Nogales and Palsson 2014; Ebrahim *et al*. 2015). The previously mentioned yeast consensus networks and platforms as MetaNetX or Pathway Tools are good, yet scarce or underutilized, examples of this (Karp *et al*. 2010; Moretti *et al*. 2016). For example, among all the yeast genome-scale models published after the first consensus model became available, only two GSMMs of *Y. lipolytica*, iNL895 and iYali4, present the same reactions' nomenclature in the respective metabolic networks. Although one could argue that nomenclature per se cannot directly contribute to affect or even improve the model performance, the use of standard identifiers for metabolites and reactions based on general databases and string representations (such as KEGG, PubChem, InChI and so forth) would certainly facilitate the automated integration and consequent comparison of different metabolic reconstructions. This would allow not only to better understand the underlying biology of the target organism but also to avoid error propagation, while highlighting opportunities for improving the consistency of the networks and their reusability.

It is known that many predictive errors are indeed caused by inconsistencies of the network, including incorrect assignment of GPR associations, reaction directionality or reversibility, incongruous stoichiometric parameters, missing reactions and inaccurate biomass composition (Zomorrodi and Maranas 2010; Dikicioglu, Kirdar and Oliver 2015; Heavner and Price 2015a). For example, the existence of unbalanced reactions in the metabolic network can significantly affect the accuracy of predictions (Kumar, Suthers and Maranas 2012). Still, it was recently reported that several models contain a significant fraction of reactions either unbalanced or for which mass balances cannot be determined due to absence of the corresponding metabolite formula (Ravikrishnan and Raman 2015). Although some methods and tools have been applied in the metabolic model reconstruction and refinement to guarantee the consistency of the network, as reviewed by Durot, Bourguignon and Schachter (2009), additional efforts should be devoted to check the model consistency with regard to mass and charge balance, thermodynamic information and confidence of the annotations, which are crucial elements in simulations. In addition, cellular growth is often simulated by maximizing the flux through a pseudo growth reaction, known as biomass objective function, which describes the growth requirements of a cell (Feist and Palsson 2010). Hence, the biomass composition is also a critical factor when studying genotype–phenotype relationships *in silico*. Nevertheless, even though advanced analytical methods have become gradually available, the biomass composition in *S. cerevisiae* GSMMs has scarcely changed over time, being recently dubbed by Dikicioglu, Kirdar and Oliver (2015) as the 'elephant in the room' of metabolic modeling. The authors demonstrated that flux distributions are very sensitive to changes in yeast's biomass composition, which should be represented in an accurate and condition-specific manner not to compromise the predictive accuracy of the model. In that sense, for example, the most recent *Pichia pastoris* model, iMT1026, includes different biomass compositions specific for each of the alternative carbon sources used (Tomàs-Gamisans, Ferrer and Albiol 2016). Surprisingly, we found that apart from the iLL672 model, none of the published *S. cerevisiae* genome-scale models include a detailed composition for vitamins, elements and cofactors required for growth. Cofactors, in particular, are often essential to proper enzymatic function, and some *Escherichia coli* modeling studies have demonstrated the importance of their representation

in the biomass equation. An increasing level of detail in the biomass objective functions has been perceived for models of *E. coli* and prokaryotes in general, and it would be interesting to conduct similar efforts in yeast GSMMs (Feist *et al.* 2007; Xavier, Patil and Rocha 2017).

Regarding more quantitative predictions, the model performance is often assessed based on the simulation of genotype–phenotype relationships, particularly gene essentiality, utilization of different carbon sources, growth rate and product excretion, to ensure that the metabolic model can accurately represent the biological system of interest. If, on one hand, physiological data on growth rates, substrate utilization and product formation are fairly accessible for all the yeast species with an available GSMM, on the other hand genome-wide datasets comprising gene deletion phenotypic information are only available for *S. cerevisiae* and, more recently, for *Schizosaccharomyces pombe*, hindering a more complete validation process. The single gene essentiality overall predictive accuracy reported in the yeast GSMMs publications, i.e. the fraction of correct predictions either for truly essential and non-essential genes, generally exceeds 80% or even 90% in some cases which, at a first glance, is quite remarkable. Nevertheless, if we account only the fraction of correctly predicted lethal knockouts, commonly known as model specificity, the agreement rates drop significantly to nearly half of the above-mentioned values (Zomorrodi and Maranas 2010). Moreover, from our analysis, if we deep root the gene essentiality prediction analysis, we find that some true positive cases, i.e. non-essential genes correctly predicted by the model, are associated with blocked reactions, suggesting that some positive results might be somewhat biased due to structural issues of the model (manuscript in preparation).

The results of mutant phenotypic studies are dependent on strain background, growth media and other environmental conditions (Hillenmeyer *et al.* 2008; Li *et al.* 2011; Alam *et al.* 2016; Jacquier 2016; Monk *et al.* 2016). For instance, a complex—undefined—growth medium is very difficult to formulate *in silico*. Also, some studies do not take into account specificities of the strain used in the experimental procedure, such as the presence of auxotrophic markers. Thus, it can be difficult to generate a totally reliable reference set of essential genes to be used in the model validation process, since some genes might be essential only in context-specific conditions (Zhang and Ren 2015). Studies that have developed their own large-scale experimental results based on well-defined and 'simulation friendly' conditions might therefore be on an advantageous position. At the same time, to develop unbiased comparisons when evaluating different models, one should at least use the same experimental dataset and *in silico* conditions. There are other simulation features that can influence the model predictive performance, including the choice of the growth threshold and the applied constraint-based algorithms. Although FBA has been the main constraint-based method used when evaluating new GSMMs, other methods such as MOMA and ROOM have also been applied. The latter two methods use a similar biological hypothesis which aims to minimize the number of significant flux changes with respect to the wild-type strain after a certain genetic perturbation, using a quadratic or mixed-integer linear programming, respectively (Segrè, Vitkup and Church 2002; Shlomi, Berkman and Ruppin 2005). Therefore, the quality of the reference wild-type flux distribution is crucial for obtaining meaningful results with these methods. A recent publication by our group showed that some of the most commonly used yeast GSMMs predict erroneous fluxes even in the well-studied pathways of the central carbon metabolism (Pereira, Nielsen and Rocha 2016). Interestingly, it was found that the oldest GSMM of *S. cerevisiae* (iFF708) was the best predictor of central carbon fluxes, which might explain why many authors still use this model in ME studies (Asadollahi *et al.* 2009; Brochado *et al.* 2010; Otero *et al.* 2013). Hence, even if most yeast models have demonstrated to accurately predict common physiological parameters such as specific growth rates, the analysis of the internal flux distribution, which is barely taken into account, should be part of the validation process, whenever fluxomics data are available. It is also known that cells may need time to adapt to genetic perturbations or environmental variability. Thus, some *in silico* predictions based on optimality criteria might not actually be incorrect, simply need to be verified in the light of evolution. Accordingly, it might be advisable to combine genome-scale modeling and adaptive laboratory evolution in the strain development process for certain biotechnological applications.

There is a natural tendency to overemphasize improvements in the predictive capabilities of new metabolic reconstructions, with particular prominence for the evaluation of cell viability after a specific gene deletion. However, we would like to stress that incorrect model predictions can constitute an excellent opportunity for knowledge generation, including the discovery of novel gene functions or alternative pathways, through the formulation of hypotheses to address these failures (Snitkin *et al.* 2008; Szappanos *et al.* 2011). Underlining these limitations when publishing or analyzing GSMMs, instead of overfitting the model for a particular experimental dataset or not providing clear information on how it was assembled, could guide future research toward new biological discovery (Heavner and Price 2015b). In summary, there is a clear need to define minimal criteria to assess the quality and completeness of genome-scale metabolic networks, along with more transparent reconstruction and validation processes, not only to increase our understanding of the target organism but also the reproducibility and applicability of the metabolic models.

THE QUEST FOR THE INTEGRATION OF REGULATORY AND KINETIC INFORMATION INTO GSMMs

Metabolism is regulated at multiple levels and, even thinking of the best-studied unicellular prokaryotic and eukaryotic organisms, we are still significantly far from having a full understanding of their underlying biological processes. Missing knowledge of enzyme regulators and other specific factors governing flux rates across different physiological conditions are amongst the main contributors for these deficiencies (Fendt *et al.* 2010). Nonetheless, biological knowledge has been increasingly generated and, by now, there are multiple data sets available that can be integrated in the genome-scale modeling process to improve the systems-level understanding of the cellular metabolism and even to link strain-specific phenotypes to molecular features (Monk *et al.* 2016; Mülleder *et al.* 2016). In fact, discrepancies between *in silico* predictions and experimental data are commonly justified with the lack of regulatory information in the metabolic networks. The integration of multi-omics data and other phenotypic information in functional metabolic models has been applied as a way of circumventing the absence regulatory rules in GSMMs, whilst increasing their scope and predictive capabilities. Yeast GSMMs have been used as scaffolds to integrate this type of data, as extensively described elsewhere (O'Brien, Monk and Palsson 2015; Sánchez and Nielsen 2015). Accordingly, a myriad of computational methods

to perform this task have been published (Kyung and Lun 2014). By integrating this information in a quantitative or qualitative way, we are able to shrink the solution space, which in turn is expected to improve the prediction of cellular phenotypes and/or gain insights into metabolic-driven adaptations, and even gene expression noise after certain genetic or environmental perturbations (Shlomi *et al.* 2008; Cimini *et al.* 2009; Chi, Tao and Liu 2015; O'Brien, Monk and Palsson 2015). Interestingly, a systematic evaluation of different methods used to integrate transcriptomic data into constraint-based models of metabolism showed that, in most situations, none of these methods outperform FBA and that many predictions may actually be the result of artifacts of the same methods and not a consequence of integrating gene expression data (Machado and Herrgård 2014). More recently, a new integration method was developed claiming significant improvements in the prediction of growth rates in comparison with previously existing algorithms (Motamedian *et al.* 2017). Still, the development of new methods to integrate metabolic networks and different data types remains a challenge.

Transcriptional regulatory networks representing the interplay between environmental conditions, transcription factors and target genes have also been addressed as a way of extending the coverage of constraint-based metabolic models of yeast and improve their accuracy and predictive ability (Chandrasekaran and Price 2013; Liu, Marras and Nielsen 2014). However, a recent systems-level study conducted to analyze the mechanisms regulating yeast metabolic fluxes showed that changes in fluxes across different nutrient conditions occur mainly due to changes in metabolite concentrations and not enzyme levels (Hackett *et al.* 2016). Also recently, a genome-wide quantitative metabolic map of the budding yeast was established by measuring amino acid concentration changes upon deletion on non-essential *S. cerevisiae* coding genes, showing that their deletion often creates very specific concentration signatures, apparently ruling the metabolism regulatory network (Mülleder *et al.* 2016).

If, on one hand, GSMMs have the advantage of being suitable for large-scale studies of the cellular metabolism, they clearly lack a detailed characterization of the cell, including the characterization of the factors determining the rate of reactions that could explain not only which reaction can occur but also when and to what extent they take place. Kinetic modeling constitutes an excellent opportunity toward this end but, although some advanced *in silico* approaches have been developed, several challenges still remain before large-scale kinetic modeling will be routinely applied in industrial biotechnology (Almquist *et al.* 2014; Vasilakou *et al.* 2016). The scarce amount of information available for the majority of reactions, as well as the uncertainty of the existing data in terms of the kinetic rates, expressions and parameters values significantly hamper the development of high-quality kinetic models. Yet, the emergence of new knowledge on how metabolic fluxes are determined points out that efforts should rather be redirected to use constraint-based models as templates for integrating kinetic data towards the generation of genome-scale kinetic models able to characterize the mechanisms of each reaction, as stressed by Smallbone *et al.* (2010) who gave the first steps in that direction by generating the first genome-scale kinetic model of yeast using linlog kinetics. More recently, a dynamic GSMM of the yeast *Pichia pastoris* has also been generated using the iPP668 model as scaffold, leading to improved flux distributions throughout dynamic cultivations resembling industrially relevant conditions (Saitua *et al.* 2017). The combination of dynamic modeling concepts with stoichiometric models can also constitute an excellent platform for integrating kinetic parameters toward the development of robust and large-scale kinetic models. Still, the usefulness of these approaches for cell factory improvements remains to be proven, meaning that efforts should also be applied to experimentally determine kinetic parameters under well-controlled conditions and to develop new methods for reducing the level of uncertainty currently linked to the generated data (Andreozzi, Miskovic and Hatzimanikatis 2016).

GSMMs AS GUIDING TOOLS OF METABOLIC ENGINEERING APPLICATIONS

The rising interest in producing fuels, chemicals and other materials from renewable resources associated to the concerns about sustainability have been the driving forces behind the developments in the industrial biotechnology field (Dai and Nielsen 2015). Although a clear assessment regarding the impact of GSMMs as guiding tools in ME industrial applications is still missing, several biotechnology companies have already filed patent applications for producing microorganisms mentioning the use of GSMMs in the strain design process, which clearly demonstrates their usefulness (Nielsen *et al.* 2014; Maia, Rocha and Rocha 2016). So, in addition to their role in biological elucidation and knowledge discovery processes already discussed, and notwithstanding some criticisms stated before, the use of GSMMs to rationally design and optimize microbial cell factories has indeed shown to be of great value. The rising interest in this topic has concomitantly driven the development of a myriad of computational strain optimization methods (CSOMs) which allow to find *in silico* combinations of genetic modifications toward desired phenotypical traits. OptKnock established the groundwork for the conception of many other CSOMs developed further on. Based on a bilevel structure, it was formulated to search for strain designs (reaction deletions targets) maximizing simultaneously two competing objective functions: cellular growth and the overproduction of a target compound (Burgard, Pharkya and Maranas 2003). Since then, CSOMs have been developed to search for non-intuitive genetic designs in more efficient and scalable ways. From the use of metaheuristic approaches (OptGene: Patil *et al.* 2005) to the consideration of gene deletions together with heterologous insertions—using mixed-integer programming methods (OptStrain: Pharkya, Burgard and Maranas 2004)—or gene expression levels (OptReg, OptForce, EMILio: Pharkya and Maranas 2006; Ranganathan, Suthers and Maranas 2010; Yang, Cluett and Mahadevan 2011), to the exploitation of transcriptional regulatory targets—using integrated (OptORF: Kim and Reed 2010) or unintegrated (BeReTa: Kim *et al.* 2016) networks of metabolism and transcriptional regulation—today, we can find over 30 different CSOMs in the literature. For a comprehensive review on this topic, see Maia, Rocha and Rocha (2016).

Since yeast species are the focus of this review, we underline the main *in silico*-aided ME applications for the development of experimentally validated yeast cell factories, as shown in Table 1. Despite the increasing number of available GSMMs of yeast, up until now, only a few have been used to design yeast cell factories. Interestingly, the first and simpler GSMM of *S. cerevisiae*, iFF708, has been used in several ME applications ranging from the improved production of biofuels and building block chemicals, such as ethanol and succinate (Bro *et al.* 2006; Agren, Otero and Nielsen 2013; Otero *et al.* 2013), to sesquiterpenes and aromatic compounds, including cubebol and vanillin, mainly based on OptGene suggested predictions (Asadollahi *et al.* 2009; Brochado *et al.* 2010). A combination of literature mining and

Table 1. A selection of experimentally validated model-based metabolic engineering applications/studies of different yeast species.

Organism	Target product	Model/method	Results	Reference (year)
S. cerevisiae	Bioethanol	iFF708/FBA	40% reduced glycerol yield on glucose and increased ethanol yield (+3%) without affecting the maximum specific growth rate	Bro *et al.* (2006)
S. cerevisiae	Sesquiterpenes	iFF708/OptGene	85% increase in the final cubebol titer	Asadollahi *et al.* (2009)
S. cerevisiae	Vanillin	iFF708/OptGene	1.5-Fold higher vanillin β-D-glucoside yield in batch mode, 2-fold productivity improvement in continuous culture	Brochado *et al.* (2010)
S. cerevisiae	2,3-Butanediol	iMM904/Optknock	2,3-Butanediol titer: 2.29 g l^{-1}; Product yield: 0.113 g.g^{-1} under anaerobic conditions	Ng *et al.* (2012)
S. cerevisiae	Fumaric acid	iND750/literature mining + FBA	Titer: ~1.68 g l^{-1} in batch culture	Xu *et al.* (2012)
C. glabrata	Malate	iNX804/FBA	Malate titer: 8.5 g l^{-1}	Chen *et al.* (2013)
S. cerevisiae	Succinate	iFF708/OptGene	30- and 43-fold improvements in succinate titer and succinate yield on biomass, respectively	Otero *et al.* (2013)
S. cerevisiae	Amorphadiene	iMM904/FDCA[a]	8- to 10-fold greater product yield compared to the wild type	Sun *et al.* (2014)
C. glabrata	Acetoin	iNX804/FBA	Final acetoin titer: 3.67 g l^{-1}	Li *et al.* (2014)
P. pastoris	Human recombinant protein	PpaMBEL1254/MOMA and FSEOF	Enhanced recombinant protein yield up to 40%	Nocon *et al.* (2014)
C. glabrata	Fumaric acid	iNX804/NS[b]	Final fumarate titer: 8.83 g l^{-1}	Chen *et al.* (2015)
Y. lipolytica	Lipids	iMK735/dFBA[c]	Byproduct (citrate) formation was reduced and lipid production yield increased	Kavšček *et al.* (2015)
S. cerevisiae	3HP	iTO977/pFBA[d]	3HP titer: 9.8 g l^{-1}; Yield: 13 % C-mol C-mol^{-1} glucose	Kildegaard *et al.* (2016)
S. cerevisiae	β-Farnesene	iLL672 (extended version)/pFBA	Farnese yield: 17.3% g g^{-1} Productivity: 2.24 g l^{-1} h^{-1} (requiring 75% less oxygen)	Meadows *et al.* (2016)

[a]FDCA—flux distribution comparison analysis.
[b]NS—Not specified.
[c]Dynamic FBA.
[d]Parsimonious enzyme usage FBA.

FBA using the iND750 model has also demonstrated that *in silico*-aided ME for the production of fumaric acid in *S. cerevisiae* can be efficiently developed (Xu *et al.* 2012).

Various sustainable forms of alternative energy and chemicals have been sought. Accordingly, researchers have also successfully designed and constructed *S. cerevisiae* strains with improved 2,3-butanediol production, based on *in silico* predictions obtained through the OptKnock framework and using the iMM904 model (Ng *et al.* 2012). Biological synthesis of terpenoids, which are candidate drugs and fragrances, has also been on the radar of ME researchers and systems biologists (Tippmann *et al.* 2013). Compared to bacteria, yeasts are more suitable to synthetize plant terpenoids mainly due to their ability to express plant cytochrome P450 enzymes (Schoendorf *et al.* 2001; Drăgan *et al.* 2011). The iMM904 GSMM was successfully used to investigate the impact of gene deletions—predicted through metabolic flux analysis using FBA and MOMA constraint-based methods—on isoprenoids pathway fluxes, hence showing that metabolic flux analysis combined with genome-scale modeling constitutes a powerful tool to identify suitable strategies for rerouting metabolic fluxes toward the production of exogenous terpenoids (Sun *et al.* 2014). More recently, an extended version of the iLL672 model was applied by Meadows *et al.* (2016) to identify an improved farnesene biosynthetic pathway in a study where the central carbon metabolism of *S. cerevisiae* was rewired for industrial isoprenoid production, achieving higher yields and productivity rates of the heterologous compound. In turn, and although the candidate ME strategy is not directly linked to *in silico* strain optimization or simulation methods, one of the latest *S. cerevisiae* models, iTO977, was recently used in an ME application for biosynthesis of 3-hydroxypropionic acid (3HP) to gain insights of the influence of 3HP biosynthesis on the flux distribution, hence guiding further ME efforts (Kildegaard *et al.* 2016).

Regarding other yeast species beyond *S. cerevisiae*, a model of the yeast *P. pastoris* (PpaMBEL1254) was successfully used to predict deletion and overexpression of genetic targets for overproduction of cytosolic human superoxide dismutase, using MOMA and flux scanning based on enforced objective function (FSEOF) approaches, respectively (Nocon *et al.* 2014). Meanwhile, a novel fed-batch strategy to avoid citrate excretion in the lipid production phase, deduced from FBA simulations with the iMK735 model, was developed leading to increased lipid yields in *Y. lipolytica* (Kavšček *et al.* 2015). Interestingly, although there is only one published GSMM for the yeast *Candida glabrata*, their authors have been demonstrating its usefulness in the strain design of *C. glabrata* for the production of different dicarboxylic acids, including malate, fumaric acid and acetoin (Chen *et al.* 2013, 2015; Li *et al.* 2014). However, since this yeast is also considered an opportunistic pathogen, this might constitute a significant drawback in regulatory affairs regarding the industrial use of the engineered strains.

Notwithstanding these successful *in vivo* applications, there is actually much room for improvements regarding the use of model-guided ME approaches, in particular to obtain yields, titers and productivity rates similar to those obtained using

more classical or non-rational methodologies, such as the bio-based and economically viable production of succinate (Verwaal *et al.* 2014). There are several reasons that might help to explain the current discrepancies. For example, many GSMMs and CSOMs remain to be properly validated *in vivo*, in part due to the lack of high-quality phenotypic data, including experimental data on flux distributions to better approximate the predicted fluxes with the real ones. On the other hand, despite the multiple computational methods developed to integrate the increasing number of -omics data available for several organisms within GSMMs, results show that we are not yet taking full advantage of it to improve phenotype predictions, meaning that more powerful methods leading to a systems-level understanding of the metabolism are needed. At the same time, reports underlining failed efforts in validating *in silico* predictions are often overlooked. For example, Gruchattka and Kayser tried to validate *in vivo* two strategies for using yeast as a terpenoid cell factory based on constrained minimal cut sets predictions previously obtained using a central carbon metabolic network. However, high amounts of acetate were produced instead of terpenoids (Gruchattka *et al.* 2013; Gruchattka and Kayser 2015). We therefore stress that reporting failed attempts is also of extreme importance, in particular, to detect major bottlenecks of the system and guide improvements in further *in silico*-aided ME studies. Others factors such as global regulatory networks, product toxicity or metabolic burden must also be taken into account to achieve optimal production phenotypes, meaning that the use of more combinatorial approaches, including iterative rounds of ME, should also be considered to overcome disconnections between genotypes and predicted phenotypes (Woolston, Edgar and Stephanopoulos 2013).

CHALLENGES AND FUTURE PERSPECTIVES

Genome-scale modeling of yeasts has been evolving for the last 15 years, contributing both to gain insights into the biological processes of several yeast species and to develop rational approaches in ME applications. However, many challenges remain, ranging from the need of clear evaluation approaches, high-quality phenotypic data and benchmark tests to assess the performance of newer models in context-specific environments—often hindered by the absence of standard identifiers of metabolites, reactions and enzymes among the different models—to the integration of kinetic and regulatory information known to govern metabolic fluxes across different physiological conditions, as a way of attaining more precise and robust predictions.

Since metabolism is highly regulated at various levels, the integration of multi-omics data might indeed help to ensure a higher robustness of the functional system. However, this has to come along with the development of computational methods capable of properly capturing the advantages of integrating this information, which is still not clear at the moment. Before that, more emphasis should be given to increase the transparency of the reconstruction and validation approaches, highlighting rather than omitting the major bottlenecks found across these processes, which can contribute to fill some knowledge gaps through hypothesis-driven experiments and ultimately to generate more reliable predictions.

For the model prokaryote *Escherichia coli,* the expansion of metabolic models to incorporate processes of proteome synthesis and localizations (ME-Models) or protein structure information (GEM-PRO models) has been gaining momentum, presenting significant improvements in model predictions (Chang *et al.* 2013; O'Brien *et al.* 2013). However, this type of models is not yet available for yeasts. Feizi *et al.* (2012) gave the first steps toward this goal by reconstructing the protein secretory machinery in yeast; still, it is not clear if it will be possible to create a ME-Model in the near future due to missing information about several processes in the yeast cell.

With the increasing availability of fluxomics data, a more attainable approach should pass through the assessment of internal flux distribution patterns in new genome-scale reconstructions, which has not been taken into account. For example, the iFF708 is the first yet better model predicting central carbon metabolic fluxes, and this might help to justify its success in several *in silico*-aided ME applications. Concurrently, it is important to determine what do cells really want, i.e. to know and formulate realistic objective functions based on experimental evidences to accurately represent specific cellular environments, as well as to represent biomass composition in a condition-specific manner, since this has a major impact in the simulation outputs (Feist and Palsson 2016). For example, it would be interesting to study cofactor requirements for cell growth in yeast models, as recently done for other organisms (Xavier, Patil and Rocha 2017).

Recent evidences also demonstrate that flux changes are often governed by changes in metabolite concentrations rather than enzyme levels (Hackett *et al.* 2016). So, despite all the limitations and need of powerful computational tools for dynamic modeling at the genome scale, further research in this field should desirably explore dynamic environments by integrating kinetic data into yeast metabolic networks.

Additionally, it is known that yeast strains from different ecological origins might present different phenotypic responses and even distinct intracellular metabolic fluxes (Nidelet *et al.* 2016). So, if we take this into account while performing computational simulations, we will likely improve the understanding of genotype–environment–phenotype relationships and, consequently, the rational design of cell factories (Long and Reed 2017). Lastly, in nature, yeast species present several interactions with other microorganisms and the compounds they secret can influence their co-habitants (Jouhten *et al.* 2016). Therefore, the development of microbial communities' models to study yeast species interactions is also expected to emerge in the next years.

FUNDING

This work was supported by the Portuguese Foundation for Science and Technology (FCT) under the scope of a Ph.D. grant (PD/BD/52336/2013), of the strategic funding of UID/BIO/04469/2013 unit and COMPETE 2020 (POCI-01–0145-FEDER-006684) and also in the context of the EU-funded initiative ERA-NET for Industrial Biotechnology (ERA-IB-2/0003/2013), in addition to the BioTecNorte operation (NORTE-01–0145-FEDER-000004) funded by European Regional Development Fund under the scope of Norte2020 - Programa Operacional Regional do Norte.

Conflict of interest. None declared.

REFERENCES

Agren R, Liu L, Shoaie S *et al.* The RAVEN toolbox and its use for generating a genome-scale metabolic model for penicillium chrysogenum. *PLoS Comput Biol* 2013;**9**:e1002980.

Agren R, Otero JM, Nielsen J. Genome-scale modeling enables metabolic engineering of Saccharomyces cerevisiae for succinic acid production. *J Ind Microbiol Biot* 2013;**40**:735–47.

Alam MT, Zelezniak A, Mülleder M *et al.* The metabolic background is a global player in Saccharomyces gene expression epistasis. *Nat Microbiol* 2016;**1**:15030.

Almquist J, Cvijovic M, Hatzimanikatis V *et al.* Kinetic models in industrial biotechnology—improving cell factory performance. *Metab Eng* 2014;**24**:38–60.

Andreozzi S, Miskovic L, Hatzimanikatis V. iSCHRUNK—in silico approach to characterization and reduction of uncertainty in the kinetic models of genome-scale metabolic networks. *Metab Eng* 2016;**33**:158–68.

Asadollahi MA, Maury J, Patil KR *et al.* Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. *Metab Eng* 2009;**11**:328–34.

Aung HW, Henry SA, Walker LP. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Ind Biotechnol* 2013;**9**:215–28.

Balagurunathan B, Jonnalagadda S, Tan L *et al.* Reconstruction and analysis of a genome-scale metabolic model for Scheffersomyces stipitis. *Microb Cell Fact* 2012;**11**:1–18.

Bro C, Regenberg B, Förster J *et al.* In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. *Metab Eng* 2006;**8**:102–11.

Brochado AR, Matos C, Møller BL *et al.* Improved vanillin production in baker's yeast through in silico design. *Microb Cell Fact* 2010;**9**:1–15.

Büchel F, Rodriguez N, Swainston N *et al.* Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst Biol* 2013;**7**:1–19.

Burgard AP, Pharkya P, Maranas CD. OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 2003;**84**:647–57.

Caspeta L, Shoaie S, Agren R *et al.* Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in-silico evaluation of their potentials. *BMC Syst Biol* 2012;**6**:1–14.

Chandrasekaran S, Price ND. Metabolic constraint-based refinement of transcriptional regulatory networks. *PLoS Comput Biol* 2013;**9**:e1003370.

Chang RL, Andrews K, Kim D *et al.* Structural systems biology evaluation of metabolic thermotolerance in escherichia coli. *Science (80-)* 2013;**340**:1220–3.

Chen X, Wu J, Song W *et al.* Fumaric acid production by Torulopsis glabrata: engineering the urea cycle and the purine nucleotide cycle. *Biotechnol Bioeng* 2015;**112**:156–67.

Chen X, Xu G, Xu N *et al.* Metabolic engineering of Torulopsis glabrata for malate production. *Metab Eng* 2013;**19**:10–6.

Cherry JM, Ball C, Weng S *et al.* Genetic and physical maps of Saccharomyces cerevisiae. *Nature* 1997;**387**:67–73.

Chi B, Tao S, Liu Y. Physiologically shrinking the solution space of a Saccharomyces cerevisiae Genome-Scale model suggests the role of the metabolic network in shaping gene expression noise. *PLoS One* 2015;**10**:e0139590.

Chung BK, Selvarasu S, Andrea C *et al.* Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. *Microb Cell Fact* 2010;**9**:1–15.

Cimini D, Patil KR, Schiraldi C *et al.* Global transcriptional response of Saccharomyces cerevisiae to the deletion of SDH3. *BMC Syst Biol* 2009;**3**:1–12.

Covert MW, Schilling CH, Famili I *et al.* Metabolic modeling of microbial strains in silico. *Trends Biochem Sci* 2001;**26**:179–86.

Dai Z, Nielsen J. Advancing metabolic engineering through systems biology of industrial microorganisms. *Curr Opin Biotechnol* 2015;**36**:8–15.

Damasceno LM, Pla I, Chang HJ *et al.* An optimized fermentation process for high-level production of a single-chain Fv antibody fragment in Pichia pastoris. *Protein Expr Purif* 2004;**37**:18–26.

Damiani AL, He QP, Jeffries TW *et al.* Comprehensive evaluation of two genome-scale metabolic network models for Scheffersomyces stipitis. *Biotechnol Bioeng* 2015; **112**:1250–62.

Dias O, Pereira R, Gombert AK *et al.* iOD907, the first genome-scale metabolic model for the milk yeast Kluyveromyces lactis. *Biotechnol J* 2014;**9**:776–90.

Dias O, Rocha M, Ferreira EC *et al.* Reconstructing genome-scale metabolic models with merlin. *Nucleic Acids Res* 2015;**43**:3899–910.

Dikicioglu D, Kirdar B, Oliver SG. Biomass composition: the 'elephant in the room' of metabolic modelling. *Metabolomics* 2015;**11**:1690–701.

Dobson PD, Smallbone K, Jameson D *et al.* Further developments towards a genome-scale metabolic model of yeast. *BMC Syst Biol* 2010;**4**:1–7.

Drăgan CA, Peters FT, Bour P *et al.* Convenient gram-scale metabolite synthesis by engineered fission yeast strains expressing functional human P450 systems. *Appl Biochem Biotech* 2011;**163**:965–80.

Duarte NC, Herrgård MJ, Palsson BØ. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 2004;**14**:1298–309.

Durot M, Bourguignon P, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev* 2009;**33**:164–90.

Ebrahim A, Almaas E, Bauer E *et al.* Do genome-scale models need exact solvers or clearer standards? *Mol Syst Biol* 2015;**11**:1–3.

Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae Rd* metabolic genotype. *Cell Biol Metab* 1999;**274**:17410–6.

Feist AM, Henry CS, Reed JL *et al.* A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;**3**:1–18.

Feist AM, Herrgard MJ, Thiele I *et al.* Reconstruction of biochemical networks in microbial organisms. *Nat Rev Microbiol* 2009;**7**:129–43.

Feist AM, Palsson BO. The biomass objective function. *Curr Opin Microbiol* 2010;**13**:344–9.

Feist AM, Palsson BØ. What do cells actually want? *Genome Biol* 2016;**17**:1–2.

Feizi A, Österlund T, Petranovic D *et al.* Genome-scale modeling of the protein secretory machinery in yeast. *PLoS One* 2012;**8**:e63284.

Fendt S, Oliveira AP, Christen S *et al.* Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast. *Mol Syst Biol* 2010;**6**:1–11.

Förster J, Famili I, Fu P *et al.* Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res* 2003;**13**:244–53.

Goffeau A, Barrell BG, Bussey H *et al.* Life with 6000 Genes. *Science (80-)* 1996;**274**:546–67.

Green M, Karp P. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics* 2004;**5**:1–16.

Gruchattka E, Hädicke O, Klamt S *et al.* In silico profiling of Escherichia coli and Saccharomyces cerevisiae as terpenoid factories. *Microb Cell Fact* 2013;**12**:84.

Gruchattka E, Kayser O. In vivo validation of in silico predicted metabolic engineering strategies in yeast: disruption of alpha-ketoglutarate dehydrogenase and expression of ATP-citrate lyase for terpenoid production. *PLoS One* 2015;**10**:1–25.

Hackett SR, Zanotelli VRT, Xu W *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* 2016;**354**:432–49.

Heavner BD, Price ND. Comparative analysis of yeast metabolic network models highlights progress, opportunities for metabolic reconstruction. *PLoS Comput Biol* 2015a;**11**:1–26.

Heavner BD, Price ND. Transparency in metabolic network reconstruction enables scalable biological discovery. *Curr Opin Biotechnol* 2015b;**34**:105–9.

Heavner BD, Smallbone K, Barker B *et al.* Yeast 5 - an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. *BMC Syst Biol* 2012;**6**:1–13.

Heavner BD, Smallbone K, Price ND *et al.* Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database* 2013;**2013**:1–5.

Henry CS, DeJongh M, Best AA *et al.* High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 2010;**28**:977–82.

Herrgård MJ, Lee B, Portnoy V *et al.* Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome Res* 2006;**16**:627–35.

Herrgård MJ, Swainston N, Dobson P *et al.* A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 2008;**26**:1155–60.

Hillenmeyer M, Fung E, Wildenhain J *et al.* The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 2008;**320**:362–5.

Hong KK, Nielsen J. Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. *Cell Mol Life Sci* 2012;**69**:2671–90.

Irani ZA, Kerkhoven EJ, Shojaosadati SA *et al.* Genome-scale metabolic model of Pichia pastoris with native and humanized glycosylation of recombinant proteins. *Biotechnol Bioeng* 2016;**113**:961–9.

Jacquier A. Systems biology: supplementation is not sufficient. *Nat Microbiol* 2016;**1**:1–2.

Jouhten P, Ponomarova O, Gonzalez R *et al.* Saccharomyces cerevisiae metabolism in ecological context. *FEMS Yeast Res* 2016;**16**:1–8.

Karp PD, Paley SM, Krummenacker M *et al.* Pathway tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* 2010;**11**:40–79.

Kavšček M, Bhutada G, Madl T *et al.* Optimization of lipid production with a genome-scale model of Yarrowia lipolytica. *BMC Syst Biol* 2015;**9**:1–13.

Kerkhoven EJ, Pomraning KR, Baker SE *et al.* Regulation of amino-acid metabolism controls flux to lipid accumulation in Yarrowia lipolytica. *npj Syst Biol Appl* 2016;**2**:1–7.

Kildegaard KR, Jensen NB, Schneider K *et al.* Engineering and systems-level analysis of Saccharomyces cerevisiae for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway. *Microb Cell Fact* 2016; **15**:1–13.

Kim J, Reed JL. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol* 2010;**4**:1–19.

Kim J, Reed JL. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biol* 2012;**13**:1–12.

Kim M, Sun G, Lee D-Y *et al.* BeReTa: a systematic method for identifying target transcriptional regulators to enhance microbial production of chemicals. *Bioinformatics* 2016;**33**:87–94.

Kuepfer L, Sauer U, Blank LM. Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Res* 2005;**15**:1421–30.

Kumar A, Suthers PF, Maranas CD. MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics* 2012;**13**:6.

Kyung M, Lun DS. Methods for integration of transcriptomic data in genome-scale metabolic models. *Comput Struct Biotechnol J* 2014;**11**:59–65.

Lee M, Nurse P. Cell cycle control genes in fission yeast and mammalian cells. *Trends Genet* 1988;**4**:287–90.

Li Q, Sun Z, Li J *et al.* Enhancing beta-carotene production in Saccharomyces cerevisiae by metabolic engineering. *FEMS Microbiol Lett* 2013a;**345**:94–101.

Li S, Chen X, Liu L *et al.* Pyruvate production in Candida glabrata: manipulation and optimization of physiological function. *Crit Rev Biotechnol* 2013b;**8551**:1–10.

Li S, Gao X, Xu N *et al.* Enhancement of acetoin production in Candida glabrata by in silico-aided metabolic engineering. *Microb Cell Fact* 2014;**13**:1–11.

Li Z, Vizeacoumar FJ, Bahr S *et al.* Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat Biotechnol* 2011;**29**:361–9.

Liu G, Marras A, Nielsen J. The future of genome-scale modeling of yeast through integration of a transcriptional regulatory network. *Quant Biol* 2014;**2**:30–46.

Liu T, Zou W, Liu L *et al.* A constraint-based model of Scheffersomyces stipitis for improved ethanol production. *Biotechnol Biofuels* 2012;**5**:2–11.

Loira N, Dulermo T, Nicaud J-M *et al.* A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. *BMC Syst Biol* 2012;**6**:1–9.

Long MR, Ong WK, Reed JL. Computational methods in metabolic engineering for strain design. *Curr Opin Biotechnol* 2015;**34**:135–41.

Long MR, Reed JL. Improving flux predictions by integrating data from multiple strains. *Bioinformatics* 2017;**33**:893–900.

Machado D, Herrgård MJ. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput Biol* 2014;**10**:e1003580.

Magnúsdóttir S, Heinken A, Kutt L *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol* 2016;**35**:81–92.

Maia P, Rocha M, Rocha I. In silico constraint-based strain optimization methods: the quest for optimal cell factories. *Microbiol Mol Biol R* 2016;**80**:45–67.

Meadows AL, Hawkins KM, Tsegaye Y *et al.* Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* 2016;**537**:694–7.

Mishra P, Park G-Y, Lakshmanan M *et al.* Genome-scale metabolic modeling and in silico analysis of lipid accumulating yeast Candida tropicalis for dicarboxylic acid production. *Biotechnol Bioeng* 2016;**113**:1993–2004.

Mo ML, Palsson BO, Herrgård MJ. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 2009;**3**:37.

Monk J, Nogales J, Palsson BO. Optimizing genome-scale network reconstructions. *Nat Biotechnol* 2014;**32**:447–52.

Monk JM, Koza A, Campodonico MA *et al.* Multi-omics quantification of species variation of escherichia coli links molecular features with strain multi-omics quantification of species variation of escherichia coli links molecular features with strain phenotypes. *Cell Syst* 2016;**3**:238–51.

Moretti S, Martin O, Van Du Tran T *et al.* MetaNetX/MNXref - Reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res* 2016;**44**:D523–6.

Motamedian E, Mohammadi M, Shojaosadati SA *et al.* TRFBA: an algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data. *Bioinformatics* 2017;**4**:1–17.

Mülleder M, Calvani E, Alam MT *et al.* Functional metabolomics describes the yeast biosynthetic regulome. *Cell* 2016;**167**:553–65.

Ng CY, Jung M, Lee J *et al.* Production of 2,3-butanediol in Saccharomyces cerevisiae by in silico aided metabolic engineering. *Microb Cell Fact* 2012;**11**:1–14.

Nidelet T, Brial P, Camarasa C *et al.* Diversity of flux distribution in central carbon metabolism of S. cerevisiae strains from diverse environments. *Microb Cell Fact* 2016;**15**:1–13.

Nielsen J, Fussenegger M, Keasling J *et al.* Engineering synergy in biotechnology. *Nat Chem Biol* 2014;**10**:319–22.

Nielsen J, Larsson C, van Maris A *et al.* Metabolic engineering of yeast for production of fuels and chemicals. *Curr Opin Biotechnol* 2013;**24**:398–404.

Nocon J, Steiger MG, Pfeffer M *et al.* Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. *Metab Eng* 2014;**24**:129–38.

Nookaew I, Jewett MC, Meechai A *et al.* The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 2008;**2**:1–15.

O'Brien EJ, Lerman JA, Chang RL *et al.* Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* 2013;**9**:693.

O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. *Cell* 2015;**161**:971–87.

Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 2009;**5**:1–16.

Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nat Biotechnol* 2010;**28**:245–8.

Österlund T, Nookaew I, Bordel S *et al.* Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol* 2013;**7**:1–10.

Österlund T, Nookaew I, Nielsen J. Fifteen years of large scale metabolic modeling of yeast: developments and impacts. *Biotechnol Adv* 2012;**30**:979–88.

Otero JM, Cimini D, Patil KR *et al.* Industrial systems biology of Saccharomyces cerevisiae enables novel succinic acid cell factory. *PLoS One* 2013;**8**:e54144.

Paddon CJ, Westfall PJ, Pitera DJ *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 2013;**496**:528–32.

Pan P, Hua Q. Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, Yarrowia lipolytica. *PLoS One* 2012;**7**:1–11.

Park JM, Kim TY, Lee SY. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnol Adv* 2009;**27**:979–88.

Patil KR, Rocha I, Förster J *et al.* Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics* 2005;**6**:1–12.

Pereira R, Nielsen J, Rocha I. Improving the flux distributions simulated with genome-scale metabolic models of Saccharomyces cerevisiae. *Metab Eng Commun* 2016;**3**:153–63.

Petranovic D, Nielsen J. Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol* 2008;**26**:584–90.

Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res* 2004;**14**:2367–76.

Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng* 2006;**8**:1–13.

Ranganathan S, Suthers PF, Maranas CD. OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* 2010;**6**:e1000744.

Ravikrishnan A, Raman K. Critical assessment of genome-scale metabolic networks: the need for a unified standard. *Brief Bioinform* 2015;**16**:1057–68.

Saitua FJ, Torres P, Pérez-Correa JR *et al.* Dynamic genome-scale metabolic modeling of the yeast Pichia pastoris. *BMC Syst Biol* 2017;**11**:1–21.

Sánchez BJ, Nielsen J. Genome scale models of yeast: Towards standardized evaluation and consistent omic integration. *Integr Biol* 2015;**7**:846–58.

Satish Kumar V, Dasika MS, Maranas CD. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* 2007;**8**:1–16.

Savinell JM, Palsson BO. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *JTheoretical Biol* 1992;**154**:421–54.

Schoendorf A, Rithner CD, Williams RM *et al.* Molecular cloning of a cytochrome P450 taxane 10β-hydroxylase cDNA from Taxus and functional expression in yeast. *P Natl Acad Sci USA* 2001;**98**:1501–6.

Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *P Natl Acad Sci USA* 2002;**99**:15112–7.

Senger R, Yen J, Tanniche I *et al.* Designing metabolic engineering strategies with genome-scale metabolic flux modeling. *Adv Genomics Genet* 2015;**5**:93–105.

Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *P Natl Acad Sci USA* 2005;**102**:7695–700.

Shlomi T, Cabili MN, Herrgård MJ *et al.* Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* 2008;**26**:1003–10.

Smallbone K, Simeonidis E, Swainston N *et al.* Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst Biol* 2010;**4**:1–9.

Snitkin ES, Dudley AM, Janse DM *et al.* Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. *Genome Biol* 2008;**9**:R140.

Sohn SB, Graf AB, Kim TY *et al*. Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. *Biotechnol J* 2010;**5**:705–15.

Sohn SB, Kim TY, Lee JH *et al*. Genome-scale metabolic model of the fission yeast Schizosaccharomyces pombe and the reconciliation of in silico/in vivo mutant growth. *BMC Syst Biol* 2012;**6**:1–12.

Sturgeon C, Kemmer D, Anderson HJ *et al*. Yeast as a tool to uncover the cellular targets of drugs. *Biotechnol J* 2006;**1**:289–98.

Sun Z, Meng H, Li J *et al*. Identification of novel knockout targets for improving terpenoids biosynthesis in saccharomyces cerevisiae. *PLoS One* 2014;**9**:e112615.

Szappanos B, Kovács K, Szamecz B *et al*. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet* 2011;**43**:656–62.

Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 2010;**5**:93–121.

Tippmann S, Chen Y, Siewers V *et al*. From flavors and pharmaceuticals to advanced biofuels: Production of isoprenoids in Saccharomyces cerevisiae. *Biotechnol J* 2013;**8**:1435–44.

Tomàs-Gamisans M, Ferrer P, Albiol J. Integration and validation of the genome-scale metabolic models of Pichia pastoris: A comprehensive update of protein glycosylation pathways, lipid and energy metabolism. *PLoS One* 2016; **11**:1–24.

Vasilakou E, Machado D, Theorell A *et al*. Current state and challenges for dynamic metabolic modeling. *Curr Opin Microbiol* 2016;**33**:97–104.

Verwaal R, Wu L, Damveld RA *et al*. Succinic acid production in a eukaryotic cell. *European Patent Office* 2014; WO/2009/065778.

Werten MWT, Van Den Bosch TJ, Wind RD *et al*. High-yield secretion of recombinant gelatins by Pichia pastoris. *Yeast* 1999;**15**:1087–96.

Woolston BM, Edgar S, Stephanopoulos G. Metabolic engineering: past and future. *Annu Rev Chem Biomol* 2013;**4**:259–88.

Xavier JC, Patil KR, Rocha I. Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. *Metab Eng* 2017;**39**:200–8.

Xu G, Zou W, Chen X *et al*. Fumaric acid production in Saccharomyces cerevisiae by in silico aided metabolic engineering. *PLoS One* 2012;**7**:e52086.

Xu N, Liu L, Zou W *et al*. Reconstruction and analysis of the genome-scale metabolic network of Candida glabrata. *Mol Biosyst* 2013;**9**:205–16.

Yang L, Cluett WR, Mahadevan R. EMILiO: A fast algorithm for genome-scale strain design. *Metab Eng* 2011;**13**:272–81.

Zhang Z, Ren Q. Why are essential genes essential? - The essentiality of Saccharomyces genes. *Microb Cell* 2015;**2**:280–7.

Zomorrodi AR, Maranas CD. Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data. *BMC Syst Biol* 2010;**4**:1–15.