Reconstruction and analysis of genome-scale metabolic model of weak Crabtree positive yeast *Lachancea kluyveri*

Piyush Nanda
Indian Institute of Technology Kharagpur
Pradipta Patra
Indian Institute of Technology Kharagpur
Manali Das
Indian Institute of Technology Kharagpur
Amit Ghosh (amitghosh@iitkgp.ac.in)
Indian Institute of Technology Kharagpur https://orcid.org/0000-0003-3514-885X

Research article

Keywords: Metabolic Flux Analysis, Non-conventional yeast, Dynamic Flux Balance Analysis, Uracil degradation, Ethyl acetate

Posted Date: January 21st, 2020

DOI: https://doi.org/10.21203/rs.2.16651/v2

Version of Record: A version of this preprint was published at Scientific Reports on October 1st, 2020. See the published version at https://doi.org/10.1038/s41598-020-73253-3.

Abstract

Background *Lachancea kluyveri*, a weak Crabtree positive yeast, has been extensively studied for its unique URC pyrimidine catabolism pathway. It produces more biomass than *Saccharomyces cerevisiae* due to the underlying weak Crabtree effect and resorts to optimal fermentation only in oxygen limiting conditions that render it a suitable host for industrial-scale protein production. Ethyl acetate, an important industrial chemical, has been demonstrated to be a major overflow metabolite during aerobic batch cultivation with a specific rate of 0.12 g per g dry weight per hour. Here, we reconstruct a genome-scale metabolic model of the yeast to better explain the observed phenotypes and aid further hypothesis generation.

Results We report the first genome-scale metabolic model, iPN730, using Build Fungal Model in KBase workspace. The inconsistencies in the draft model were semi-automatically corrected using literature and published datasets. The curated model comprises of 1235 reactions, 1179 metabolites, and 730 genes distributed in 8 compartments (organelles). The *in silico* viability in different media conditions and the growth characteristics in various carbon sources show good agreement with experimental data. Dynamic flux balance analysis describes the growth dynamics, substrate utilization and product formation kinetics in various oxygen-limited conditions. The URC pyrimidine degradation pathway incorporated into the model enables it to grow on uracil or urea as the sole nitrogen source.

Conclusion The genome-scale metabolic construction of *L. kluyveri* will provide a better understanding of metabolism, particularly that of pyrimidine metabolism and ethyl acetate production. Metabolic flux analysis using the model will enable hypotheses generation to gain a deeper understanding of metabolism in weakly Crabtree positive yeast and in fungal biodiversity in general.

Background

Lachancea kluyveri, previously known as *Saccharomyces kluyveri*, is a weak Crabtree positive yeast and has presented in numerous studies due to its unique metabolic properties[1–3]. Its unique pathway of pyrimidine catabolism has been well characterized and its importance has been discussed[1, 4]. It has the ability to utilize purines and pyrimidines as sole nitrogen sources which makes it a great interest to the pharmaceutical industry[5]. The ability to assimilate and ferment melibiose by *L. kluyveri* is also a differentiating feature from other characterized yeasts[3]. The organism diverged from *Saccharomyces cerevisiae* after the Whole Genome Duplication (WGD) event. Previously, using searchDOGS algorithm, the position of the whole genome duplication event has been described for 11 yeast species and they marked the location of the whole genome duplication (WGD) event which separated the *Saccharomyces* from the *Lachancea* genus [6]. The effect of the WGD event is reflected in the number of ortholog duplicates that *L. kluyveri* genes have in *S. cerevisiae.* The single knockout simulation would also explain the effect of such duplicate genes on the viability of the model and metabolic flux. Genome-scale metabolic models (GEMs) are one of the important systems biology tools as they integrate genomic information, transcriptomic, metabolomics and associated experimental data for a given organism into a

whole-cell metabolic network[7–9]. Moreover, complementary to general omics analysis, GEMs provide a scope to integrate omics data into metabolic networks and help to predict condition-specific metabolic capabilities. To date, GEMs have been successfully applied to design microbial strains for bioproduction and evaluate their capabilities in different conditions[10]. Specifically, the model yeast *S. cerevisiae* was the first choice among the eukaryotic organisms to be fully sequenced [11, 12], and it is one of the conventional workhorses in cell factory engineering for bio-production of several compounds with various applications in chemical [13, 14], food[15, 16]and pharmaceutical industries[17]. These models are widely being used for various biotechnological predictions and experimental designs. However, non-conventional yeast species include important human pathogens or have also been reported to be suitable platforms for several biotechnological applications and various models have been therefore reconstructed for many yeasts.

Due to the weak Crabtree energy metabolism of *L. kluyveri*, it resorts to minimal fermentation in aerobic conditions[18]. This enables it to produce more biomass compared to *S. cerevisiae*, which diverts substantial flux towards ethanol production even in aerobic conditions [2]. This makes it a lucrative host for protein production[19]. Ethyl acetate has also been measured to be a major overflow metabolite in growth studies[2]. *L. kluyveri*, therefore, could be leveraged to uncover the mechanisms behind these metabolic routes and also serves as a potential industrial organism.

When genome-scale metabolic models are combined with constraint-based algorithms, they offer an excellent chance to analyze metabolism and genotype-phenotype relationships[10]. Genome-scale metabolic reconstructions (GEM) are expanding our understanding of cellular metabolism and its dynamics. It supplies us with a systems-level picture of the metabolism derived from its genotype. GEMs have been used to analyze disease phenotypes and improvise the production of chemicals from cell factories. Fungal GEMs have been widely used for metabolic engineering efforts to overproduce target metabolites like fuels and chemicals[20, 21, 14, 13]. As an obvious target, *S. cerevisiae*'s GEM was constructed iteratively due to its immense industrial and academic importance. Beginning from iFF708i[9], other metabolic models like iND750[22], iMH805/775[23] and iMM904[24] have been reconstructed. iMH805 has gene regulatory information making it a powerful model to analyze the coordination of metabolic pathways with gene circuits[23]. This has begun the quest for exploring related yeast strains that can furnish novel phenotypes. In this regard, numerous yeast GEMs have been published for *Kluyveromyces lactis*[25], *Pichia pastoris*[26] and *Yarrowia lipolytica*[27]. GEMs can be used to visualize a systems-level effect of environmental conditions and induced perturbations.

To understand the metabolic capabilities of *L. kluyveri*, here we built the first GEM named iPN730, which comprises of 1235 reactions, 1179 metabolites, 730 genes, and 8 compartments. *Lachancea kluyveri* NRRL-12651 strain was selected for the genome annotation due to the extensive experimental analysis that has been conducted on this organism particularly with respect to pyrimidine degradation and ethyl acetate production. Single knockout analysis using Flux Balance Analysis (FBA) [8] led to a significant agreement with experimental knockout data [28] for *S. cerevisiae* in various *in silico* media. Dynamic Flux Balance Analysis (DFBA) [29] resulted in growth dynamics, product formation that agreed with

experimental observation as reported in previous studies. The reported pathways for pyrimidine catabolism and ethyl acetate production are also incorporated into the model to enable further studies. To our knowledge, it is the first attempt to develop a GEM for *L. kluyveri* through which we anticipate a better understanding of the metabolic routes in the organism and streamlining its industrial applications.

Results And Discussions

Genome-scale metabolic reconstruction of *Lachancea kluyveri*

The genome annotation of *Lachancea kluyveri* NRRL 12651 (Assembly: ASM14922v1, GenBank assembly accession: GCA_000149225.1) using Yeast Genome Annotation Pipeline [30] yielded 5,505 ORFs. The draft model from the annotated genome using Build Fungal Model in Kbase contained 1180 reactions, 1232 metabolites, and 730 genes. The draft model harbored the integration of data from 13 well-curated fungal genome-scale reconstructions. *S. cerevisiae* was used as the template model due to its high quality and reliability. The homology table generated from the BDBH algorithm for proteome comparison has been incorporated (Table S4, Supplementary Information). This can be used to predict possible duplication, inversion, deletion, insertion or synteny loci between the organisms. This can be used to predict possible duplication, inversion, deletion, insertion and synteny loci between the organisms i.e. *L. kluyveri* and *S. cerevisiae*. The details about the fungal templates used for reconstruction have been described in Materials and Methods. In the gene-level data integration statistics obtained from Kbase Build fungal model (Fig 1A), the percentage represented the fraction of protein-coding genes that have an ortholog in a specific fungal template model. *S. cerevisiae* showed the highest similarity (17%) followed by *Candida tropicalis* (13.6%) with *L. kluyveri.* Although it is known that *E. gossypii* and *K. lactis* are evolutionarily closer to *L. kluyveri* than *S. cerevisiae* and *C. tropicalis*, we present here a phylogenetic analysis at the gene level to explain the phenomena. The physiophenotypic similarities of *L. kluyveri* with *S. cerevisiae* and *C. tropicalis* are also justified given that both are Crabtree positive and fermenting yeasts. *E. gossypii* is non-fermenting filamentous fungi [31, 32] and *K. lactis* is a budding Crabtree negative yeast[25]. The 18S rRNA based phylogenetic tree (Fig 2A) built using the NCBI Taxonomy browser with iTOL visualization and genome-level phylogenetic tree (Fig 2B) constructed through REALPHY server describes the proximity of *S. cerevisiae* and *C. tropicalis* to *L. kluyveri*.

In a metabolic model, correct mass balance in the reactions is critical for the quality of predictive power[7]**.** Elemental and charge balances can be calculated and analyzed only if the model contains the formula and charge for the metabolites respectively. All the metabolites were assigned formula and charge with reference from CHEBI (Chemical Entities of Biological Interest), PubChem (NIH) and BiGG Database (UCSD)[33]. For metabolites having ambiguous alkyl group (-R), the alkyl group was considered as a conserved moiety “R” in the metabolite formula. The default metabolite ID was replaced by standard nomenclature (MIRIAM) for genome-scale metabolic reconstruction as used elsewhere for well-established models iMM904, iAF1260, etc. The assigned ID captured both the information regarding the chemical name and the compartments of the metabolites. Similar assignment was made for reactions in

the reconstruction. Further refinement of the model was necessary with regard to the gaps in the biosynthetic pathways and the ambiguous reversibility of the reactions.

To assign the correct reversibility of the metabolic reactions, the iMM904 model was used as a template and homologous reactions were mapped to the draft model based on modelSEED reaction (Table S5, Supplementary Information). This corrected 82.5 % (1018 of 1232 reactions) of the inconsistencies in the reversibility of the reactions. The remaining inconsistencies were resolved by manual curation using COBRApy[34]. Upper and lower bound for internal reversible reactions were set to a high flux value i.e. 1000 mmol gdw^{-1} hr^{-1} to allow unconstrained flux through the internal metabolic network. For irreversible reactions, a lower flux bound of zero was assigned to restrict any flux in the reverse direction. The reversibility information on such ambiguous reactions was obtained from MetaCyc, KEGG and previously published datasets[35] on metabolic reactions. The irreversible reactions comprised of reactions that involve i) ATP as reactant ii) Oxygen as the product and iii) Reactions involving NADH and NADPH. (Fig 1B). These are due to the high enthalpy of the reactions associated with ATP breakdown, NADH/NADPH consumption and the fact that the cell cannot produce oxygen through its metabolism[7]. Through this semi-automated correction of the reversibility, 879 reactions were subject to irreversible flux bounds.

In GEMs, the biomass equation describes the growth of the cells from the biomass precursors like amino acids, carbohydrates, lipids, and ions. There was no natural auxotrophy reported for the organism. Hence, these biomass precursors should be produced from the central carbon pathway rather than uptake reactions from *in silico* media. The biomass equation comprised of 47 metabolites involved in biomass synthesis. The defects or gaps in biosynthetic pathways for these metabolites would block any biomass synthesis. 54 reactions were added to the model with reference from the KEGG pathway database. This enabled the organism to produce all the biomass precursors i.e. nucleotides, amino acids, carbohydrates, lipids and energy equivalents just from glucose, oxygen, ammonium salt and trace elements. The distribution of these gap filled reactions in different compartments of the metabolic model has been presented in Fig 1C. In internal reactions, the major inconsistencies were in mannan biosynthesis and fatty acid production pathways. Mannan biosynthesis takes place in the endoplasmic reticulum facilitated by dolichol phosphate anchoring[36]. The major gaps in this pathway comprised of transport reactions between the cytoplasm and endoplasmic reticulum. It was rectified manually with reference from KEGG and MetaCyc. Similarly, the fatty acid pathway comprises elongation of long-chain fatty acid with the addition of an even number of carbon through malonyl-CoA [37]. Due to this, any reaction blockage owing to incorrect flux direction or missing metabolites could render no flux through the whole pathway. We therefore, manually checked through the whole pathway to ensure that no reaction is blocked. There were significant inconsistencies in the transport of gases across the compartments. The diffusion of gases from extracellular space to a specific compartment and vice versa was important. The transport of oxygen from the extracellular space to the mitochondria through the cytosol was corrected by adding the transport reactions at the cell membrane and mitochondrial membrane. Any erroneous reaction which consumes oxygen (fatty acids desaturase reactions) were removed or given the proper reaction directionality. This was to avoid the diversion of oxygen to these reactions in the cytosol which is

physiologically irrelevant. The numerical confidence on the reactions in the model was assigned based on the gene-protein reaction (GPR) and observed flux through it (Table S2, Supplementary Information). This will help the users to ascertain the confidence in predictions made using the model for metabolic flux analysis.

Although *S. cerevisiae* is phylogenetically very close to *L. kluyveri*, it has significantly different model statistics (Table 1). *S. cerevisiae* has a higher number of duplicate genes compared to *L. kluyveri* which has been reflected in the total number of genes in their respective metabolic models i.e. 904 vs 730. For example, *S. cerevisiae* has fatty acyl-CoA synthetases i.e. FAA1 (Gene symbol: YOR317) and FAA4 (Gene symbol: YMR246W) whereas *L. kluyveri* has just one that encodes for a fatty acyl-CoA synthetase. Similarly in the annotated genome, 527 instances showed that multiple *S. cerevisiae* genes were mapped to a single locus in *L. kluyveri* (Fig 1D).

The GEM for *L. kluyveri* was derived from the consensus reconstruction of different published metabolic models (Materials and Methods). The physiologically relevant characteristics i.e. formula, charge, and reversibility were assigned to metabolites and reactions. Incomplete biosynthetic pathways derived from the genome annotation were corrected to enable the synthesis of all essential biomass precursors. This qualified the model for further downstream applications to predict the flux distribution in the pathway in physiological conditions. The species-specific reactions like ethyl acetate metabolism and pyrimidine degradation routes were incorporated into the model. The model also mapped the genotype of the organism to a complex phenotype like flux distribution in the metabolic network which can be used to predict the phenotypic changes in the organism in response to combinatorial genetic perturbations.

Model validation through flux balance analysis

Flux Balance Analysis (FBA) solves a linear optimization problem with the objective as the specific growth of the cell i.e. flux through the biomass equation. This assigns fluxes to active reactions in the model given the constraints. The constraints of the optimization are defined by the bounds on the flux through the reactions and the uptake rate of substrates and other media components. This is further dictated by the thermodynamic Gibbs free energy (ΔG) of the metabolic reactions.

From the previously generated experimental data[18, 41, 42], the accuracy of the prediction of the model was evaluated. Using the similar substrate and oxygen uptake rate as measured in the experiment, the specific growth rate and the product secretion rates were evaluated *in silico* by FBA. In the experiments reported previously, for a glucose uptake rate of 2.28 mmol $gdw^{-1}hr^{-1}$ and an oxygen uptake rate of 6.2 mmol $gdw^{-1}hr^{-1}$, a specific growth rate of 0.2 hr^{-1} were observed[18]. This corresponded to a carbon dioxide production rate of 6.4 mmol $gdw^{-1}hr^{-1}$ and a respiratory quotient (RQ) (the ratio of the volume of carbon dioxide evolved to that of oxygen consumed) of 1.03. To this, the predicted specific growth rate and the respiratory quotient were compared to the experimentally measured values[18, 43] (Table S1, Supplementary Information). For the specific growth rate, the analysis yielded a Pearson's correlation coefficient of 0.96 and a p-value of 0.0021 (Fig 3A). For the respiratory quotient, the Pearson's correlation

coefficient value was 0.97 and a p-value of 0.033 (Fig 3B). Usually, the respiratory quotient depends on the respiratory and the fermentative metabolism of the cell. The RQ values vary in aerobic and anaerobic conditions depending on the metabolic state of the cell. A high correlation between experimental and *in silico* measurements validated the correct behaviour of the model in the span of all possible metabolic states. The bi-modal nature of the distribution in histograms of growth rate and RQ values is due to the growth of the organism lying in the aerobic and anaerobic growth regimes (Fig 3A and Fig 3B). The growth rate in aerobic conditions is higher compared to anaerobic conditions.

The *in-silico* growth and the viability of the model were evaluated in different carbon sources (Fig 4A). The carbon sources considered were glucose, galactose, maltose, sucrose, trehalose, melibiose, arabinose, ethanol, and citrate. The *in vivo* ability of *L. kluyveri* to grow in multiple carbon sources was procured from MycoBank[44]. The *in-silico* viability showed agreement with the reported *in vivo* viability (Fig 4A). With an uptake rate of 1 g gdw^{-1} hr^{-1}, the *in-silico* growth rate on various carbon sources like glucose, galactose, maltose, sucrose, melibiose, ethanol and glycerol was evaluated in aerobic and anaerobic conditions (Fig 4B). The highest *in silico* growth rate in aerobic conditions was for ethanol followed by glycerol and glucose. The lowest growth rate was for galactose. The growth rate in anaerobic conditions was proportional to aerobic conditions for glucose, galactose, maltose, sucrose and melibiose. In an interesting observation, the growth rate on glycerol and ethanol was higher compared to others in aerobic conditions but the lowest in anaerobic conditions. This can be explained by the redox chemistry of ethanol and glycerol utilization. Both of them are products of the central carbon metabolism formed by reduction of glycolytic intermediates i.e. dihydroxyacetone phosphate[45] and acetaldehyde respectively. Ethanol and glycerol are non-fermentable carbon sources due to the redox imbalance caused during the anaerobic conditions. NADH redox imbalance in anaerobic glycerol metabolism impedes the growth[46]. Similarly, metabolic assimilation of ethanol in anaerobic conditions causes NADH imbalances due to the production of excess NADH which cannot be oxidized due to the absence of oxygen. We also performed batch growth experiments to validate the prediction of the growth rates on various carbon sources. The organism failed to grow on glycerol as the sole carbon source in shake flask experiment. This could be due to requirement of complete aerobic conditions in shake flask experiment. We could grow the organism on maltose, sucrose and galactose (Table S6, Supplementary Information) and measured the specific growth rate and substrate uptake rates (materials and methods). Using this, we showed a high correlation between the experimental growth rate and the predicted growth rate (Fig 4C).

The substrate consumption, growth and product formation profiles were obtained from dynamic flux balance analysis (DFBA). DFBA [29] comprises of iterative flux balance analysis on a given amount of carbon source. It was important to confirm whether the experimentally observed metabolite production and growth rate tallied with the simulation results. A glucose uptake rate of 2.28 mmol $gdw^{-1}hr^{-1}$ was considered as reported to be a realistic uptake rate [18]. In aerobic conditions with an oxygen uptake rate of 1000 mmol $gdw^{-1}hr^{-1}$, the model did not produce any significant metabolite apart from biomass (Fig 5A). It agrees with the experimental results as *L. kluyveri* in aerobic conditions produces insignificant amounts of ethanol due to weak Crabtree positive nature. In semi-aerobic conditions with an oxygen

uptake rate of 2 mmol $gdw^{-1}hr^{-1}$, the model produced ethyl acetate and ethanol apart from biomass (Fig 5B). The production of ethyl acetate as a major overflow has been reported in previous studies[2]. The ethanol concentration increased to a point until glucose was present in the medium after which it was consumed as a carbon source to fuel the growth. In anaerobic conditions with a very low oxygen uptake rate of 0.25 mmol $gdw^{-1}hr^{-1}$, the ethanol production increased significantly while ethyl acetate production was lowered compared to semi-aerobic conditions (Fig 5C). Similar to semi-aerobic conditions, the ethanol was used as the sole carbon source once glucose was exhausted in the medium. The reduction in the production of ethyl acetate both in complete aerobic conditions and in anaerobic conditions can be attributed to a reduction in the production of either of its precursors i.e. acetate and ethanol. Ethanol production was reduced in aerobic condition and acetate production was reduced in anaerobic conditions due to diversion of glycolytic flux towards ethanol. Rate of ethanol production increased and acetate production decreased with a decrease in oxygen uptake rate (Fig 5D). An optimal point was reached where ethyl acetate production was maximized against an uptake rate of about 3.125 mmol gdw^{-1} hr^{-1}.

The model was able to predict the growth rate and the respiratory quotient with good accuracy in comparison to the experiments. The ability to grow and ferment on melibiose differentiates the organism from the related yeasts[3]. The variation in growth rate while growing on different carbon sources can be attributed to the energetic cost and the benefit of metabolizing them. Ethyl acetate and ethanol production phenotype observed in *in silico* growth kinetics showed agreement with the reported overflow metabolite production. This model can be potentially leveraged to generate hypotheses in relation to the metabolism and physiology of the organism.

In silico single knockout analysis shows an agreement with experimental data.

Genome-scale metabolic models map the genotype of an organism i.e. gene information to the metabolic flux phenotypes of the cell. Each gene encodes for protein/subunit of a protein that catalyzes a specific reaction[22]. Each of the reactions is assigned a flux solution space through FBA. This enables the metabolic model to predict the changes in flux phenotypes with respect to perturbation in genotype i.e. single knockout or combinatorial knockout analysis. To gain a high degree of confidence, it was imperative to validate such predictive power of the model in comparison to experimentally established datasets. The genotype information was mapped to the reaction information through Boolean logic. For encoding a protein complex, either all genes or only some of them are required. This is well presented by the Boolean logic using AND or OR to map the genes to the protein. The viability should remain unaffected by the deletion of genes that do not have obligatory functions. Two large scale gene knockout studies in *S. cerevisiae* were used to evaluate the predictive power of the model. Comparative gene knockout analysis of the model with reference to the published datasets[47, 48] for homologous genes in *S. cerevisiae* was performed. Statistical analysis was performed in terms of sensitivity, specificity, F_1 score and Mathews correlation coefficient (MCC) (Fig 6A). FBA and MOMA[49] (Minimization of Metabolic Adjustment) were performed to assess the viability of the model based on single-gene knockouts. Due to the similar performance of both the approaches i.e. FBA and MOMA, only FBA was considered for any further discussion.

When cells grow in minimal medium where only carbon source, ammonium salts, phosphate and trace elements are present, it requires all the enzymes necessary for synthesizing amino acids, nucleotides, lipids and carbohydrates. In rich media, the cells are supplied with all the growth precursors in a platter so the cells relax the biosynthesis of those precursors [50]. In the genome-wide single-gene knockout analysis, the metabolic model supplemented with all growth precursors has significantly fewer essential genes i.e. true positives compared to growth on minimal media (Fig 6B). The total number of true essential genes were calculated with reference to the corresponding single knockout dataset of *S. cerevisiae* (Table S3, Supplementary Information) [28, 48]. Only 74.93 % (547) genes had a homolog in *S. cerevisiae*. The model had a sensitivity of 0.58 for rich media and 0.32 for the minimal media. It had almost equal specificity scores of 0.84 in rich and minimal media. For evaluating the performance of the model in classifying essential and non-essential genes, F_1 score and Matthews Correlation Coefficient (MCC) were used. F_1 score for predictions was 0.324 in rich media, 0.317 in minimal media with glucose. MCC for predictions was 0.28 in rich media, 0.162 in minimal media with glucose. The radar plot (Fig 6A) compared the statistical figures of merit when the single knockout analysis was performed in minimal and rich media. *S. cerevisiae* single-gene knockout database was used due to the unavailability of extensive knockout analysis for *L. kluyveri*. A considerable F_1 score and MCC being on the higher side for the model validated the confidence in the predictions.

Integration of reported metabolic routes in the model iPN730

In the previous published literature, important metabolic phenotypes for *L. kluyveri* have been reported. Of the four possible modes of pyrimidine degradation known, the uracil catabolism (URC) pathway has been uniquely characterized in *L. kluyveri*[1, 4, 51]. Locus URC1-6 and URC8 have been experimentally characterized to be an essential component of the URC pathway for uracil degradation. This enables *L. kluyveri* to grow on uracil and the intermediates of the degradation pathway as the sole nitrogen source. This ability is supposedly lost in *S. cerevisiae* after the whole genome duplication event [4]. In the URC pathway, the gene products of URC1 and URC6 degrade Uracil to 3-hydroxypropionate and ribosylurea through multiple steps that have not been characterized. The gene product of URC4 degrades ribosylurea to urea. The urea is further broken down by the action of gene products of URC3 and URC5 to ammonia and carbon dioxide. The ammonia/ammonium then replenishes the nitrogen pool for amino acid and nucleotide biosynthesis. This enables the organism to grow on uracil and urea as the sole nitrogen source. The whole pathway (Fig 7A) was incorporated into the metabolic model iPN730. The permeases/transporters for the intermediates of the pathway were incorporated into the model to enable their uptake. The ability of the model to show *in silico* growth on uracil and urea as a nitrogen source was evaluated. The model was able to incorporate nitrogen from uracil and urea into its metabolism. The uptake rate for uracil and urea per unit uptake flux of glucose in complete aerobic conditions was 1.009 mmol gdw^{-1} hr^{-1} and 0.261 mmol gdw^{-1} hr^{-1} respectively. With ammonium as the sole nitrogen source, the uptake rate per unit uptake flux of glucose was 0.526 mmol gdw^{-1} hr^{-1} (Fig 7B). We also found a high agreement between the *in silico* growth rate (0.097 hr^{-1}) and the experimental growth rate (0.0962 hr^{-1}) (Table S6, Supporting Information).

The ethyl acetate biosynthesis pathway was reconstructed in iPN730 which was reported earlier. Its production was highest in the semi-aerobic condition when ethanol production and acetate production were optimal. The production of ethyl acetate from its precursors was putatively facilitated by the cytoplasmic esterase. A transport reaction for ethyl acetate was incorporated into the metabolic model that enabled its secretion. The pathways for phenyl acetate, 2-methylbutanol and isobutanol were also present in the reconstructed metabolic model which are important industrial chemicals[52]. Their yields can be estimated *in silico* per unit mole of glucose consumed for metabolic engineering purposes.

Pyrimidine degradation pathway in *L. kluyveri* has been discussed in relation to anti-cancer drug design as understanding pyrimidine homeostasis directly affects the DNA replication machinery[4]. Although the pyrimidine catabolism pathway has been sufficiently characterized, iPN730 will give an opportunity to understand the genotype to flux phenotype relations and its interaction with other pathways at a systems level.

Conclusion

L. kluyveri's first genome-scale metabolic model reflects important phenotypes that have been experimentally characterized earlier. The phylogenetic similarity at the gene level and 18S rRNA level has been discussed. Model integration statistics reveal *L. kluyveri* to be closer to *S. cerevisiae* and *C. tropicalis.* There are 527 instances of duplicate genes in *S. cerevisiae* compared to *L. kluyveri* owing to the whole genome duplication event that took place in the evolutionary time course. The knowledge of the pyrimidine degradation pathway has been utilized for understanding the drug metabolism in the case of cancer cells particularly in relation to DNA replication machinery. The metabolic flux analysis of such pathways can be quintessential in understanding the dynamics of homologous pathways in higher eukaryotes. Due to the weak Crabtree positive metabolism, *L. kluyveri* is able to produce more biomass than *S. cerevisiae* that makes it a lucrative organism for the production of proteins and other growth associated commercially important macromolecules. The *in-silico* viability and the growth rate in multiple carbon sources also show good agreement with the experimental data. The single-gene knockout analysis also validated the predictive power of the model iPN730 for alterations in flux phenotypes in varying genetic backgrounds. We anticipate the use of this model for deriving predictions on the flux distribution in the metabolic network in varying growth conditions. This model will add up to the non-conventional yeast genome-scale metabolic reconstructions that will enable hypotheses generation for the exploration of diverse phenotypes of fungi. This will potentially lead to the discovery of new aspects of metabolism in lower eukaryotes and the design of engineering strategies to overproduce fine chemicals of industrial relevance from *L. kluyveri.*

Methods

Genome-scale reconstruction of *Lachancea kluyveri.*

KBase (US Department of Energy Knowledge Database)[53] workspace was used for annotation and generation of draft metabolic model iPN730. For the genome annotation using YGA pipeline, we have applied BDBH algorithm (Bidirectional Base Hit) with an E-value cutoff of 10^{-10} and Saccharomyces cerevisiae (S288C) as the base template model. The contigs were uploaded in a multi-FASTA format into the server. The homology table generated from the BDBH algorithm for proteome comparison can be used to predict possible duplication, inversion, deletion, insertion or synteny loci between the organisms (Table S4, Supplementary Information). Build Fungal Model was used to reconstruct the genome-scale metabolic model using 13 previously published fungal metabolic models as templates. The template metabolic models were *S. cerevisiae*-iMM904[38], *Aspergillus oryzae*-iWV1314[54], *Mucor circinelloides*-iWV1213[55], *Y. lipolytica*-iNL895[27], *Scheffersomyces stipitis*-iSS884[56], *Penicillium rubensi*-iAL1006[57], *Eremothecium gosypii*-iRL766, *Kluyveromyces lactis*-iOD907[25], *Komagataella phaffii*-iLC915[58], *Neurospora crassa*-iJDZ836[59], *C. tropicalis*-iCT646[39], *C. glabrata*-iNX804[55], *A. terreus*-iJL1454[60]. Among them, *S. cerevisiae*-iMM904 was chosen as the template for the reconstruction. COBRApy[34] was used in Spyder, Anaconda (Python 3.6) for reading and manipulating the draft model in Systems Biology Markup Language (SBML) format. COBRA Toolbox v3.0 was used in MATLAB R2017b (Mathworks, Natick, MA, USA) for identifying the inconsistencies in the model. Gurobi version 8.0 (Academic license version) and GLPK solver were used for linear programing for Flux Balance Analysis in both MATLAB and Python. The metabolite formula and charge were added using PubChem, CHEBI (Chemical Entities of Biological Interest) and BiGG database. The reaction bounds by default were reversible. The reaction bounds were corrected using iMM904 as a reference and similar reactions were mapped. Remaining reactions were manually corrected with reference from MetaCyc, KEGG and previously published reaction reversibility database[35]. KEGG pathway database was used as a reference to find gaps in the biomass precursor synthesis pathways. Inferred reactions were added wherever necessary to enable biosynthesis of all precursors[61]. Due to insufficient information on the biomass composition of *L. kluyveri,* the biomass equation for *S. cerevisiae* (iMM904)[24] was used for all purposes. A confidence score was assigned to each reaction considered in the model. The confidence score was assigned based on the existence of a GPR for the reaction and flux through it. A high confidence score of ‘4’ was assigned to the reactions which had a GPR assignment along with significant flux through the network calculated by flux balance analysis. The confidence score of ‘4’ was also assigned for exchange reactions necessary for growth on defined media. A score of ‘3’ was assigned for reactions having significant flux through them but no GPR assigned for them. A score of ‘2’ was assigned for reactions having GPR assignment but no flux through them for growth in defined medium. A low confidence score of ‘1’ was assigned for reactions having no GPR assignment or flux through them.

Model validation through flux balance analysis.

Flux Balance Analysis (FBA) was performed using Gurobi solver in Python 3.6 (Spyder, Anaconda). *In silico* media was assigned to the model for FBA by assigning experimentally reported uptake rates to the exchange reactions. The glucose and oxygen uptake rates were assigned as per experimental data. The bounds for ammonium ion and trace metals were set to very high values i.e. 1000 mmol gdw^{-1} hr^{-1} to

remove any limitation. *In silico* growth rate and respiratory quotient were compared to experimental data based on the Pearson correlation coefficient. We used the 'jointplot' function in the seaborn visualization and analysis library in python 3.7 for Fig 3. The Pearson's correlation coefficient was calculated and the p-value corresponding to the hypothesis "The experimental and in silico values are positively correlated" is given. The respiratory quotient (RQ) was calculated as:

See Formula 1 in the supplemental files.

The information regarding the growth on various carbon sources was procured from Westerdijk Fungal Biodiversity Institute's MycoBank[44] database. The data regarding the growth of the strain on the mentioned carbon sources can be accessed from the BioloMICS database of MycoBank using *"Lachancea kluyveri*" as the identifier. The *in-silico* viability was qualitatively compared with the reported experimental growth on glucose, galactose, maltose, sucrose, trehalose, melibiose, arabinose, ethanol and citrate. The *in-silico* growth rate was simulated for glucose, galactose, maltose, sucrose, melibiose, ethanol and glycerol in anaerobic and aerobic conditions. For aerobic condition, an oxygen uptake rate of 1000 mmol gdw^{-1} hr^{-1} was considered and for anaerobic condition, an uptake rate of 0.25 mmol gdw^{-1} hr^{-1} was considered.

Dynamic flux balance analysis was implemented for the model iPN730 in COBRA Toolbox v3.0 (MATLAB, USA). The glucose uptake rate was kept fixed at 2.28 mmol gdw^{-1} hr^{-1} as measured in previous studies[18]. The uptake rate of oxygen was varied from 1000 mmol gdw^{-1} hr^{-1} i.e. aerobic condition (Glucose limited regime) to 0.25 mmol gdw^{-1} hr^{-1} i.e. anaerobic condition (Oxygen limited regime). An uptake rate just below the optimal uptake rate in the aerobic condition of 3 mmol gdw^{-1} hr^{-1} was considered as the semi-aerobic condition. The parameters considered for the dynamic flux balance (Table 2) were chosen based on experimental conditions in which batch fermentation experiments were performed[2].

The DFBA algorithm applies Flux Balance Analysis (FBA) at regular intervals i.e. DFBA sampling intervals on the model and assumes that the substrate uptake rate remains constant throughout the fermentation[29]. The *in-silico* media composition was recalculated based on the uptake and growth rate after every sampling interval. The following equations define the growth kinetics, substrate uptake and product formation in an *in-silico* batch fermentation:

See Formula 2 in the supplemental files.

X(t), S(t) and P(t) denote the concentration of biomass, substrate and the product respectively over the course of *in silico* batch fermentation. μ is the specific growth rate in hr^{-1}. v_{EX_S} and v_{EX_P} are the uptake rate and the production rate for substrate and products respectively. The products considered for dFBA were ethanol and ethyl acetate. To explain the production of ethyl acetate, the fluxes associated with the precursors compounds ethanol and acetate were analyzed in oxygen-limited conditions. The alterations in the production fluxes for ethanol i.e. R_ALCD2ir_c0 (Alcohol Dehydrogenase 1.1.1.1) and acetate i.e.

rxn00507_c0 (Aldehyde Dehydrogenase, EC 1.2.1.3) with the change in the uptake rate of the oxygen was evaluated through FBA. For example, model.reactions. R_ALCD2ir_c0.flux was used to access the flux through the production reaction of ethanol in COBRApy.

Growth experiments on various carbon sources

Lachancea kluyveri NRRL 12651 strain was used in all the experiments. The yeast was grown in Yeast Extract-Peptone-Glucose (YEPD) media (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) overnight at 30^0 C shaking at 240 RPM. The cells were washed thrice in Yeast Nitrogen Base buffer (0.67%) to remove any remaining media with centrifugation at 3000 RPM for 3 minutes. The growth experiments were conducted in shake flask. The media containing different carbon sources comprised of Yeast Nitrogen Base with Ammonium sulphate and without amino acids (Difco, BD); 6.7 g/L, 20 g/L of respective carbon source. The carbon sources used in the experiments were Maltose (Merck), Sucrose (Merck), and Galactose (SRL). The cells were diluted to 0.2 OD in the growth media. The optical density was measured at 600 nm using Nanodrop ONE (Thermo Scientific, USA). The substrate and extracellular product concentrations were measured in High Performance Liquid Chromatography (Agilent Technologies, HPLC 1260 Infinity) equipped with an Aminex HPX-87H ion exchange column (Biorad, USA) where 5mM H_2SO_4 was used as mobile phase and flowrate was 0.6ml/minute. Briefly, 20 uL syringe filter purified sample was injected in the column. The specific growth rate ( ) was calculated as,

See Formula 3 in the supplemental files.

For growth on uracil as the sole nitrogen source, we used a concentration of 0.1%(w/v) [51]. The media contained 1.7 g/L Yeast Nitrogen Base without Ammonium Sulphate and without amino acids.

Single Knockout Landscape and Statistical Analysis.

Genome-scale metabolic models link the metabolic phenotypes to the genotype of the organism through Gene-Protein-Reaction (GPR) association. This enables the prediction of the alterations in the metabolic phenotype of the organism in response to genetic perturbations. To gain confidence in the accuracy of the model, it is imperative to validate the predictions with reference to experimental data. Unfortunately, such a genome-scale single gene deletion dataset was not available for *L. kluyveri.* Hence, for this purpose, the *S. cerevisiae* gene essentiality dataset was considered due to the evolutionary closeness between them and high confidence in the dataset[47, 48]. Gene essentiality information for *S. cerevisiae* genes was transferred to homologous genes in *L. kluyveri. In silico* single-gene knockout was performed on iPN730 by constraining the flux to zero for the reactions mapped from the genes by the GPR association information. For this, cobra.flux_analysis.single_gene_deletion() function was used in COBRApy in Python 3.6 (Anaconda, USA) and Gurobi solver was used for the optimization. The viability of the cell in terms of the growth rate was estimated by FBA for each *in silico* gene deletion in the model. The *in-silico* essentiality of the genes were compared with experimental data mapped from the gene deletion studies on *S. cerevisiae* and the statistics of the predictions were calculated as follows:

See Formula 4 in the supplemental files.

Where TP: True Positives; FP: False Positives; TN: True Negative; FN: False Negative. The single-gene deletion analysis was performed in different *in silico* media. Aerobic conditions in which original experiments for *S. cerevisiae* were performed were assumed for the analysis. For rich media i.e. YPD, the model was allowed to uptake all the trace elements, nitrogen sources, 20 amino acids, all the nucleotides and carbohydrates. Whereas in the minimal media, the model was allowed to uptake only glucose (carbon source), ammonium ions (nitrogen source) and trace elements.

Reconstructing reported pathways in iPN730

In *L. kluyveri* (previously known as *Saccharomyces kluyveri* in many studies) metabolic pathways have been reported with a particular focus on pyrimidine metabolism and overflow metabolism. URC pathway for pyrimidine degradation allowed the model iPN730 to grow on uracil, urea, and intermediates of the pathway as the sole nitrogen source. The pathway information on the URC pathway was obtained from the literature[51]. The metabolites and reactions for this pathway were incorporated into the model in COBRApy using reaction.add_metabolites() and model.add_reactions() respectively. URC6 gene locus encodes uracil phosphoribosyl transferase (EC: 2.4.2.9) which breaks down uracil to phosphoribosyl and 3-hydropropionate. URC3 and URC5 encode urea amidolyase (EC 3.5.1.54) which breaks down urea to ammonia and carbon dioxide. To check for the uptake of uracil and urea as a sole nitrogen source, the uptake for ammonium ion was restricted. The uptake rate for uracil and urea was set to 1000 mmol gdw^{-1} hr^{-1}. The carbon source was kept as glucose while setting up an oxygen uptake rate very high for aerobic metabolism. FBA was performed to check for any *in silico* growth.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The metabolic model and associated data are supplied in the additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors contributions

PN and AG planned the project. PN, PP, and MD conducted the analysis. PP, PN, MD and AG wrote the paper.

Acknowledgment

This work was supported by a research grant from the Department of Science and Technology (No. ECR/2016/001096) and DBT-Ramalingaswami Re-entry fellowship (No.BT/RLF/Re-entry/06/2013) to A.G. PP appreciates the support from DST-Inspire. MD thanks the support from CSIR.

Additional files:

Table S1: Literature reported experimental data for growth kinetics, product formation, and substrate uptake.

Table S2: Confidence scores on the reactions in the model based on GPR assignment and flux through the reactions calculated by FBA.

Table S3: Single-gene knockout analysis of *L. kluyveri* in comparison to *S. cerevisiae.*

Table S4: BDBH proteome mapping table between *S. cerevisiae* and *L. kluyveri.*

Table S5: Kbase/modelSEED reaction database.

Table S6: Batch growth experiment to validate the in silico growth simulations on various carbon sources and Uracil.

Metabolic Model in SBML format: iPN730.xml

Abbreviations

GEM- Genome-Scale Metabolic Model

FBA-Flux Balance Analysis

COBRA-Constraint Based Reconstruction and Analysis

GPR-Gene-Protein-Reaction association

DFBA-Dynamic Flux Balance Analysis

References

1. Andersen G, Björnberg O, Polakova S, Pynyaha Y, Rasmussen A, Møller K, et al. A Second Pathway to Degrade Pyrimidine Nucleic Acid Precursors in Eukaryotes. J Mol Biol. 2008.
2. Møller K, Christensen B, Förster J, Pikšur J, Nielsen J, Olsson L. Aerobic glucose metabolism of Saccharomyces kluyveri: Growth, metabolite production, and quantification of metabolic fluxes. Biotechnol Bioeng. 2002.
3. Naumova ES, Serpova E, Korshunova I, Naumov GI. Molecular genetic characterization of the yeast Lachancea kluyveri. Microbiology. 2007.
4. Beck H, Dobritzsch D, Piškur J. Saccharomyces kluyveri as a model organism to study pyrimidine degradation. FEMS Yeast Research. 2008.
5. Gojkovic Z, Paracchini S, Piskur J. A New Model Organism for Studying the Catabolism of Pyrimidines and Purines. 2011.
6. Óhéigeartaigh SS, Armisén D, Byrne KP, Wolfe KH. Systematic discovery of unannotated genes in 11 yeast species using a database of orthologous genomic segments. BMC Genomics. 2011.
7. Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010;5:93–121.
8. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28:245–8. doi:10.1038/nbt.1614.
9. Forster J, Famili I, Fu P, Palsson BÃ, Nielsen J. Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network. Genome Res. 2003;13:244–53. doi:10.1101/gr.234503.
10. O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. Cell. 2015.
11. Goffeau A, Barrell G, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. Science (80-). 1996.
12. Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, Adler C, et al. Genetic and physical maps of Saccharomyces cerevisiae. Nature. 1997.
13. Shi S, Chen T, Zhang Z, Chen X, Zhao X. Transcriptome analysis guided metabolic engineering of Bacillus subtilis for riboflavin production. Metab Eng. 2009.
14. Shabestary K, Hudson EP. Computational metabolic engineering strategies for growth-coupled biofuel production by Synechocystis. Metab Eng Commun. 2016.
15. Brochado AR, Matos C, Møller BL, Hansen J, Mortensen UH, Patil KR. Improved vanillin production in baker's yeast through in silico design. Microb Cell Fact. 2010.
16. Li Q, Sun Z, Li J, Zhang Y. Enhancing beta-carotene production in Saccharomyces cerevisiae by metabolic engineering. FEMS Microbiology Letters. 2013.

17. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature. 2013;496:528–32.
18. Møller K, Bro C, Piškur J, Nielsen J, Olsson L. Steady-state and transient-state analyses of aerobic fermentation in Saccharomyces kluyveri. In: FEMS Yeast Research. 2002.
19. Rasmussen A, Lv Y, Schnackerz KD, Piškur J. A new expression vector for production of enzymes in the yeast Saccharomyces (Lachancea) kluyveri. Nucleosides, Nucleotides and Nucleic Acids. 2011.
20. Acevedo A, Conejeros R, Aroca G. Ethanol production improvement driven by genome-scale metabolic modeling and sensitivity analysis in Scheffersomyces stipitis. PLoS One. 2017.
21. Ng CY, Jung M, Lee J, Oh M-K. Production of 2,3-butanediol in Saccharomyces cerevisiae by in silico aided metabolic engineering. Microb Cell Fact. 2012;11:68.
22. Duarte NC, Herrgård MJ Markus J., Palsson BØ. Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. Genome Res. 2004;14:1298–309. http://genome.cshlp.org/content/14/7/1298.abstract.
23. Herrgård MJ, Lee BS, Portnoy V, Palsson B. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. Genome Res. 2006.
24. Mo M, Palsson BØ, Herrgård M. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst Biol. 2009;3:37. http://www.biomedcentral.com/1752-0509/3/37.
25. Dias O, Pereira R, Gombert AK, Ferreira EC, Rocha I. iOD907, the first genome-scale metabolic model for the milk yeast Kluyveromyces lactis. Biotechnol J. 2014.
26. Saitua F, Torres P, Pérez-Correa JR, Agosin E. Dynamic genome-scale metabolic modeling of the yeast Pichia pastoris. BMC Syst Biol. 2017.
27. Loira N, Dulermo T, Nicaud JM, Sherman DJ. A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. BMC Syst Biol. 2012.
28. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science (80-). 1999;285:901–6.
29. Mahadevan R, Edwards JS, Doyle III FJ. Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. Biophys J. 2002;83:1331–40. doi:http://dx.doi.org/10.1016/S0006-3495(02)73903-9.
30. Proux-Wéra E, Armisén D, Byrne KP, Wolfe KH. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. BMC Bioinformatics. 2012.
31. Souciet JL, Dujon B, Gaillardin C, Johnston M, Baret P V., Cliften P, et al. Comparative genomics of protoploid Saccharomycetaceae. Genome Res. 2009.
32. Wendland J, Walther A. Genome evolution in the eremothecium clade of the saccharomyces complex revealed by comparative genomics. G3 Genes, Genomes, Genet. 2011.
33. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 2016.

34. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol. 2013.
35. Stelzer M, Sun J, Kamphans T, Fekete SP, Zeng AP. An extended bioreaction database that significantly improves reconstruction and analysis of genome-scale metabolic networks. Integr Biol. 2011.
36. Orlean P. Architecture and biosynthesis of the Saccharomyces cerevisiae cell wall. Genetics. 2012.
37. Beld J, Lee DJ, Burkart MD. Fatty acid biosynthesis revisited: Structure elucidation and metabolic engineering. Molecular BioSystems. 2015.
38. Zomorrodi AR, Maranas CD. Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data. BMC Syst Biol. 2010.
39. Mishra P, Park GY, Lakshmanan M, Lee HS, Lee H, Chang MW, et al. Genome-scale metabolic modeling and in silico analysis of lipid accumulating yeast Candida tropicalis for dicarboxylic acid production. Biotechnol Bioeng. 2016.
40. Mishra P, Lee NR, Lakshmanan M, Kim M, Kim BG, Lee DY. Genome-scale model-driven strain design for dicarboxylic acid production in Yarrowia lipolytica. BMC Syst Biol. 2018.
41. Zhou N, Bottagisi S, Katz M, Schacherer J, Friedrich A, Gojkovic Z, et al. Yeast-bacteria competition induced new metabolic traits through large-scale genomic rearrangements in Lachancea kluyveri. FEMS Yeast Res. 2017.
42. Møller K, Olsson L, Piškur J. Ability for anaerobic growth is not sufficient for development of the petite phenotype in Saccharomyces kluyveri. J Bacteriol. 2001.
43. Hagman A, Piškur J. A study on the fundamental mechanism and the evolutionary driving forces behind aerobic fermentation in yeast. PLoS One. 2015.
44. Robert V, Vu D, Amor ABH, van de Wiele N, Brouwer C, Jabas B, et al. MycoBank gearing up for new horizons. IMA Fungus. 2013.
45. Sprague GF, Cronan JE. Isolation and characterization of Saccharomyces cerevisiae mutants defective in glycerol catabolism. J Bacteriol. 1977.
46. Turcotte B, Liang XB, Robert F, Soontorngun N. Transcriptional regulation of nonfermentable carbon utilization in budding yeast. FEMS Yeast Research. 2010.
47. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science (80-). 1999.
48. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002;418:387–91.
49. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A. 2002;99:15112–7.
50. Österlund T, Nookaew I, Bordel S, Nielsen J. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. BMC Syst Biol. 2013.

51. Rasmussen AA, Kandasamy D, Beck H, Crosby SD, Björnberg O, Schnackerz KD, et al. Global expression analysis of the yeast Lachancea (saccharomyces) kluyveri reveals new URC genes involved in pyrimidine catabolism. Eukaryot Cell. 2014.
52. Nielsen J, Larsson C, van Maris A, Pronk J. Metabolic engineering of yeast for production of fuels and chemicals. Current Opinion in Biotechnology. 2013.
53. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol. 2018;36:566–9. doi:10.1038/nbt.4163.
54. Vongsangnak W, Olsen P, Hansen K, Krogsgaard S, Nielsen J. Improved annotation through genome-scale metabolic modeling of Aspergillus oryzae. BMC Genomics. 2008.
55. Xu N, Liu L, Zou W, Liu J, Hua Q, Chen J. Reconstruction and analysis of the genome-scale metabolic network of Candida glabrata. Mol Biosyst. 2013.
56. Liu T, Zou W, Liu L, Chen J. A constraint-based model of Scheffersomyces stipitis for improved ethanol production. Biotechnol Biofuels. 2012.
57. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. PLoS Comput Biol. 2013;9.
58. Caspeta L, Shoaie S, Agren R, Nookaew I, Nielsen J. Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. BMC Syst Biol. 2012.
59. Dreyfuss JM, Zucker JD, Hood HM, Ocasio LR, Sachs MS, Galagan JE. Reconstruction and Validation of a Genome-Scale Metabolic Model for the Filamentous Fungus Neurospora crassa Using FARM. PLoS Comput Biol. 2013.
60. Liu J, Gao Q, Xu N, Liu L. Genome-scale reconstruction and in silico analysis of Aspergillus terreus metabolism. Mol Biosyst. 2013.
61. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotech. 2010;28:977–82. http://dx.doi.org/10.1038/nbt.1672.

Tables

Table 1: Comparative analysis of *L. kluyveri* (iPN730) with other reported genome-scale metabolic reconstructions such as *S. cerevisiae* (iMM904), *C. tropicalis* (iCT646), *Y. lipolytica* (iYL647) and *K. lactis* (iOD907).

	iPN730This study	iMM904[38]	iCT646[39]	iYL647[40]	iOD907[25]
Reactions	1235	1577	945	1347	1867
Metabolites	1179	1226	712	1119	1476
Genes	730	905	646	647	907
Compartments	8	8	4	8	4

Table 2: Parameters considered for dynamic flux balance analysis using COBRA Toolbox v3.0 in MATLAB 2017(b).

dFBA parameters	Parameter values
Initial Concentration of Glucose	20 g/L
Initial Biomass Concentration	0.05 g/L
Fermentation Time	50 hours
dFBA sampling interval	0.5 hours

Figures

Figure 1

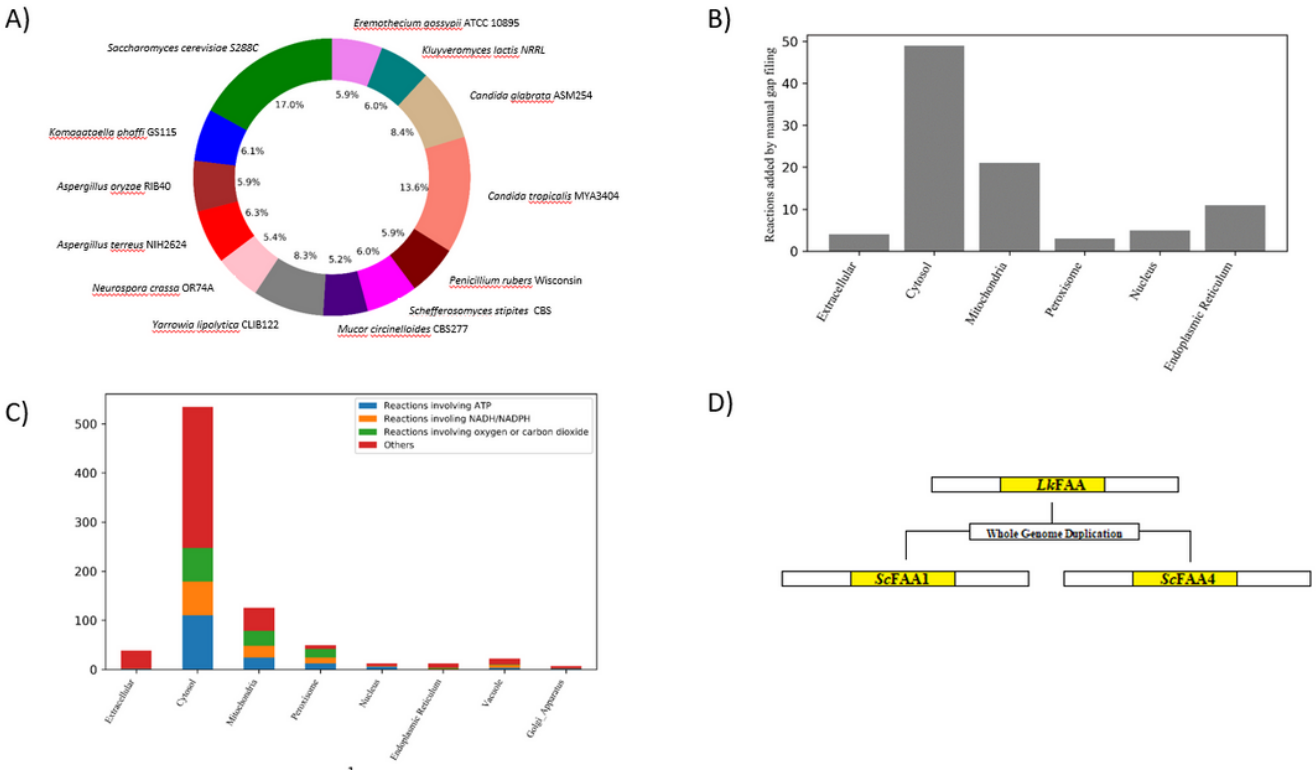

Figure 1

Genome-scale metabolic reconstruction of Lachancea kluyveri. (A) Distribution of the orthologs of Lachancea kluyveri protein coding genes associated with metabolism in various fungi in percentage of total. (B) The number of reactions added to the model by manual gap filling with reference to MetaCyc and KEGG contained in various compartments. (C) Irreversible reactions in the iPN730 distributed across various compartments. The colour indicates the reactions belonging specific categories involving ATP, NADH/NADPH or metabolic gases like oxygen or carbon dioxide. (D) Duplication of the prospective FAA1 (Long chain fatty acyl-CoA synthetase) in L. kluyveri to FAA1 and FAA4 in S. cerevisiae as an example of duplicate homologs.

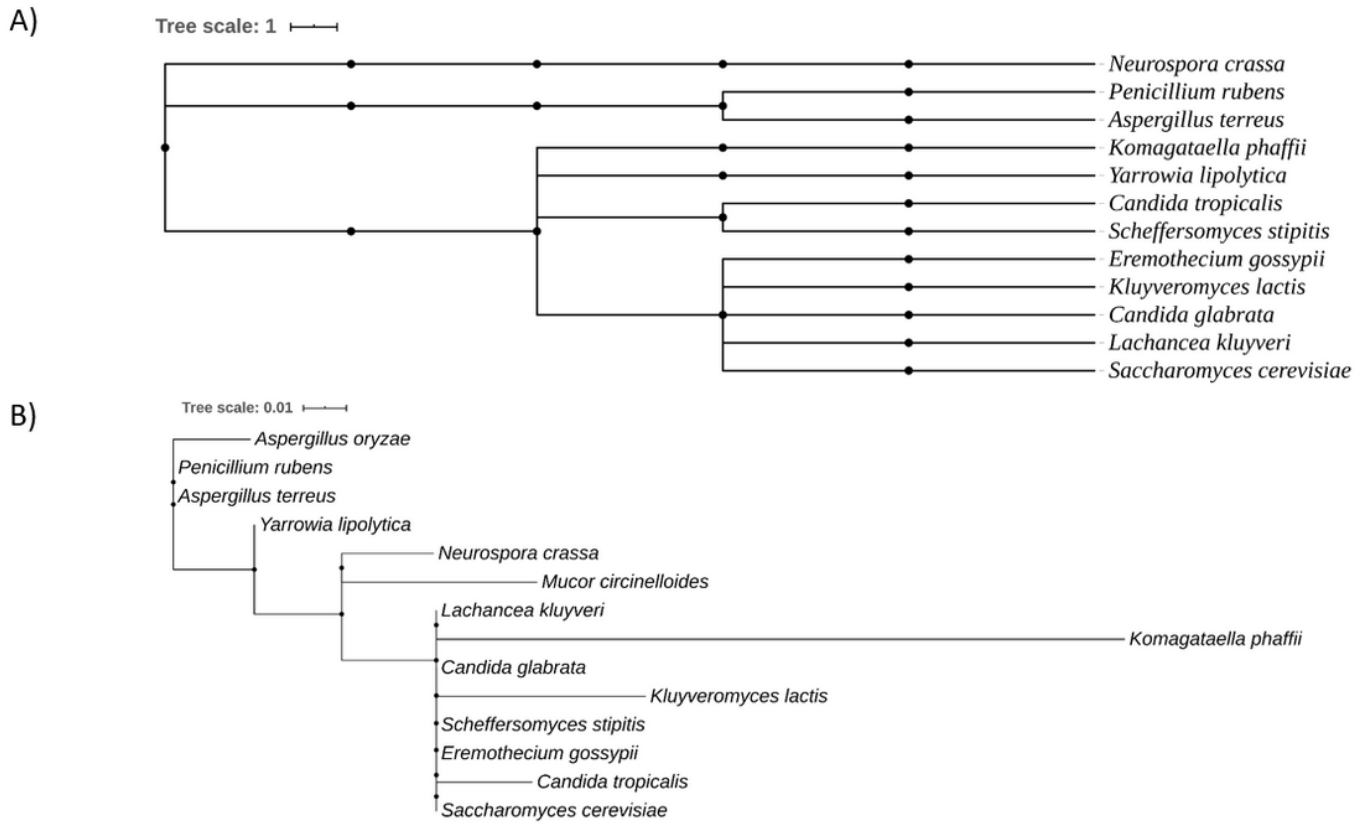

Figure 2

Phylogenetic analysis of the yeasts species involved in reconstruction at 18S rRNA and gene level. (A) The phylogenetic tree was built using NCBI Taxonomy browser and iTOL visualization based on 18S rRNA. (B) The phylogenetic tree was built using REALPHY server and iTOL visualization based on whole genome sequence.

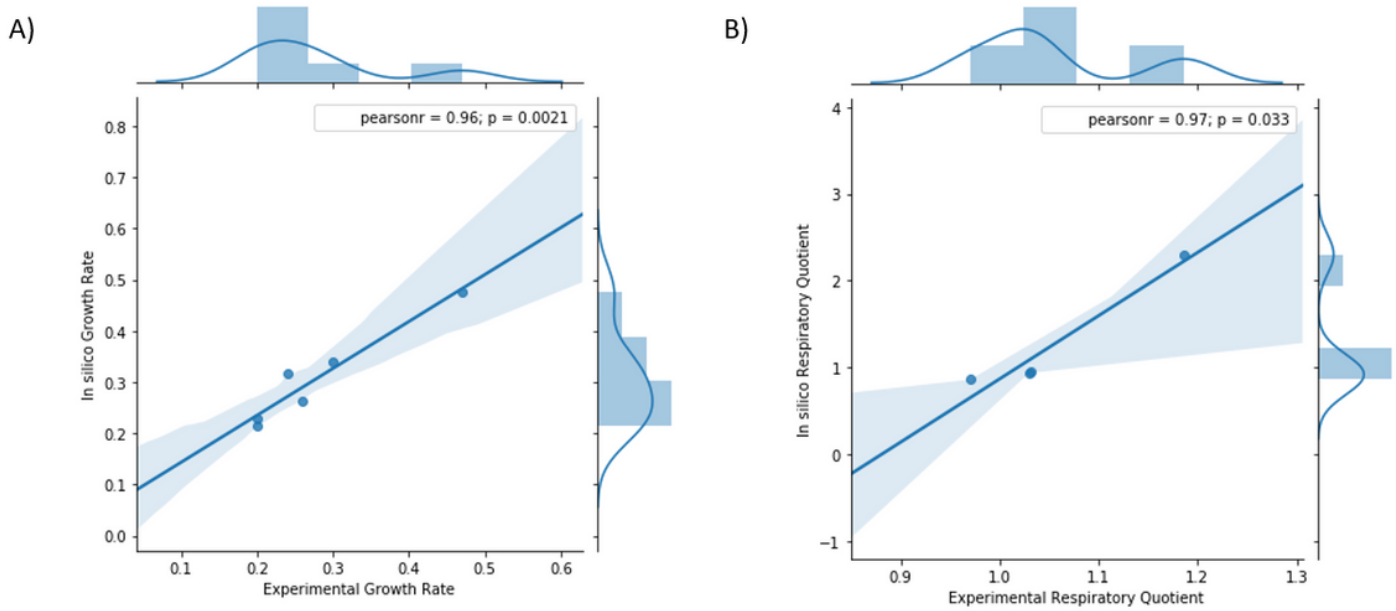

Figure 3

Computational predictions of growth characteristics show good agreement with experimental reports. (A) The correlation between the experimental growth rate observed in previous studies and the in silico growth rate from Flux Balance Analysis (FBA) of iPN730. (B) The correlation between the experimental respiratory quotient (moles of carbon dioxide released per mole of oxygen consumed) and the in silico values obtained from iPN730.

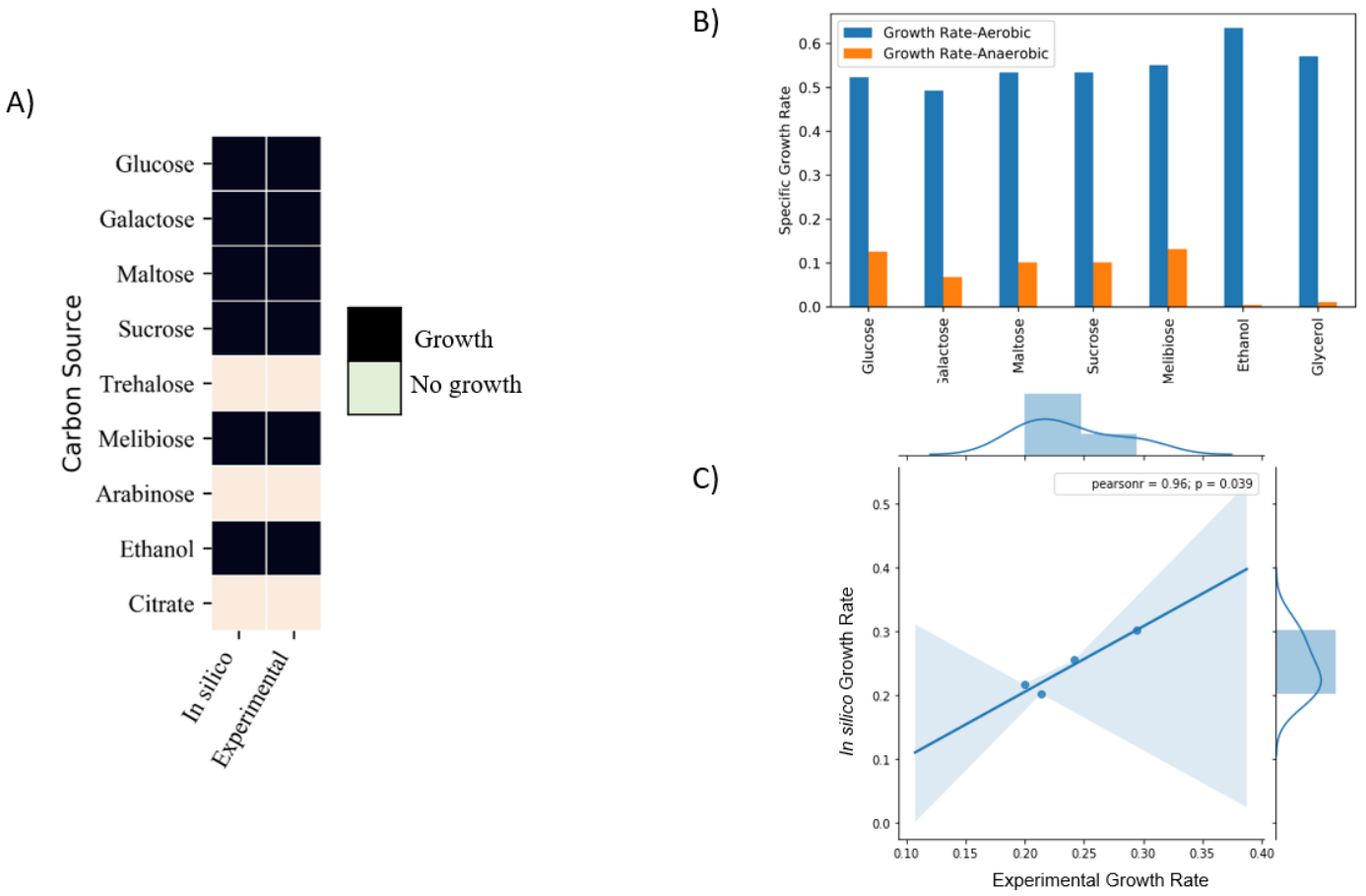

Figure 4

Viability and growth on various carbon sources in comparison to experimental reports. (A) Comparative analysis of in silico viability on multiple carbon sources and the reported data in MycoBank and other literature. (B) The in silico specific growth rate (hr-1) simulated by FBA of L. kluyveri on different carbon sources in both aerobic and anaerobic conditions. (C) The correlation between the experimental growth rate and in silico growth rate on various carbon sources.

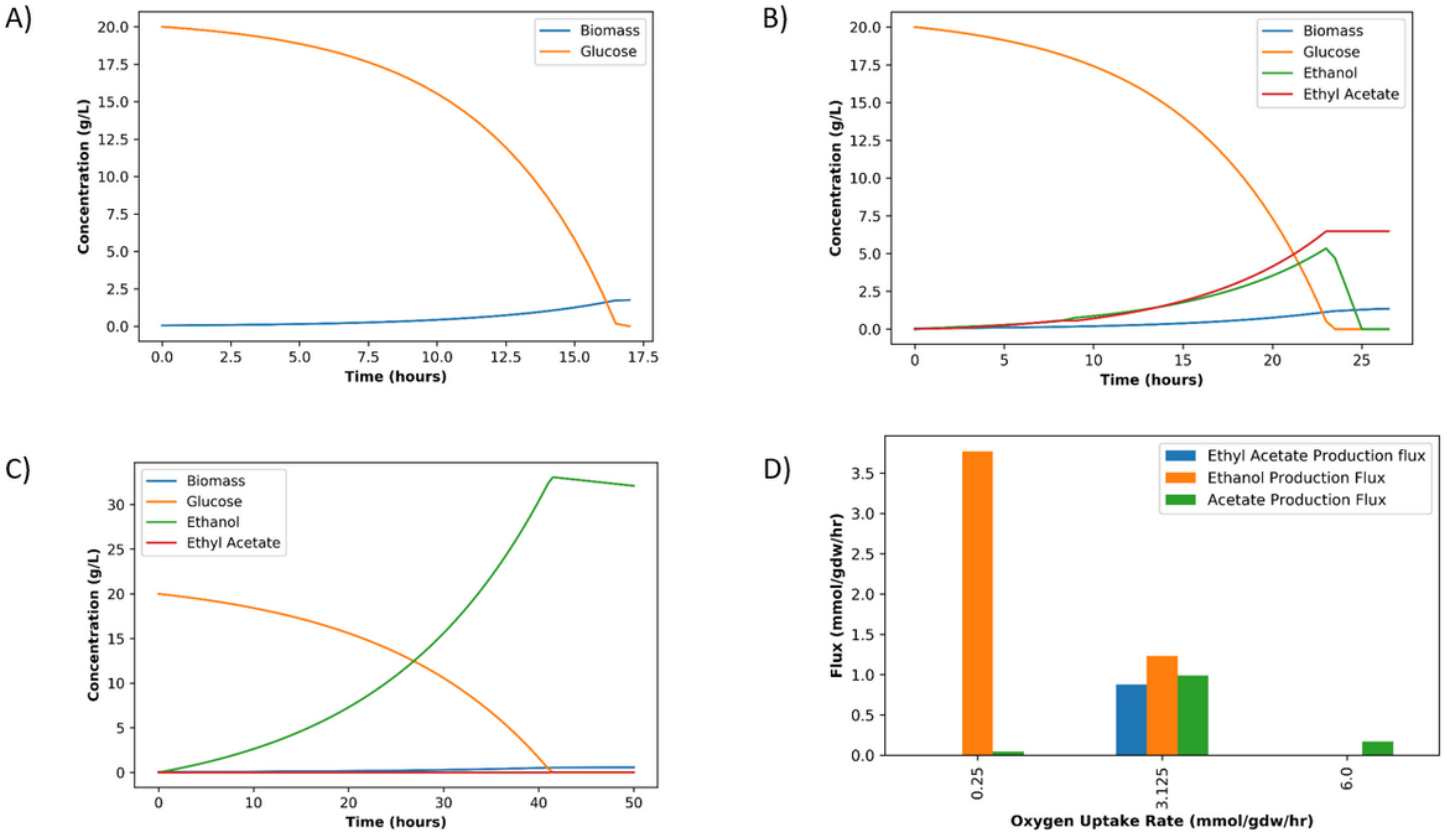

Figure 5

The dynamic flux balance analysis of iPN730 in oxygen limited conditions. DFBA on glucose as carbon source and oxygen uptake in (A) Aerobic (B) Semi-aerobic and (C) Anaerobic regime. (D) The shift from respiratory metabolism (6 mmol gdw-1 hr-1) to fermentative metabolism (0.25 mmol gdw-1 hr-1) shows the variation in production of acetate and ethanol which are precursor for the major overflow metabolite ethyl acetate.

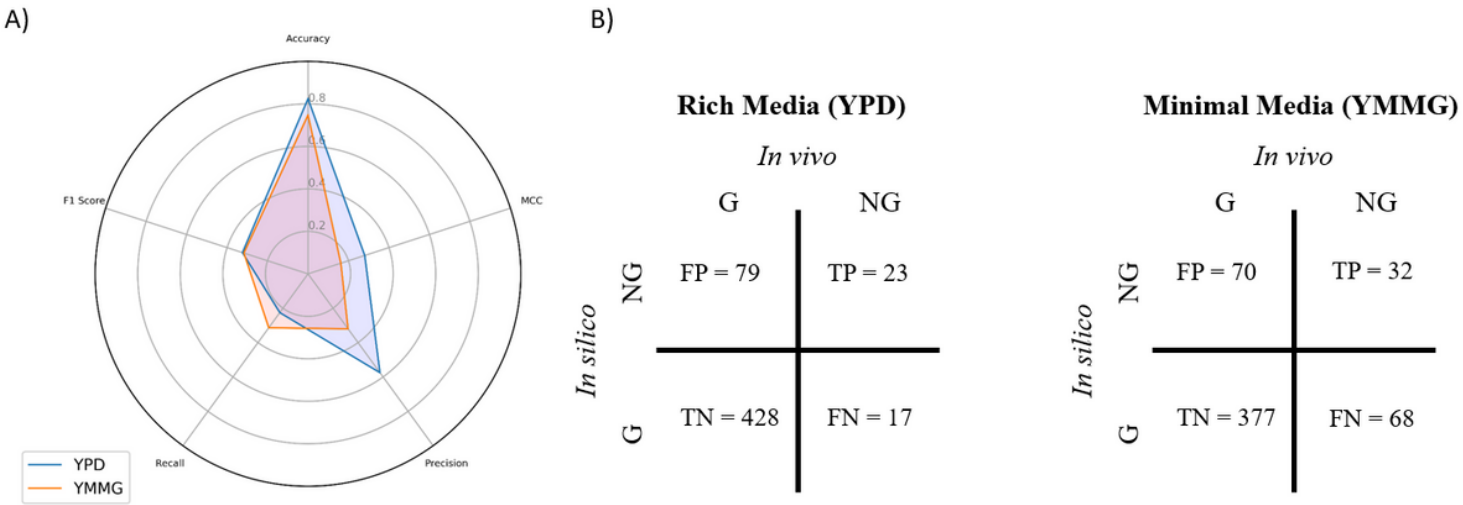

Figure 6

Single knockout analysis of viability in rich and minimal media. (A) Radar plot for comparative analysis of the model in rich and minimal media for classifying essential and non-essential genes over multiple statistical parameters i.e. F1 score, MCC, Precision, Recall and Accuracy. (B) Cross comparison between ground truth (Saccharomyces cerevisiae gene essentiality data) and the model predictions in rich and minimal media. NG and G stands for No Growth and Growth respectively for each in silico knockout.

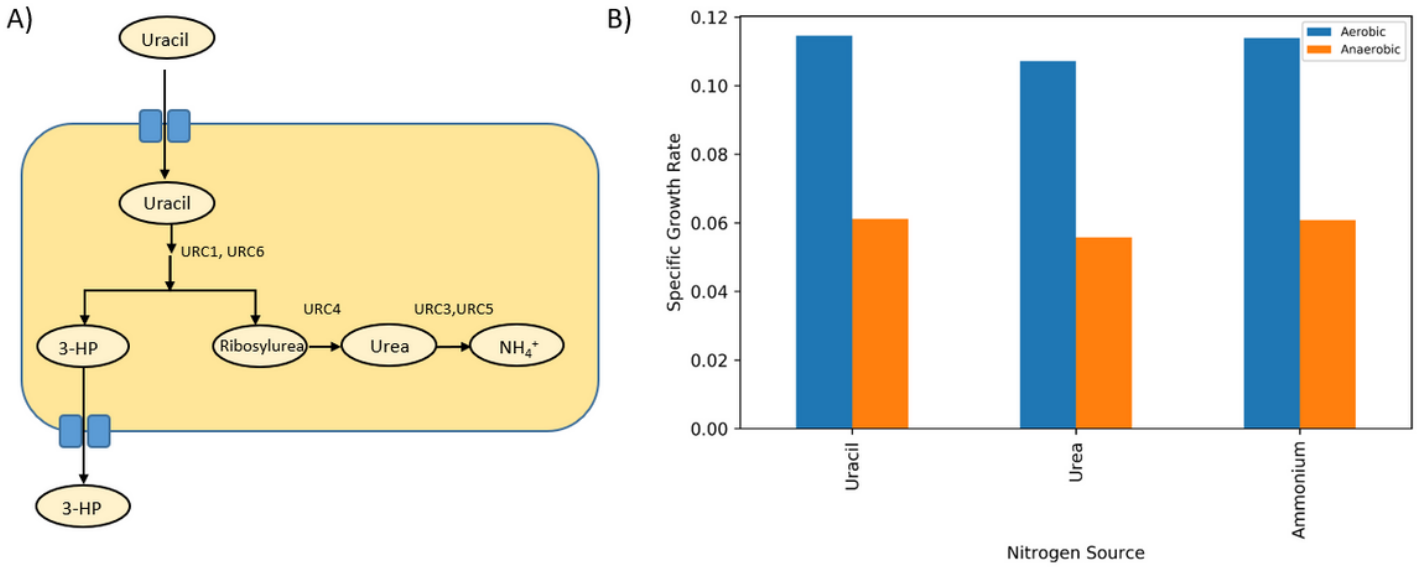

Figure 7

URC pathway for pyrimidine degradation in Lachancea kluyveri. (A) URC pathway in L. kluyveri reported in literature. This enables the organism to assimilate Uracil and other pathway intermediates as it's sole nitrogen source. (B) The fitness in terms of specific growth rate for iPN730 simulated on uracil only or urea only or ammonium only as the sole nitrogen source.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- formula2.PNG
- formula1.PNG
- LkiPN730.xml
- SupportingInformation.xlsx
- formula3.PNG
- formula4.PNG