Genome-scale model reconstruction of the methylotrophic yeast Ogataea polymorpha

Ulf W Liebal[1], Brigida A Fabry[1], Aarthi Ravikrishnan[2], Constantin VL Schedel[1], Simone Schmitz[1], 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
[1]Institute 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 arabinofermentans* [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.

O. polymorpha and *K. phaffii* share many similarities especially regarding their methanol metabolism. However, several metabolic features are distinctive. *O. polymorpha* is capable of growth at 50°C and is therefore one of the most thermotolerant eukaryotic microorganisms [15]. In contrast to *K. phaffii*, *O. polymorpha* can ferment xylose to ethanol [16] and can assimilate nitrate [17]. Moreover, high activity of the AOX promoter with glycerol as the sole carbon source or under glucose starvation is a unique feature of *O. polymorpha*. In *K. phaffii* activity of AOX strictly depends on the presence of methanol [18].

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 [19].

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 Figure 1 and Table 1. We mapped genes from *K. phaffii* to *O. polymorpha* using a homolog-search (see Material and Methods) which failed to find 84 genes (see Additional File 2). However, 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. iRY1243 differs from the reference model iMT1026v3 only in reactions associated with transcription/translation, signaling and protein turnover. Since iUL909 covers only metabolic reactions it was not amended with these additional reactions from iRY1243. Overall, the common origin and high overlap of the three GSMMs is visible from the comparable number of the compartmentalized reactions (Figure 1).

Comparison of iUL909 with iMT1026

The *O. polymorpha* GSMM reproduces experimental growth rates with growth parameters from iMT1026. All growth parameters were taken from the reference model iMT1026v3, namely growth- and non-growth-associated maintenance energy (GAM, NGAM) as well as the biomass reaction [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 [20, 5, 21, 22]. The Pearson correlation coefficient of 0.9996 between predicted and experimental growth rate supports using GAM, NGAM and biomass equation from iMT1026v3 (Figure 2). In particular, the growth characteristics on methanol reported by Van Dijken et al. [20] for *O. polymorpha* (Figure 2, 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 were 84 genes in iMT1026 without a detected homolog in *O. polymorpha*. Because many reactions were catalyzed by alternative enzymes, the gene-protein-reaction relationships (GPR) was still satisfied for 62 genes, but 22 gene products were sole catalysts for the associated reactions (see Additional File 2). These 22 reactions lacked a GPR, caused by gene function replacement or loss. Replacement is likely when upstream and downstream reactions are annotated, loss (or misannotation in iMT1026) is likely for orphan reactions without connections to the metabolism. Among those orphan reactions, nine were deleted because they could not carry flux. For example, a potential misannotation was identified for 4a-hydroxytetrahydrobiopterin dehydratase (EC number 4.2.1.96) in iMT1026 for which neither KEGG nor the model itself would allow metabolite flux. However, thirteen reactions were retained because they were metabolically well connected, and their deletion would result in reaction gaps. For example, 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.

Substrate identification 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 Microarray™ 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 Microarray™ 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. Indeed, we verified growth support in shake flask experiments (Additional File 1). There are possibly additional false-negative results among the 134 substrates identified by the Biolog Phenotype Microarrays as non-growth supporting, if, simultaneously, the reference model iMT1026v3 would falsely lack the corresponding metabolic activities. 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.

Gene annotation and pathway correction

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

(MGOB) [23] and KEGG [24]. We accessed the UniProt TrEMBL [25] sequence
source of *O. polymorpha*[1] which listed 5167 proteins with only 422 EC number
associated enzymes. Due to the incomplete enzyme coverage by the UniProt an-
notation we relied on KEGG and the MGOB, and could functionally annotate
$> 97\%$ of the genes. We identified enzymes within the biotin pathway converting
8-amino-7oxononanoate to 7,8-diaminononoate 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 bal-
ance. 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 nec-
essary 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 reconstruc-
tion 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* [26].
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 com-
pounds from methanol and glucose, and applied the FSEOF approach to identify
reactions whose increased activity stimulates target production [28]. The optimized
synthesis in iUL909 for lactate is routed via methylglyoxal generated from dihydrox-
yacetone phosphate [29]. Hence, stimulation of the glyoxalase system is predicted to
enhance production (Figure 3, 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 3,
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

---

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

[30].

iUL909 fulfills key standards for reconstructed models and was tested with Memote and the SBML validator [19, 31]. 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 4, 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 Microarray™ 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 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 [30] and OD was tested after two, seven and thirteen 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 [25]. 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 for homolog detection [32]. ProteinOrtho reports a single best homolog candidate for a *K. phaffii* input protein sequence. *K. phaffii* GPRs were replaced by the *O. polymorpha* homolog, while maintaining isoenzyme and multi-protein complex GPRs, which was possible for 908 GPRs. In the case of failed homolog mappings, manual Blast searches were conducted, and for the reactions of the central carbon metabolism manual comparison of gene annotations were performed with the JGI linked KEGG database of *O. polymorpha* [33] and the MGOB database [23].

The MGOB shows the local genomic organization among methylotrophic yeasts. *K. phaffii* MGOB-IDs of reactions in the central carbon metabolism were derived by querying the MGOB database with KEGG gene IDs. The MGOB-IDs were then used to locate the genes in their genomic context. The alignment with the homologs of all methylotrophic yeasts within MGOB provided information on annotation including subcellular localization. In this way, we could characterize 518 additional genes compared to UniProt. The model completeness and quality was tested with Memote [19]. SBML Level and Version tags were additionally validated 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. A Jupyter Notebook guide for the simulation of iUL909 and reproduction of experimental data as in Figure 2 can be downloaded from GitHub (https://github.com/iAMB-RWTH-Aachen/Opol-GSMM). Flux scanning based on enforced objective flux (FSEOF) [28] was performed in Matlab using the functions of COBRA toolbox to identify the targets for overexpression. Briefly, FSEOF was carried out in two stages. In the first stage, the maximum product flux (lactate or succinate) was computed by setting the biomass flux to zero and maximizing the objective function for product formation. In the second stage, the product flux was enforced in steps, and the biomass growth was set as an objective function and maximized. Reactions fluxes increasing monotonically represent targets for overexpression.

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
[1]Institute of Applied Microbiology-iAMB, Aachen Biology and Biotechnology-ABBt, RWTH Aachen University, Worringer Weg 1, 52074 Aachen, Germany. [2]Genome Institute of Singapore, 60 Biopolis Street, Genome, 03-01, 138672 Singapore, Singapore. [3]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane City QLD 4072, Australia. [4]CSIRO 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., Others: 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

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

3. Kunze, G., Kang, H.A., Gellissen, G.: *Hansenula polymorpha (Pichia angusta)*: Biology and Applications. Yeast Biotechnol Divers Appl, 47–64 (2009). doi:10.1007/978-1-4020-8292-4-3

4. Scheidle, M., Jeude, 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

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 Commun **5**, 34–44 (2017). doi:10.1016/j.meteno.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 Commun **7** (2018). doi: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., al., E.: Comparative genomics of biotechnologically important yeasts. Proc Natl Acad Sci USA **113**(35), 9882–9887 (2016). doi: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 *Peterozyma 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

9. Suh, S.-O., Zhou, J.J.: Methylotrophic yeasts near *Ogataea (Hansenula) polymorpha*: a proposal of *Ogataea angusta comb. nov.* and *Candida parapolymorpha sp. nov.*. FEMS Yeast Res **10**(5), 631–638 (2010). doi:10.1111/j.1567-1364.2010.00634.x

10. Ravin, N.V., Eldarov, M.A., Kadnikov, V.V., Beletsky, A.V., Schneider, J., Mardanova, E.S., Smekalova, E.M., Zvereva, M.I., Dontsova, O.A., Mardanov, A.V., al., E.: Genome sequence and analysis of methylotrophic yeast *Hansenula polymorpha* DL1. BMC Genomics **14**(1), 837 (2013). doi: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

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

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 (2018). doi: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

15. Cabeç-Silva, C., Madeira-Lopes, A.: Temperature relations of yield, growth and thermal death in the yeast *Hansenula polymorpha*. Z Allg Mikrobiol **24**(2), 129–132 (1984). doi:10.1002/jobm.19840240216

16. Ryabova, O., Chmil, O., Sibirny, A.: Xylose and cellobiose fermentation to ethanol by the thermotolerant methylotrophic yeast. FEMS Yeast Res **4**(2), 157–164 (2003). doi:10.1016/S1567-1356(03)00146-6

17. Siverio, J.M.: Assimilation of nitrate by yeasts. FEMS Microbiol Rev **26**(3), 277–284 (2002). doi:10.1111/j.1574-6976.2002.tb00615.x

18. Gellissen, G., Kunze, G., Gaillardin, C., Cregg, J., Berardi, E., Veenhuis, M., VanderKlei, I.: New yeast expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adeninivorans* and *Yarrowia lipolytica* – A comparison. FEMS Yeast Res **5**(11), 1079–1096 (2005). doi:10.1016/j.femsyr.2005.06.004

19. Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., Bartell, J.A., Blank, L.M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B.E., Edirisinghe, J.N., Faria, J.P., Feist, A.M., Fengos, G., Fleming, R.M.T., García-Jiménez, B., Hatzimanikatis, V., van Helvoirt, W., Henry, C.S., Hermjakob, H., Herrgård, M.J., Kaafarani, A., Kim, H.U., King, Z., Klamt, S., Klipp, E., Koehorst, J.J., König, M., Lakshmanan, M., Lee, D.-Y., Lee, S.Y., Lee, S., Lewis, N.E., Liu, F., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G.L., Monk, J.M., Nielsen, J., Nielsen, L.K., Nogales, J., Nookaew, I., Palsson, B.O., Papin, J.A., Patil, K.R., Poolman, M., Price, N.D., Resendis-Antonio, O., Richelle, A., Rocha, I., Sánchez, B.J., Schaap, P.J., Malik Sheriff, R.S., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J.O., Wodke, J.A.H., Xavier, J.C., Yuan, Q., Zakhartsev, M., Zhang, C.: MEMOTE for standardized genome-scale metabolic model testing. Nat Biotechnol **38**(3), 272–276 (2020). doi:10.1038/s41587-020-0446-y

20. 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

21. 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

22. Moon, H., Kim, S.W., Lee, J., Rhee, S.K., Choi, E.S., Kang, H.A., Kim, I.H., Hong, S.I.: 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

23. 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 Res **19**(6), 058 (2019). doi:10.1093/femsyr/foz058

24. 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. http://www.genome.ad.jp/kegg/

25. UniProt Consortium: UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res **47**(D1), 506–515 (2019). doi:10.1093/nar/gky1049

26. 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

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**(1), 197 (2018). doi:10.1186/s13068-018-1203-z

28. 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 Env Microbiol **76**(10), 3097–3105 (2010). doi:10.1128/AEM.00115-10

29. 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

30. 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

31. 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

32. 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

33. Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I.: MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res **42**(D1), 699–704 (2014). doi:10.1093/nar/gkt1183

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
O. polymorpha: NCYC 495 leu1.1
Genome Accession: AECK01000000
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
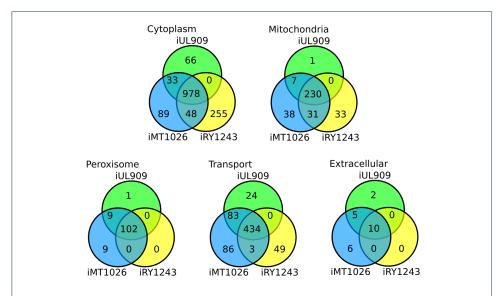
Figures

Figure 1 Venn diagrams of compartment specific reactions overlap. The genome-scale metabolic model (GSMM) of *K. phaffii* iMT1026 [13] was used as a reference to construct the GSMM for *O. polymorpha* because of metabolic and phylogenetic similarities (see text). iMT1026 was further adapted for biotechnology in the GSMM iRY1243 [14]. The reactions are fully overlapping for the compartments Golgi, nucleus, endoplasmatic reticulum and vacuole.
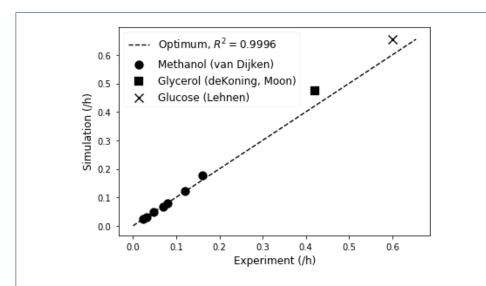

Figure 2 Comparison of experimental and simulated growth rates. Comparison of experimentally measured and simulated growth rates on methanol, glucose, and glycerol. Experimental data were taken from literature [5, 20, 21, 22]. Biomass composition, growth-, and non-growth associated maintenance (GAM, NGAM) are based on [13] for the respective substrate.
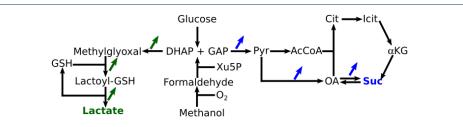
Figure 3 Reaction activation candidates. Reactions to be activated according to FSEOF for overproduction of lactate (green) and succinate (blue). DHAP: Dihydroxyacetonephosphate, GAP: Glyceraldehydephosphate, GSH: Glutathione, Pyr: Pyruvate, AcCoA: Acetyl-CoA, Cit: Citrate, Icit: Isocitrate, αKG: αKetoglutarate, Suc: Succinate, OA: Oxaloacetate.
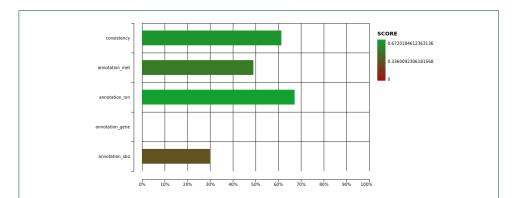

Figure 4 Memote scores of iUL909 regarding quality of reactions, metabolites and annotation degree. The quality of the reconstruction is in total evaluated with 45 % (see also GitHub repository result file). The y-axis represents different model aspects pertaining to structural correctness and the degree of annotation. The consistency is evaluated according to stoichiometric consistency, mass and charge balance, metabolite connectivity and correct exchange reaction definitions. Because approximately 3% of the reactions are not fully balanced, the stoichiometric consistency is reduced and results in a 60% consistency performance. Annotations for metabolites, reactions and SBO-terms were added, however, the gene annotations available in the excel sheet could not integrated to the SBML file. Nonetheless, iUL909 was evaluated nearly double the annotation quality compared to the reference GSMM of iMT1026.

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

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 Microarray$^{\text{TM}}$ 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.