

RESEARCH

Genome-scale model reconstruction of the methylotrophic yeast *Ogataea polymorpha*

Ulf W Liebal¹, Brigida A Fabry¹, Aarthi Ravikrishnan^{1,2}, Constantin VL Schedel¹, Simone Schmitz¹, Lars M Blank^{1*} and Birgitta E Ebert^{1,3,4}

Abstract

Background: *Ogataea polymorpha* is a thermotolerant, methylotrophic yeast with significant industrial applications. While previously mainly used for protein synthesis, it also holds promise for producing platform chemicals. *O. polymorpha* has the distinct advantage of using methanol as a substrate, which could be potentially derived from carbon capture and utilization streams. Full development of the organism into a production strain and estimation of the metabolic capabilities require additional strain design, guided by metabolic modeling with a genome-scale metabolic model. However, to date, no genome-scale metabolic model is available for *O. polymorpha*.

Results: To overcome this limitation, we used a published reconstruction of the closely related yeast *Komagataella phaffii* as a reference and corrected reactions based on KEGG and MGOB annotation. Additionally, we conducted phenotype microarray experiments to test the suitability of 190 substrates as carbon sources. Over three-quarter of the substrate use was correctly reproduced by the model and 27 new substrates were added, that were not present in the *K. phaffii* reference model.

Conclusion: The developed genome-scale metabolic model of *O. polymorpha* will support the engineering of synthetic metabolic capabilities and enable the optimization of production processes, thereby supporting a sustainable future methanol economy.

Keywords: Biotechnology; Genome-Scale Metabolic Model; Metabolic Reconstruction; Metabolic Engineering; COBRA; Methylotrophy

*Correspondence:

lars.blank@rwth-aachen.de

¹Institute of Applied

Microbiology-iAMB, Aachen
Biology and Biotechnology-ABBt,
RWTH Aachen University,
Worringer Weg 1, 52074 Aachen,
Germany

Full list of author information is
available at the end of the article

Background

Ogataea polymorpha (*Hansenula polymorpha*; *Pichia angusta*) is a widely used yeast for biotechnological applications. It is environmentally ubiquitous and has been isolated, among others, from orange juice, maize meal, and insect guts [1]. The ability to metabolize methanol stems from its close association to the phyllosphere [2]. *O. polymorpha* has some unique features among methylotrophic yeasts such as high growth rate, temperature tolerance, and nitrate fixation [3] and has been optimized to express peptides and proteins [4]. A particular benefit for biotechnological applications is the lack of byproducts even on high glucose feeds [4, 5]. Furthermore, the conversion of methanol to succinic acid via methylotrophic yeasts was found to provide a competitive alternative to conventional petrochemical approaches in a computational comparison of various microbial carbon fixation strategies [6].

Methylotrophy in yeast is restricted to a single multigenus clade, with its most prominent members being *Komagataella phaffii*, *O. polymorpha*, and *Candida arabi-fermentans* [7]. The phylogeny of methylotrophic yeast is complex and subject to recent updates [8]. The *Ogataea* strains most frequently mentioned in the literature

are DL-1 (ATCC 26012), CBS4732 (ATCC 34438, NRRL-Y-5445, TB-3) and NCYC 495 (ATCC 14754, NRRL Y-1798) [9, 1, 10]. The genomes of all three strains have been sequenced and annotated [1, 7, 10]. A genomic comparison of *O. polymorpha* DL-1 and *K. phaffii* showed that they share three-quarters of enzymes whereas the non-overlapping proteome contains mostly hypothetical, uncharacterized proteins, indicating their close relationship and overlapping metabolic features [10].

The methylotrophic model organism *K. phaffii* is one of the few yeasts that have the ability to use methanol as the only carbon source for energy production. *K. phaffii* was developed into an effective producer of recombinant proteins based on the strength of the native methanol-responsive promoter expression system. The wealth of knowledge has lead to the development of metabolic models on the genome level (GSMM) for simulation and strain engineering [11, 12]. The most recent models are iMT1026v3 and iRY1243 [13, 14]. Based on the close relationship among methylotrophic yeasts and the availability of sequencing data in genomic databases, the development of a GSMM for the metabolism of *O. polymorpha* is desirable.

Here, we present iUL909 as the first GSMM of the biotechnological relevant methylotrophic yeast *O. polymorpha* NCYC 495. The model is based on existing models for *K. phaffii* extended by species-specific substrate utilization identified in phenotype microarrays. Model predictions of growth rates were found to represent experimental growth in different conditions. We tested the performance for overproduction of lactate, and succinate with methanol and glucose as substrates for biotechnological applications and identified potential targets for amplification of reaction activities.

Results and discussion

Here, we report on the construction of a genome-scale metabolic model of the methylotrophic yeast *O. polymorpha* named iUL909. iUL909 was generated on the basis of an existing metabolic model of *K. phaffii* iMT1026v3 [13, 14]. Gene identifiers from *K. phaffii* in iMT1026v3 were replaced by those of homologs from *O. polymorpha*. The model was validated against physiological data from substrate utilization tests with phenotype microarrays and shake flask experiments, as well as growth rates from the literature. We further simulated the production of industrially relevant molecules to identify future biotechnological applications. The validity of the SBML model was tested with Memote [18].

General properties of iUL909

The close relationship among methylotrophic yeasts is reflected by a high overlap of homologs between *O. polymorpha* and *K. phaffii* [7]. The details of our reconstruction and a comparison with the reference GSMMs are shown in Table 1. The reconstruction iUL909 has a lower number of genes associated with reactions because we mapped genes from *K. phaffii* to *O. polymorpha* by means of a homolog-search (see Material and Methods) which failed to find 84 genes (see Additional File 2). With respect to the reference model iMT1026, we added 114 new reactions and 39 new genes (see Additional File 2) from two sources: (i) additionally annotated GPRs from the iRY1243 GSMM, and (ii) genes identified for reactions required

for the metabolization of carbon substrates identified with the Biolog[®] Phenotype microarray. iUL909 has fewer reactions compared to iRY1243 because it does not contain reactions from transcription/translation, signaling and protein turnover, which account for most of the surplus reactions in iRY1243. Overall, the common origin and high overlap of the three GSMMs is visible from the comparable number of the compartmentalized reactions (Table 1).

A direct homolog search between *O. polymorpha* and *K. phaffii* was conducted because of the poor annotation quality of the former. The UniProt TrEMBL [20] sequence source of *O. polymorpha*^[1] listed 5167 proteins, of which 3852 had been predicted by sequence alignments and 1314 had been inferred from additional information about orthologs of closely related species. Only one entry had evidence on protein level. Overall, 422 protein entries were associated with an EC number, of which, however, only 70 proteins had been assigned to a protein name, with the remaining entries classified as 'uncharacterized.' Among the characterized proteins, the majority were connected to processes of information transformation like genetic maintenance and protein phosphorylation. Others were related to structural components like actin-related processes, while still others dealt with the cell cycle, e.g., subunits of the anaphase-promoting complex. Our GSMM of *O. polymorpha* captures 80 EC-annotated protein entries in UniProt, of which 51 are 'uncharacterized' and 29 have identified protein names. In conclusion, the published annotated sequence alone would not have led to a high-quality GSMM.

Comparison of iUL909 with iMT1026

The *O. polymorpha* GSMM reproduces experimental growth rates with growth parameters from iMT1026. The growth parameters include growth- and non-growth-associated maintenance energy (GAM, NGAM) as well as the biomass reaction all transferred from *P. pastoris* for different substrates [13]. The substrates tested were methanol, glucose, and glycerol, industrially relevant sources for which literature values of growth rate and substrate uptake rate are available [21, 5, 22, 23]. As Figure 1 shows, there is good correlation between the experimental and the simulated growth rate, supporting the use of corresponding parameters from iMT1026v3. In particular, the growth characteristics on methanol reported by Van Dijken et al. [21] for *O. polymorpha* (Figure 1, circles) is reproduced, underscoring the similarity of biomass composition and methanol metabolism between *O. polymorpha* and *K. phaffii*.

Notwithstanding the high degree of similarity between *K. phaffii* and *O. polymorpha* there are 84 genes in iMT1026 without a homolog in *O. polymorpha*. The majority of missing genes in *O. polymorpha* affected reactions for which additional enzymes had been annotated with gene-protein-reaction relationships for the associated reaction, but 22 gene products were sole catalysts for the associated reactions (see Additional File 2). Among those orphan reactions, nine were deleted because of

^[1]accessed on 07/31/2018 at <https://www.uniprot.org/uniprot/?query=reviewed:no%20taxonomy:460523>

their marginal integration to the metabolism. However, 13 reactions were retained because the associated enzymes belong to otherwise fully annotated pathways. The identities and the location of all non-proton pumping mitochondrial NADH dehydrogenase system is currently not sufficiently well known. We decided to retain dehydrogenase reactions from the reference model iMT1026. More experiments and data will guide improvements regarding the actual electron and proton homeostasis.

Additional substrates identified with Biolog plates

Substrate tests with phenotype-microarray plates were conducted to compare predictions of the reference GSMM of *K. phaffii* iMT1026 with the actual growth phenotype of *O. polymorpha*. Biolog's Phenotype MicroarrayTM plates test was used to analyze the metabolic utilization of 190 carbon substrates (see Material and Methods). Table 2 shows the overlap and differences between experiment and simulation run with the *K. phaffii* GSMM iMT1026. Correct growth phenotypes were predicted in 77 % cases (13 positives and 134 negatives). In eight cases no growth was experimentally measured, whereas it was predicted by the simulation. Substrates of this class comprised mainly organic acids associated with the TCA cycle (see Additional File 1). We tested growth separately in shake flask experiments and observed growth after two days for all of the eight substrates (Additional File 1). We hypothesize, that the base medium of the Biolog Phenotype MicroarrayTM was inappropriate for supporting growth with the eight substrates and/or that the cultivation time was too short for achieving adaptation of *O. polymorpha* to the specific environmental condition. Comparing our growth phenotypes with literature reports [9], we confirmed 17 out of 19 common substrates. As additional carbon sources, we identified raffinose and maltose (see Additional File 1).

Substrates with positive growth not predicted by simulations were added to the genome-scale reconstruction of *O. polymorpha*. As Table 2 shows, *O. polymorpha* grew on 35 substrates although simulations with iMT1026 predicted no growth. Of these previously unsuspected metabolized substrates, we added 27 substrates to the reconstruction iUL909. The remaining eight substrates were omitted because the annotation of metabolic pathways was unlikely. For example alanine amide has no associated pathway, whereas the degradation pathways for fucose and dulcitol involve galactose, which did not display growth in our test and was shown to be not metabolized by *O. polymorpha* [8]. We omitted gentiobiose because we only observed weak growth, although as a D-glucose disaccharide vigorous growth would be expected. *O. polymorpha* grew on L-leucine, and the model had all pathways required for metabolic activity, but a successful growth simulation was technically not achieved.

Correction of pathways

New metabolic features were added and existing reactions corrected. All genes in the reconstruction iUL909 were checked in the methylotroph gene order browser

(MGOB) [28] and KEGG [29]. We identified enzymes within the biotin pathway converting 8-amino-7-oxononanoate to 7,8-diaminononanoate while using *S*-adenosylmethionine as a co-substrate. To reconstitute the co-substrate we hypothesized that activity of *S*-adenosyl-4-methylsulfanyl-2-oxobutanoate transaminase (EC number 2.6.1.12) would be present. We corrected the equation of over 30 reactions, for example fatty acid synthesis and uridine kinase reactions were corrected to maintain proton balance. There are still 59 unbalanced reactions due to multiple allowed charges in BIGG database and complex interrelations. Overall, the mass and charge balance is correct for more than 96% of all reactions as testified by the Memote report (see GitHub repository) and allows for faithful simulations.

The largest connected path we added was connected with the metabolism of erythritol. iUL909 contains four successive reactions that represent the further processing of erythritol. Erythritol is phosphorylated by a kinase to D-erythritol-1-phosphate, which in turn is converted to L-erythrulose-1-phosphate by means of a dehydrogenase and an epimerase. The L-erythrulose-1-phosphate is then split into dihydroxyacetone phosphate and formaldehyde. These cytosolic reactions were necessary to enable the observed growth on the sugar alcohol erythritol. The hydrolysis of many sugars with growth in the phenotype assays is catalyzed in our reconstruction by the maltase enzyme (EC-number 3.2.1.20). Maltase is known to display broad substrate specificity, which was also shown explicitly for *O. polymorpha* [24]. Two reactions were integrated which describe the activities of transketolase and transaldolase in the peroxisome of *O. polymorpha*. These enzymes are essential for the alcoholic fermentation of xylose [27].

Computational strain engineering tests

Overproduction of important platform chemicals can be achieved with a limited number of genetic manipulations. We chose lactate and succinate as target compounds from methanol and glucose, and applied the FSEOF approach to identify reactions whose increased activity stimulates target production [19]. The optimized synthesis in iUL909 for lactate is routed via methylglyoxal generated from dihydroxyacetone phosphate [25]. Hence, stimulation of the glyoxalase system is predicted to enhance production (Figure 2, green). Succinate production was increased *in silico* when the reactions of the lower glycolysis were more active, e.g., glyceraldehyde-3-phosphate dehydrogenase, pyruvate kinase, and pyruvate carboxylase (Figure 2, blue). The strategies are similar for glucose and methanol. It is interesting to note that low uptake rates of methanol were more sensitive to reactions of the Xyl5P-pathway in the peroxisome. The strategy of reductive TCA cycle stimulation by anaplerotic reactions was experimentally shown to increase succinate production [15].

iUL909 fulfills key standards for reconstructed models and was tested with Memote and the SBML validator [18, 26]. The SBML validator identifies a valid SBML Level 3v1 file, with flux balance constraints (FBC) in version 2. The detected warnings relate to missing initial concentrations. The total Memote score is 45 % (see

Figure 3, and the GitHub repository for the test result file). This is significantly improved with respect to the reference reconstruction for *K. phaffii* with a score of only 24 %.

Conclusions

A genome-scale metabolic model of *O. polymorpha* was constructed using homolog identification based on the existing GSMM of *K. phaffii*. The resulting GSMM iUL909 was further tailored for appropriate substrate utilization on the basis of phenotypic microarray and shake flask experiments. Although biomass composition and energetic parameters were adopted from *K. phaffii*, the resulting growth predictions were in good agreement with chemostat experiments for various industrially relevant substrates. The biotechnological applicability was explored by testing overproduction of lactate, succinate, showing that high productivity can be achieved with limited genetic manipulation. Thus, a metabolic model is now available for further strain engineering.

Methods

Cultivation Experiments

The substrate utilization tests were conducted with the Biolog's Phenotype MicroarrayTM plates P1 and P2 with a total of 190 carbon sources. The medium was prepared with the manufacturer's inoculation medium IFY-0 and the dye mix H according to the guideline. The organism was cultivated in the microarrays in the Growth Profiler (EnzyScreen BV, Heemstede, Netherlands) at 37°C and 150 rpm, growth and respiratory activity was monitored by measuring the optical density (OD) at 490 nm for dye reduction and 750 nm for biomass in a plate reader (Synergy MX, BioTek Instruments Inc., USA) at 0 h and 72 h.

To discriminate growth from non-growth, we considered the distribution of the OD increase for each plate at 490 nm and derived a suitable cut-off. The majority of the OD-increases clustered in a normal distribution at the lower OD-end, followed by a long tail with larger OD increases. The tail of the distribution characterizes explicit substrate respiration, while the normal distribution contains metabolized as well as non-utilized substrates. To identify metabolized substrates within the normal distribution, we first separated the normal distribution from the tail with the clear positive substrates. This was achieved by removing all values above the arithmetic mean within a plate. Then, a normal distribution was fitted over the remaining measurements, and one standard deviation above the mean was used as the cut-off for metabolized substrates (see Additional File 1). Moreover, growth was associated with a substrate only when at least 50 % of replicates surpassed the threshold (duplicates for PM1, triplicates for PM2).

Separate shake flask experiments were performed for the carbon substrates succinate, α -ketoglutarate, α -ketobutyrate, citrate, fumarate, L-malate and for the carbohydrates D-xylose and D-ribose that failed to grow in the phenotype assays. The cultivation took place in CM2-medium [15] and OD was tested after five and eleven days (see Additional File 1).

Computational Genome Comparison

The *O. polymorpha* genome sequence used for the model construction was based on [7] for strain NCYC 495 retrieved from Uniprot [16]. The *K. phaffii* GSMM iMT1026 [13] was used as reference model for the *O. polymorpha* reconstruction. Replacement of gene-protein-reaction (GPR) relationships was conducted by identification of homologs among *K. phaffii* and *O. polymorpha*. We used the genome sequence of *K. phaffii* GS115, the foundation of iMT1026, extracted from UniProt. The homologs were identified using the software package *ProteinOrtho* using the default parameter settings [17]. The model was additionally compared to the *K. phaffii* GSMM iRY1243 [14]. Manual comparison of gene annotations were performed with the JGI linked KEGG database of *O. polymorpha* and for the central carbon metabolism on the MGOB database. The model completeness and quality was tested with Memote [18]. Because Memote returned an error on SBML Level and Version tags, we additionally validated the SBML file with the SBML validator (<http://sbml.org/Facilities/Validator>).

Flux Balance Analysis

Simulations of the GSMM were performed with COBRAv3 on Matlab and with COBRApy. Simulation results differ for COBRAv2, with more and higher exchange reactions and stronger unbalanced reaction rates (≈ -9000 for COBRAv2 compared to ≈ 300 for COBRAv3). A Jupyter Notebook guide for the simulation of iUL909 and reproduction of experimental data as in Figure 1 can be downloaded from GitHub (<https://github.com/iAMB-RWTH-Aachen/Opol-GSMM>). Flux scanning based on enforced objective flux (FSEOF) [19] was performed in Matlab by first maximizing fluxes for the objective of biomass growth. Subsequently, the maximum production rate of the test metabolite was calculated and in the last steps the optimization was performed with the objective on growth while the metabolite production rate was gradually increased close to 90 % of its maximum. All reactions that display a high flux increase are targets to support metabolite overproduction.

Abbreviations

FSEOF: flux scanning based on enforced objective flux

GAM: growth associated maintenance

GPR: gene-protein-reaction relationship

GSMM: genome scale metabolic network

NGAM: non-growth associated maintenance

OD: optical density

SBML: systems biology markup language

TCA: tricarboxylic acid cycle

Author details

¹Institute of Applied Microbiology-iAMB, Aachen Biology and Biotechnology-ABBt, RWTH Aachen University, Worringer Weg 1, 52074 Aachen, Germany. ²Genome Institute of Singapore, 60 Biopolis Street, Genome, 03-01, 138672 Singapore, Singapore. ³Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane City QLD 4072, Australia. ⁴CSIRO Future Science Platform in Synthetic Biology, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Black Mountain, ACT 2601, Australia.

References

1. Ramezani-Rad, M., Hollenberg, C.P., Lauber, J., Wedler, H., Griess, E., Wagner, C., Albermann, K., Hani, J., Piontek, M., Dahlems, U., et al.: The *Hansenula polymorpha* (strain CBS4732) genome sequencing and analysis. *FEMS Yeast Res* **4**(2), 207–215 (2003). doi:[10.1016/S1567-1356\(03\)00125-9](https://doi.org/10.1016/S1567-1356(03)00125-9)
2. Iguchi, H., Yurimoto, H., Sakai, Y.: Interactions of methylotrophs with plants and other heterotrophic bacteria. *Microorganisms* **3**(2), 137–151 (2015). doi:[10.3390/microorganisms3020137](https://doi.org/10.3390/microorganisms3020137)
3. Kunze, G., Kang, H.A., Gellissen, G.: *Hansenula polymorpha* (*Pichia angusta*): Biology and applications. *Yeast Biotechnology: Diversity and Applications*, 47–64 (2009). doi:[10.1007/978-1-4020-8292-4_](https://doi.org/10.1007/978-1-4020-8292-4_)
4. Scheidle, M., Jude, M., Dittrich, B., Denter, S., Kensy, F., Suckow, M., Klee, D., Büchs, J.: High-throughput screening of *Hansenula polymorpha* clones in the batch compared with the controlled-release fed-batch mode on a small scale. *FEMS Yeast Res* **10**(1), 83–92 (2010). doi:[10.1111/j.1567-1364.2009.00586.x](https://doi.org/10.1111/j.1567-1364.2009.00586.x)
5. Lehnen, M., Ebert, B.E., Blank, L.M.: A comprehensive evaluation of constraining amino acid biosynthesis in compartmented models for metabolic flux analysis. *Metab. Eng. Comm.* **5**, 34–44 (2017). doi:[10.1016/j.meten.2017.07.001](https://doi.org/10.1016/j.meten.2017.07.001)
6. Liebal, U.W., Blank, L.M., Ebert, B.E.: CO₂ to succinic acid—estimating the potential of biocatalytic routes. *Metab Eng Comm* **7**, 00075 (2018). doi:[10.1016/j.mec.2018.e00075](https://doi.org/10.1016/j.mec.2018.e00075)
7. Riley, R., Haridas, S., Wolfe, K.H., Lopes, M.R., Hittinger, C.T., Göker, M., Salamov, A.A., Wisecaver, J.H., Long, T.M., Calvey, C.H., et al.: Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A* **113**(35), 9882–9887 (2016). doi:[10.1073/pnas.1603941113](https://doi.org/10.1073/pnas.1603941113)
8. Kurtzman, C.P., Robnett, C.J.: Systematics of methanol assimilating yeasts and neighboring taxa from multigene sequence analysis and the proposal of *Petozozyma gen. nov.*, a new member of the *Saccharomycetales*. *FEMS Yeast Res* **10**(3), 353–361 (2010). doi:[10.1111/j.1567-1364.2010.00625.x](https://doi.org/10.1111/j.1567-1364.2010.00625.x)
9. Suh, S.-O., Zhou, J.J.: Methylotrophic yeasts near *Ogataea* (*Hansenula*) *polymorpha*: a proposal of *Ogataea angusta comb. nov.* and *Candida parapolyomorpha sp. nov.* *FEMS Yeast Res* **10**(5), 631–638 (2010). doi:[10.1111/j.1567-1364.2010.00634.x](https://doi.org/10.1111/j.1567-1364.2010.00634.x)
10. Ravin, N.V., Eldarov, M.A., Kadnikov, V.V., Beletsky, A.V., Schneider, J., Mardanov, E.S., Smekalova, E.M., Zvereva, M.I., Dontsova, O.A., Mardanov, A.V., et al.: Genome sequence and analysis of methylotrophic yeast *Hansenula polymorpha* DL1. *BMC Genomics* **14**(1), 837 (2013). doi:[10.1186/1471-2164-14-837](https://doi.org/10.1186/1471-2164-14-837)
11. Theron, C.W., Berrios, J., Delvigne, F., Fickers, P.: Integrating metabolic modeling and population heterogeneity analysis into optimizing recombinant protein production by *Komagataella* (*Pichia*) *pastoris*. *Appl Microbiol Biotechnol* **102**(1), 63–80 (2018). doi:[10.1007/s00253-017-8612-y](https://doi.org/10.1007/s00253-017-8612-y)
12. Peña, D.A., Gasser, B., Zanghellini, J., Steiger, M.G., Mattanovich, D.: Metabolic engineering of *Pichia pastoris*. *Metab Eng* (2018). doi:[10.1016/j.ymben.2018.04.017](https://doi.org/10.1016/j.ymben.2018.04.017)
13. Tomàs-Gamisans, M., Ferrer, P., Albiol, J.: Fine-tuning the *P. pastoris* iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources. *Microb Biotechnol* **11**(1), 224–237 (2017). doi:[10.1111/1751-7915.12871](https://doi.org/10.1111/1751-7915.12871)
14. Ye, R., Huang, M., Lu, H., Qian, J., Lin, W., Chu, J., Zhuang, Y., Zhang, S.: Comprehensive reconstruction and evaluation of *Pichia pastoris* genome-scale metabolic model that accounts for 1243 ORFs. *Bioresour. Bioprocess.* **4**(1), 22 (2017). doi:[10.1186/s40643-017-0152-x](https://doi.org/10.1186/s40643-017-0152-x)
15. Yan, D., Wang, C., Zhou, J., Liu, Y., Yang, M., Xing, J.: Construction of reductive pathway in *Saccharomyces cerevisiae* for effective succinic acid fermentation at low pH value. *Bioresour Technol* **156**, 232–239 (2014). doi:[10.1016/j.biortech.2014.01.053](https://doi.org/10.1016/j.biortech.2014.01.053)
16. UniProt Consortium: Uniprot: the universal protein knowledgebase. *Nucleic Acids Res* **46**(5), 2699 (2018). doi:[10.1093/nar/gkw1099](https://doi.org/10.1093/nar/gkw1099)
17. Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P.F., Prohaska, S.J.: Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC bioinformatics* **12**(1), 124 (2011). doi:[10.1186/1471-2105-12-124](https://doi.org/10.1186/1471-2105-12-124)
18. Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Babaei, P., Bartell, J.A., Blank, L.M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B.E., Edirisinghe, J.N., Fleming, R.M.T., Garcia-Jimenez, B., van Helvoirt, W., Henry, C., Hermjakob, H., Herrgard, M.J., Kim, H.U., King, Z., Koehorst, J.J., Klamt, S., Klipp, E., Lakshmanan, M., Le Novère, N., Lee, D.-Y., Lee, S.Y., Lee, S., Lewis, N.E., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G.L., Monk, J., Nielsen, J., Nielsen, L.K., Nogales, J., Nookaew, I., Resendis, O., Palsson, B., Papin, J.A., Patil, K.R., Price, N.D., Richelle, A., Rocha, I., Schaap, P., Malik Sheriff, R.S., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaca, P., Vik, J.O., Wodke, J.A., Xavier, J.C., Yuan, Q., Zakhartsev, M., Zhang, C.: MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol* **38**, 272–276(2020). doi:[10.1038/s41587-020-0446-y](https://doi.org/10.1038/s41587-020-0446-y)
19. Choi, H.S., Lee, S.Y., Kim, T.Y., Woo, H.M.: In silico identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol* **76**(10), 3097–3105 (2010). doi:[10.1128/AEM.00115-10](https://doi.org/10.1128/AEM.00115-10)
20. UniProt Consortium: Uniprot: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**(D1), 506–515 (2019). doi:[10.1093/nar/gky1049](https://doi.org/10.1093/nar/gky1049)
21. Van Dijken, L.P., Otto, R., Harder, W.: Growth of *Hansenula polymorpha* in a methanol-limited chemostat. *Arch Microbiol* **111**(1–2), 137–144 (1976). doi:[10.1007/bf00446560](https://doi.org/10.1007/bf00446560)
22. De Koning, W., Harder, W., Dijkhuizen, L.: Glycerol metabolism in the methylotrophic yeast *Hansenula polymorpha*: phosphorylation as the initial step. *Arch Microbiol* **148**(4), 314–320 (1987). doi:[10.1007/BF00456710](https://doi.org/10.1007/BF00456710)
23. Moon, H., Kim, S., Lee, J., Rhee, S., Choi, E., Kang, H., Kim, I., Hong, S.: Independent exponential feeding of glycerol and methanol for fed-batch culture of recombinant *Hansenula polymorpha* DL-1. *Appl Biochem Biotechnol* **111**(2), 65–79 (2003). doi:[10.1385/ABAB:111:2:65](https://doi.org/10.1385/ABAB:111:2:65)
24. Viigand, K., Visnapuu, T., Mardo, K., Aasamets, A., Alamäe, T.: Maltase protein of *Ogataea* (*Hansenula*) *polymorpha* is a counterpart to the resurrected ancestor protein ancMALS of yeast maltases and isomaltases. *Yeast* **33**(8), 415–432 (2016). doi:[10.1002/yea.3157](https://doi.org/10.1002/yea.3157)

25. Inoue, Y., Kimura, A.: Methylglyoxal and regulation of its metabolism in microorganisms. *Adv Microb Physiol* **37**, 177–227 (1995). doi:[10.1016/S0065-2911\(08\)60146-0](https://doi.org/10.1016/S0065-2911(08)60146-0)
26. Bornstein, B.J., Keating, S.M., Jouraku, A., Hucka, M.: LibSBML: an API library for SBML. *Bioinformatics* **24**(6), 880–881 (2008). doi:[10.1093/bioinformatics/btn051](https://doi.org/10.1093/bioinformatics/btn051)
27. Kurylenko, O.O., Ruchala, J., Vasylyshyn, R.V., Stasyk, O.V., Dmytruk, O.V., Dmytruk, K.V., Sibirny, A.A., Peroxisomes and peroxisomal transketolase and transaldolase enzymes are essential for xylose alcoholic fermentation by the methylotrophic thermotolerant yeast, *Ogataea (Hansenula) polymorpha* *Biotechnol Biofuels* **11**, 197 (2018). doi:<https://doi.org/10.1186/s13068-018-1203-z>
28. Douglass, A.P., Byrne, K.P., Wolfe, K.H., The Methylotroph Gene Order Browser (MGOB) reveals conserved synteny and ancestral centromere locations in the yeast family Pichiaceae *FEMS Yeast Research* **19**, foz058 (2019). doi: [10.1093/femsyr/foz058](https://doi.org/10.1093/femsyr/foz058)
29. Kanehisa, M., Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes *Nucleic Acids Res.* **28**(1);, 27–30 (2000). doi: [10.1093/nar/28.1.27](https://doi.org/10.1093/nar/28.1.27)

Acknowledgments

Frank Bergmann supported the annotation of the SBML-file by supplying scripts. The authors are grateful to Ahmed Zahoor-ul-Hassan for stimulating discussions, Felix Küttner for experimental guidance, and Jan Engels for experimental support. Maria Bulimaga, Florencia Casanova, and Lisanne Jente developed and tested initial model versions.

Funding

This work was funded by the Biotechnology Research and Information Network AG (BRAIN AG) and by the German Federal Ministry of Education and Research (BMBF) as part of the Strategic Alliance Zero Carbon Footprint (Grant No. FKZ 031A217F), and by the Excellence Initiative of the German federal and state governments (G:(DE-82)EXS-PF-PFSDS015). The laboratory of L.M.B. is partially funded by the DFG (German Research Foundation) under Germany's Excellence Strategy — Exzellenzcluster 2186, 'The Fuel Science Center' ID: 390919832. The funders had no influence on the design and publishing of the study.

Author's contributions

UWL, LMB and BEE designed the original conception of the study. The biological experiments were performed by CVLS and SS. Simulations were performed by UWL, BAF, CVLS, and AR. Analysis of data was conducted by UWL, CVLS, SS, AR and BEE. The article was written by UWL, BAF, AR, SS, LMB, BEE. All authors have agreed to the final manuscript version.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

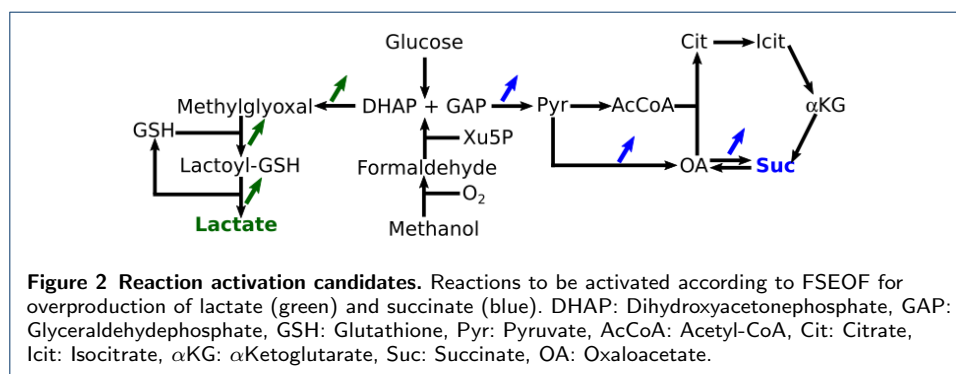
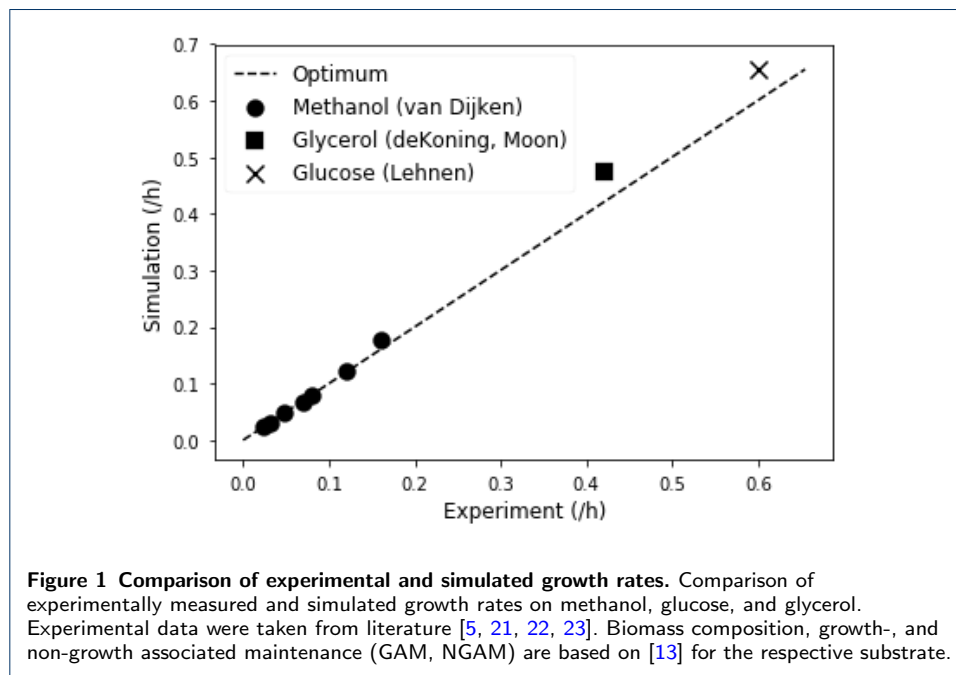
Competing interests

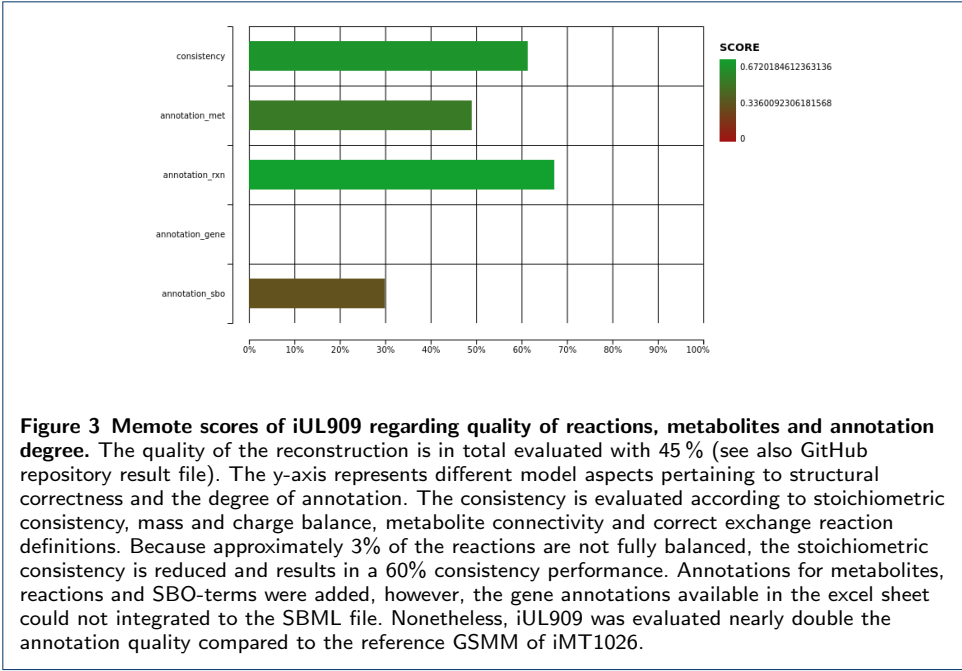
The authors declare that they have no competing interests.

Availability of data and materials

The genome scale model iUL909 generated during the current study is available from the corresponding author on reasonable request. A Jupyter Notebook for simulations, the Memote test result, and the reproduction of Figure 1 is available on GitHub: <https://github.com/iAMB-RWTH-Aachen/Opol-GSMM>

Figures





Tables

Table 1 Feature comparison of the *O. polymorpha* GSMM with respect to the reference GSMM of *K. phaffii*.

	iUL909	iMT1026v3	iRY1243
Genes	909	1026	1243
Metabolites	1639	1706	1740
Reactions	2263	2237	2407
Cytosolic	1077	1067	1281
Mitochondrial	238	268	294
Peroxisomal	112	111	102
Extracellular	17	16	10
Endoplasmatic reticulum	41	41	41
Golgi apparatus	13	13	13
Vacuolar	9	9	9
Nuclear	17	17	17
Transport	541	520	486
Exchange	198	175	154

Table 2 Confusion matrix overview of the growth overlap between *O. polymorpha* in Biolog's Phenotype Microarray™ plates test and prediction according to the iMT1026v3 GSMM for *K. phaffii*.

	iMT1026 growth	iMT1026 no growth
Experiment growth	13	35
Experiment no growth	8	134

Additional Files**Additional File 1 — Experimental data**

Excel sheet with comparison of substrate utilization tests conducted in this paper and Suh et al. [9]. Detailed information on the results of Biolog's Phenotype MicroarrayTM plates test, namely (i) a list of substrates with abiotic activity, (ii) for plate PM1: OD490/600 and selection of metabolized substrates, (iii) for plate PM1: OD490/600 and selection of metabolized substrates, (iv) shake flask growth results for the eight substrates not grown in phenotype assay (succinate, D-xylose, D-ribose, α -ketoglutarate, α -ketobutyrate, citrate, fumarate, malate).

Additional File 2 — Genome scale metabolic model details

Information to the genome scale metabolic model. It contains a list of reactions with only one associated gene-protein-reaction association for which no homolog in *O. polymorpha* was identified. A list of all genes in iMT1026 for which no homologs were identified in *O. polymorpha*. Lists of new reactions and new genes added to the model. A table of literature values for growth rates and substrate uptake rates and finally the results of the FSEOF analysis.