

# *Ustilago maydis* Metabolic Characterization and Growth Quantification with a Genome-Scale Metabolic Model

Ulf W. Liebal<sup>1\*</sup>, Lena Ullmann<sup>1</sup>, Christian Lieven<sup>2,3</sup>, Philipp Kohl<sup>1</sup>, Daniel Wibberg<sup>4,5</sup>, Thiemo Zambanini<sup>1</sup>, Lars M. Blank<sup>1\*</sup>

<sup>1</sup> iAMB - Institute of Applied Microbiology, ABBt, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany; ulf.liebal@iamb.rwth-aachen.de (U.W.L.), lena.ullmann@rwth-aachen.de (L.U.), Philipp.Kohl@rwth-aachen.de (P.K.), tiemo.zambanini@rwth-aachen.de (T.Z.), lars.blank@rwth-aachen.de (L.M.B)

<sup>2</sup> Unseen Biometrics ApS, Copenhagen, Denmark

<sup>3</sup> Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

<sup>4</sup> Genome Research of Industrial Microorganisms, CeBiTec, Bielefeld University, D-33501 Bielefeld, Germany; dwibberg@cebitec.uni-bielefeld.de (D.W.)

<sup>5</sup> Institute of Bio- and Geosciences IBG-5, Computational Metagenomics, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany.

\* Correspondence: ulf.liebal@iamb.rwth-aachen.de (U.W.L.); lars.blank@rwth-aachen.de (L.M.B)

**Abstract:** *Ustilago maydis* is an important plant pathogen causing corn-smut disease and an effective biotechnological production host. The lack of a comprehensive metabolic overview hinders a full understanding of the organism's environmental adaptation and a full use of its metabolic potential. Here, we report the first genome scale metabolic model (GSMM) of *Ustilago maydis* (iUma22) for the simulation of metabolic activities. iUma22 was reconstructed from sequencing and annotation using PathwayTools, the biomass equation was derived from literature values and from the codon composition. The final model contains over 25% of annotated genes (6,909) in the sequenced genome. Substrate utilization was corrected by Biolog-Phenotype arrays and exponential batch cultivations were used to test growth predictions. The growth data revealed a metabolic phenotype shift at high glucose uptake rates and the model allowed its quantification. A pan-genome of four different *U. maydis* strains revealed missing metabolic pathways in iUma22. The new model allows studies of metabolic adaptations to different environmental niches as well as for biotechnological applications.



**Copyright:** © 2022 by the authors. Terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** *Ustilago maydis*; Genome Scale Metabolic Model; Constraint Based Model; Biotechnology; COBRA; FBA; Metabolism; Itaconate

## 1. Introduction

*Ustilago maydis* is a model organism and economically important fungus from the division of *Basidiomycota*. The associated corn smut disease affects maize harvest but is also used as food itself [1]. As a parasite, *U. maydis* grows into the plant tissue to extract substrates for its own metabolic activity. *Ustilaginaceae* show a versatile product spectrum such as organic acids (e.g., itaconate, malate, succinate), polyols (e.g., erythritol, mannitol), and extracellular glycolipids, which are considered value-added chemicals with potential applications in the pharmaceutical, food, and chemical industries. *U. maydis* has developed an effective native production of itaconic acid, an important platform chemical. Indeed, the itaconic acid production in *U. maydis* was improved to surpass the current biotechnological route of *Aspergillus terreus* based production. The advantages are yeast-like growth, high productivities, yields and titer and reduced byproduct formation [2–4], and since it is a model organism, efficient genetic tools are available [5].

The annotated genome sequence of *Ustilago maydis* strain 512 enabled a deeper understanding of the pathogenic mechanisms as well as the metabolic competencies [6]. Annotated genomes can be used to construct genome-scale metabolic models (GSMM), which serve as a knowledge-base of metabolic capacities and allow rational biotechnological engineering [7]. GSMM can be optimized to identify genetic modifications for metabolic engineering that maximize the production of metabolic intermediates. In addition, optimal biotechnological production routes regarding different organisms and metabolic pathways can be computationally evaluated using their respective GSMM [8]. The performance of metabolic microbial and cross-kingdom interactions can be interrogated, to identify exchange metabolites, community stability and metabolic properties that mark the transitions from mutualism to parasitism [9–11].

Here, we present the first high-quality genome-scale metabolic model for *U. maydis* called iUma22. The initial draft model was constructed by the automated PathwayTools workflow with substantial manual adjustments including pathway adjustments derived from Biolog phenotype arrays with 190 carbon substrates. The model quality was assessed using the community evaluation with Memote and the accuracy of rate predictions was tested with glucose growth experiments. The metabolic capacity of the model iUma22 was compared to a pan-genome of different *U. maydis* strains annotated by means of KAAS and enabling reconstruction of KEGG pathways and BRITE hierarchies.

## 2. Materials and Methods

### 2.1. Draft GSMM from Pathway Tools

The genomic DNA sequence of *Ustilago maydis* (Strain 521 FGSC 9021) [6] was obtained from NCBI's RefSeq project [12]. A corresponding annotation file was then exported from the MIPS *Ustilago Maydis* Database via the PEDANT Interface [13]. Using the PathoLogic Tool [14], the sequence and annotation files were parsed and in combination with MetaCyc reactions database a new Pathway/Genome Database (PGDB) was created. During pathway cleaning reactions from other taxa are pruned, unless there are enzymes matching to all of the reactions. Additional metabolic activity was identified using the 'Pathway Hole Filler' function and sequence information of isoenzymes was used to query the proteome of *U. maydis* via pBlast. Protein sequences were queried on PEDANT, MUMDB, MetaCyc or KEGG [15,16] and manually curated while inconclusive polypeptides as well as those that are involved in signaling and other non-metabolic pathways were discarded.

### 2.2 Strains sequenced, pan-genome, KEGG pathway enrichment

To identify metabolic differences within the *U. maydis* strain family, a pan-genome consisting of five *Ustilago maydis* strains was assembled, including strains 198, 482, 485, 512 [17]. The Nanopore Rapid DNA Sequencing kit (SQK-RAD04, Oxford Nanopore Technologies, Oxford, UK) was used for preparation and sequencing was performed on an Oxford Nanopore GridION Mk1 sequencer using a R9.4.1 flow cell. The Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA, USA) was used for whole-genome-shotgun PCR-free libraries from 5 µg of gDNA. The library quality was assessed by an Agilent 2000 Bioanalyzer with Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA) for fragment sizes of 500–1000 bp. Paired end sequencing was performed on the Illumina MiSeq platform (2 × 300 bp, v3 chemistry). Adapters and low-quality reads were removed by an in-house software pipeline prior to polishing as recently described [18]. Run control was based on MinKNOW (Oxford Nanopore Technologies) with the 48 h sequencing run protocol. Base calling was performed offline using Bonito, assembly with canu v2.1.1 [19], contigs were polished with Pilon [20] for ten iterative cycles, and for read mapping BWA-MEM [21] and Bowtie2 v2.3.2 [22] in the first and second five iterations, respectively.

Genes were predicted using GeneMark-ES 4.6.2. [23], and functionally annotated using a modified version of the genome annotation platform GenDB 2.0 [24] for eukaryotic genomes [25]. Similarity searches were conducted against COG [26], KEGG [16] and SWISS-PROT [27]. Identification of putative tRNA genes was conducted with tRNAscan-SE [28]. Completeness, contamination, and strain heterogeneity were estimated with BUSCO (v3.0.2 [29]), using the fungi-specific single-copy marker genes database (odb9). The obtained genome sequences are compared and documented in more detail in Ullmann et al. (2022) [17]. The pan-genome of all available *U. maydis* strains was calculated by means of EDGAR 3.0 [30]. The KEGG pathway annotation was performed by comparison of E.C. numbers in the pan-genome annotation and E.C. numbers in the reaction description of iUma22. The comparison resulted in three lists, E.C. numbers only present in iUma22 (iUmaNOTpan), present in the pan-genome and iUma (iUmaANDpan), and only present in pan-genome (panNOTiUma). The panNOTiUma list was exported as fasta-file and KAAS [31] was used to annotate the list with KEGG pathway information. The annotation of the genes in the SBML-file was achieved with the BioServices Python package [32].

### 2.3 Biomass equation and growth/non-growth maintenance

The composition of proteins, RNA and DNA was estimated based on the protein and genome sequence respectively, whereas the composition of lipids and the cell wall were results mined from scientific articles [33,34]. The exact biomass composition of *U. maydis* is not available, however, the specific elemental composition [35] and the biomass composition for fungi in general [36] was used as a starting point and linear programming was applied to approximate the total biomass composition (Supplementary 2). The composition values of each monomer were converted into stoichiometric values [37]. For example, to determine the AA composition contribution (in mol<sub>AA</sub>/g<sub>Prot</sub>), first the AA-protein molarity (MP<sub>AA</sub> in g<sub>AA</sub>/mol<sub>Prot</sub>) was calculated by multiplying the AA codon frequency with the AA molar mass (minus the molar mass of water released during polymerization) (Eq. 1). Normalizing each AA-protein molarity by the overall sum yields the weight fraction of each AA (WP<sub>AA</sub>) (Eq. 2). Division of the AA weight fraction (WP<sub>AA</sub>) by its molar mass and multiplication with the weight fraction of protein to the dry weight (X in g<sub>Prot</sub>/g<sub>CDW</sub>) and conversion from mol to mmol (factor 1000) gives the stoichiometric factor (SF<sub>AA</sub>) (Eq. 3). To calculate the stoichiometric factor of an AA (SF<sub>AA</sub>) the molar percentage (MP) is multiplied with the fractional protein mass per biomass (X) (Eq. 2).

$$MP_{AA} [g_{AA}/mol_{Prot}] = CDN_{AA} / \sum CDN \cdot (M_{AA} - M_{H_2O}) \quad (1)$$

$$WP_{AA} [g_{AA}/g_{Prot}] = MP_{AA} / \sum MP \quad (2)$$

$$SF_{AA} [mol_{AA}/g_{CDW}] = WP_{AA} / (M_{AA} - M_{H_2O}) \cdot X \cdot 1000 \quad (3)$$

The fraction of protein on total biomass (X) is unknown and was determined by linear optimization. The average elemental composition of each macromolecule (protein, DNA, RNA, lipid, cell wall) was determined by summing up the products of the absolute amount of each element. For each macromolecule the C-mole content was calculated by division with the carbon elemental composition [35]. Phosphorous and Sulfur were added from the elemental composition of *S. cerevisiae*. The optimization followed the formula:

$$A \cdot X = b \quad (4)$$

$$\text{subject to: } X_{lbi} \leq X_i \leq X_{ubi} \quad (5)$$

The rows of matrix  $A$  correspond to the elements  $C, H, O, N, P, S$  whereas the columns correspond to the macromolecule types (protein, DNA, RNA, lipids, cell wall). The vector  $b$  represents the measured elemental biomass  $C, H, O, N$  supplemented by the elemental content of  $P$  and  $S$  of *S. cerevisiae*. The equation was solved for vector  $X$ , the biomass fractions of each of the macromolecules (Table 1 and Supplementary 2). The ATP-associated growth-maintenance (GAM) with ATP-dependent glucose uptake was calculated as 31 mmol/g<sub>CDW</sub> by optimization of sum of squared errors of growth experiments (Table 3). The non-growth associated maintenance (NGAM) was calculated as 0.75 mmol/g<sub>CDW</sub>/h by maximization of NGAM with the lowest not infeasible glucose uptake rate of 0.22 mmol/g<sub>CDW</sub>/h close to the experimentally determined glucose maintenance requirement of 0.2 mmol/g<sub>CDW</sub>/h (Figure 3A).

**Table 1.** Macromolecular composition of *U. maydis* calculated by linear optimization. The full composition is provided as Supplementary 1.

Component	Protein	DNA	RNA	Lipids	Cell Wall
g/100 g <sub>CDW</sub>	30	0.3	10	40	16

#### 2.4 Substrate and growth experiments

For the substrate utilization experiments the Biolog Phenotype Microplates™ PM1 and PM2A were used with *Ustilago maydis* strain FB1. Cultures were first grown on YEPS-agar plates at 30°C for at least 24 hours. To prepare the pre-cultures 25 mL of YEPS medium were inoculated from the plates of each strain then performed in 100 mL Erlenmeyer flasks and incubated at 30°C, 200 rpm for 24 hours. (Ecotron Incubation shaker, Infors HT AG, Switzerland). The inoculation fluid was prepared with IFY-0 (1.2x), cell suspension and sterile water to obtain a starting turbidity of 62% T, with 100 µL for each well. The inoculated plates were shaken at 200 rpm with a shaking diameter of 50 mm, at 30°C and with a humidity of 70% up to 168 hours (Multitron Incubation shaker, Infors HT AG, Switzerland). Microbial growth was measured with the SynergyMX (BioTek Instruments, USA) with the optical density at 600 nm. The Biolog raw data is available in Supplementary 4 (PM1) and 5 (PM2A).

The threshold for positive growth was determined by examining the OD histograms for each plate. A normal distribution at low OD values represents the OD range below positive growth (Figure 2). The final growth threshold of 0.4 a.u. was empirically determined to maximize logic consistency and to minimize the integration of false positive metabolic activity. The value approximates the end of a normal distribution of non-growth at low ODs. Separate glucose shake-flask batch experiments in modified Tabuchi medium [38] were conducted with strains MB215, FB1 Mating type a1b1 as well as literature data was used to estimate growth rates and glucose uptake rates. The OD measurements of the growth data was converted into g<sub>CDW</sub>/L using the empirical relation from yeast of 0.62 g<sub>CDW</sub>/L /OD (BNID 111182, [39]). The growth rates were identified using a nonlinear fit of the biomass to the Verhulst equation,

$$X(t) = X_0 * C / (X_0 + (C - X_0) * \exp(-\mu * t)), \quad (1)$$

which calculates the biomass from the initial biomass ( $X_0$ ), the max biomass capacity ( $C$ ), the growth rate ( $\mu$ ) and time ( $t$ ). The substrate uptake rate was estimated with a linear equation [40].

### 3. Results and discussion

#### 3.1. Description of iUma22

The genome scale model for *Ustilago maydis* was constructed based on the genome sequence and annotation of strain 521 [6]. Table 2 shows the number of represented genes, metabolites and reactions in the new reconstruction and a comparison to the community yeast model ([41], version 8.5.0). Whereas the community yeast model is more comprehensive, iUma22 has a higher gene-to-reaction ratio as well as gene-protein-reaction relationships (GPR), as we aimed to include well connected metabolic pathways with (predicted) annotated genes. *U. maydis* and *S. cerevisiae* have similar number of predicted genes and, when assuming the yeast 8.5.0 as a benchmark of metabolic representation, the iUma22 has reached 70% of genes completeness. There are likely gaps in the secondary metabolism discussed in Ullmann et al. (2022) [17] as well as adaptations to the pathogenic life style.

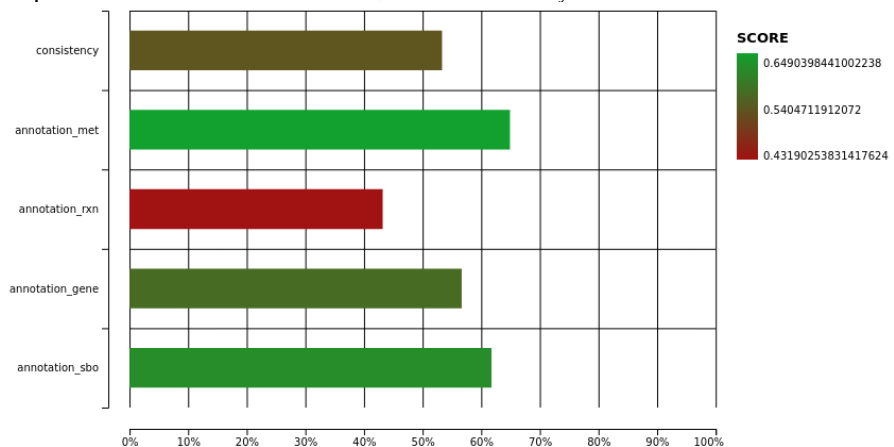
**Table 2.** Number of metabolites, reactions and genes of the genome scale metabolic model of *U. maydis* iUma22 and in comparison, the community yeast model (8.5.0 [41]).

Component	iUma22	Yeast 7.6 <sup>1</sup>
Genes	814	1,150
Metabolites	1,233	2,742
Reactions	1,856	4,058
Reactions with GPR	1434	2633
Pred. genes <sup>2</sup>	6.909	6,464

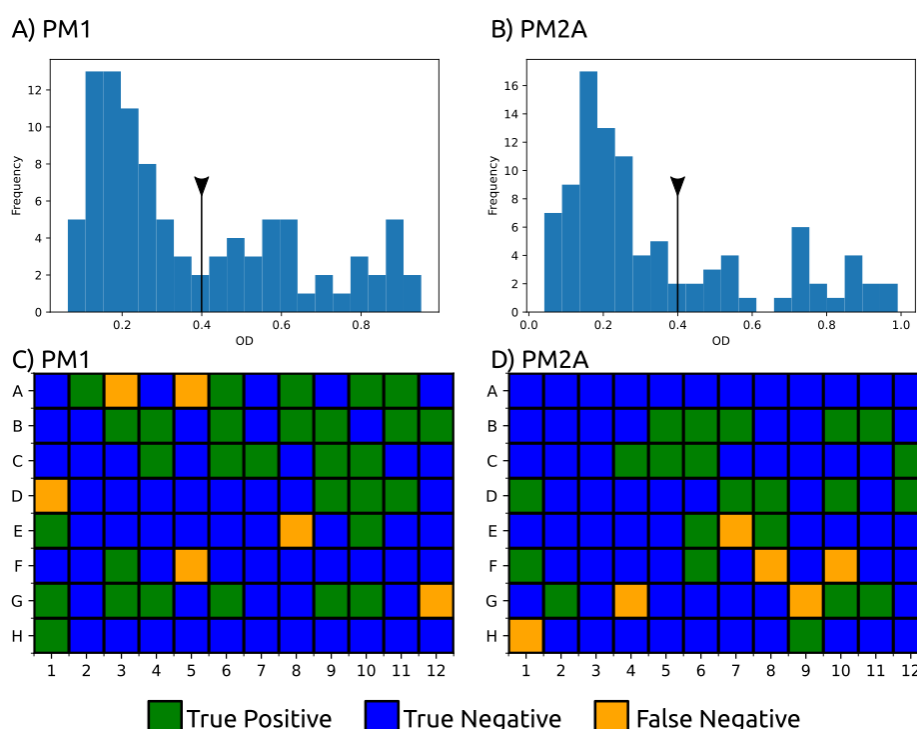
<sup>1</sup> <https://github.com/SysBioChalmers/yeast-GEM>, update 2021-06-24.

<sup>2</sup> <https://www.ncbi.nlm.nih.gov/datasets>, accessed 2022-02-04

The quality of iUma22 was tested with Memote with an overall performance of 57% [42] (Figure 1). Mass and charge balance as well as metabolite connectivity show high quality with scores of over 98%. Memote detects unbounded fluxes that reach boundary conditions during flux variability analysis for 203 reactions on standard media. The stoichiometric consistency of the model could not be evaluated thus decreasing the overall consistency quality to 53%. Note, however, that also for the *S. cerevisiae* community model ([41], version 8.5.0) the stoichiometric consistency test fails. Annotations for metabolites, reactions and genes contain detailed unique annotations. The community yeast model, developed since more than a decade, is evaluated by Memote with a total score of 65%.



**Figure 1.** Memote quality report of iUma22 with total score of 57%. The full html report is provided as Supplementary 3.



**Figure 2.** BIOLOG phenotype experiments with carbon sources from PM1 and PM2A. Growth was evaluated by OD 600 after 144h for PM1 (A) and 288h for PM2A (B) with a threshold of 0.4 a.u (black line with triangle) which excludes the normal distribution at low ODs representing no growth. 52 substrates were correctly predicted to growth (true positive, green) and 128 correctly assigned to non-growth by iUma22 (true negative, yellow) in the plates PM1 (C) and PM2A (D). Twelve substrates could not be balanced to enable growth in iUma22 (false negative). Results of PM1 and PM2A are provided as Supplementary 4 and 5.

### 3.2. Carbon Substrate tests with BIOLOG Phenotype arrays

The model iUma22 correctly reproduces 96% of growth phenotypes tested in BIOLOG carbon source assays. We chose an OD threshold of 0.4 a.u. for growth in the plates PM1 and PM2A, a value right after the apparent normal distribution of non-growth for low final OD values (Figure 2). The threshold is a compromise to include growth for glycine dipeptides (PM1: E1, G1, G6, H1), but also includes TCA cycle intermediates (succinate (PM1:A5), fumarate (PM1:F5), aspartate (PM1:A7), malate (PM1:G12)). These TCA cycle intermediates could not be enabled for growth in the model. The largest set of reactions added because of the BIOLOG plates includes di- and oligosaccharide metabolism and methylated central carbon metabolites. Overall, growth took place in 52 wells (34 in PM1 and 19 in PM2A). iUma22 was manually adjusted to reproduce the majority of the growth phenotypes (Figure 2).

While the majority of substrate were correctly reproduced, some metabolites failed to support growth in iUma22. Many intermediates from the TCA cycle did not support growth in the Biolog results (2-oxoglutarate, fumarate, succinate, aspartate) while others like lactate (PM1:B9), malate (PM1:G12) and succinamate (PM2A:F10) were positive (Figure 2). The degradation of arginine (PM2A:G4), isoleucine (PM2A:G9) and ornithine (H1) is dysfunctional in iUma22, despite the ability of the model to grow on proline, which is a carbon intermediate of the common carbon metabolism following the ornithine-glutamate aminotransferase reaction (E.C. 2.6.1.13, ORNTArm). The metabolism of sebacic acid (PM2A:F8) is not represented in the model, because no information is available

in the databases KEGG and Metacyc. Finally, note that false positive results were corrected by removing exchange reactions of the associated metabolites.

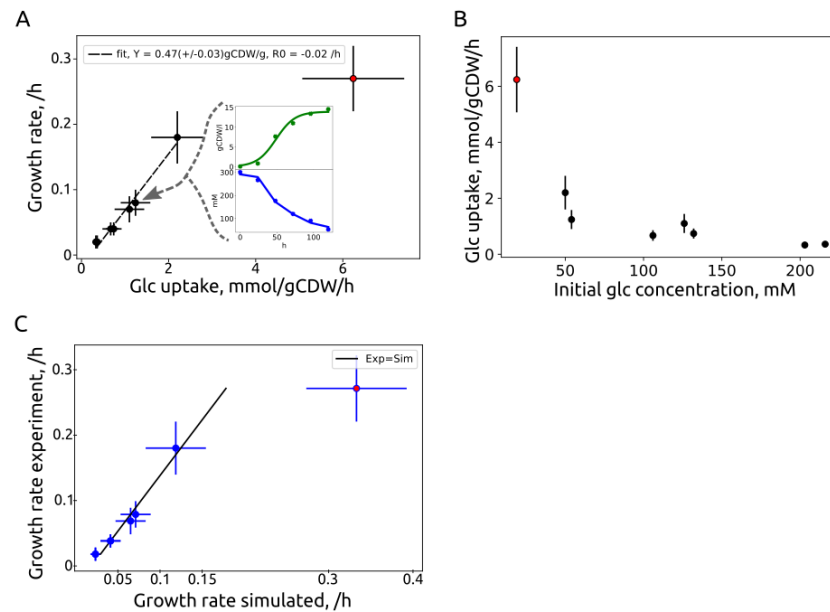
### 3.3. Growth rate correlation

We evaluated glucose growth characteristics and compared the results with iUma22 predictions. We used published batch data of *U. maydis* MB215 control strains [4], that were genetically modified for optimized biotechnological performance, as well as newly generated data with *U. maydis* strain 521 (Table 3). Note, that the model considers ATP-consuming glucose uptake via Hxt1 [43]. The experiments can be categorized into low and high yield growth phenotypes, respectively. Experiments with high initial glucose concentrations >50 g/L resulted in lower glucose uptake rates and displayed high yields, close to reported yields in *S. cerevisiae* (0.51 g<sub>CDW</sub>/g<sub>glc</sub> [44]). Osmotic stress is absent because the yield stayed high with increasing glucose levels. The correlation between substrate uptake rate and growth rate (Figure 3A), for the high yield experiments, was strong, R<sup>2</sup>=0.99, with a realistic growth associated yield (slope) of 0.47 g<sub>CDW</sub>/g<sub>glc</sub>. While experiments ‘130v1’ and ‘130v2’ displayed disparate substrate uptake and growth rate despite similar initial glucose concentrations, their yield is comparable indicating a similar metabolic state. By contrast, cells of the low yield experiment (ID ‘2196\_glc’) imported glucose with the highest rate but grew with a low yield suggesting a distinct metabolic state.

The maintenance parameter of the high yield experiments was calculated as the x-axis interception, i.e., glucose uptake in the absence of growth, and resulted in 0.2 mmol/g<sub>CDW</sub>/h, comparable to *S. cerevisiae* (0.2 mmol/g<sub>CDW</sub>/h [45]). While the growth predictions of iUma22 are correct for high yield phenotypes, the growth of the low yield condition is over-predicted (ID: ‘2196\_glc’) (Figure 3C). The optimal yield is reached at a glucose uptake rate of ~2 mmol/g<sub>CDW</sub>/h, higher uptake results in overflow or secondary metabolite reactions (e.g., itaconate) [46]. Simulations to reproduce growth yield at glucose uptake rates at 6 mmol/g<sub>CDW</sub>/h, required an increase of NGAM flux to ~9.5 mmol/g<sub>CDW</sub>/h and represents the ATP investment in the new metabolic state (not shown). Such an NGAM flux is growth prohibiting for the high yield states. Glucose sensing by the Hxt1 Glucose transporter contributes to virulence [43] and a high uptake rate potentially reports host penetration. This initiates a phenotypic adaptation with lower yields and secondary metabolite production to prevent and respond to host immune system attacks. With higher glucose concentrations the dominance of Hxt1 mediated glucose uptake might decrease, preventing Hxt1 induced virulence transition [47] and contributing to the observed effect of reduced glucose uptake with higher concentrations.

**Table 3.** Glucose batch growth experiments were performed and used from the literature [4]. The data provided growth and substrate uptake rates for testing iUma22 predictions. Growth results for each experiment is provided in Supplementary 6.

Source	ID	Strain	Initial Glc, g/L	Growth rate, /h	Substrate rate, mmol/g <sub>CDW</sub> /h	Yield, g <sub>CDW</sub> /g <sub>glc</sub>
This work	2196_glc	FB1	20	0.27+/-0.05	6.2+/-1.2	0.24
This work	2229v1	MB215	50	0.18+/-0.04	2.2+/-0.6	0.45
Becker et al.	50glc	MB215	54	0.08+/-0.02	1.22+/-0.36	0.36
This work	130v1	MB215	126	0.07+/-0.02	1.1+/-0.34	0.33
This work	130v2	MB215	132	0.04+/-0.01	0.74+/-0.18	0.3
Becker et al.	100glc	MB215	106	0.04+/-0.01	0.67+/-0.19	0.33
This work	200v1	MB215	203	0.02+/-0.01	0.33+/-0.08	0.33
This work	200v2	MB215	216	0.02+/-0.01	0.55+/-0.1	0.33



**Figure 3.** Growth characteristics of *U. maydis* glucose batch cultures from Table 3 and similarity to iUma22 predictions. A) Eight batch experiments on glucose were analyzed to extract growth- and glucose-uptake rates. The linear least squares correlation provides the biomass yield on glucose with  $0.47 \pm 0.03 \text{ gCDW/g}_{\text{glc}}$ , the interception of the x-axis provides the glucose maintenance uptake rate with  $0.2 \pm 0.01 \text{ mmol/gCDW/h}$ . The red-marked measurement (experiment ID '2196\_glc') displays a distinct phenotype and is not included in the correlation. The two inset figures exemplify the growth rate estimation by a logistic Verhulst equation for growth (green) and linear substrate uptake (blue) for experiment ID '50glc'. B) Glucose-uptake rate as a function of the initial glucose level. Initial substrate concentrations were inversely correlated with glucose uptake rates. C) Simulated and experimental growth rates, optimal predictions represented by the black line. The individual growth rate data is provided in Supplementary 6.

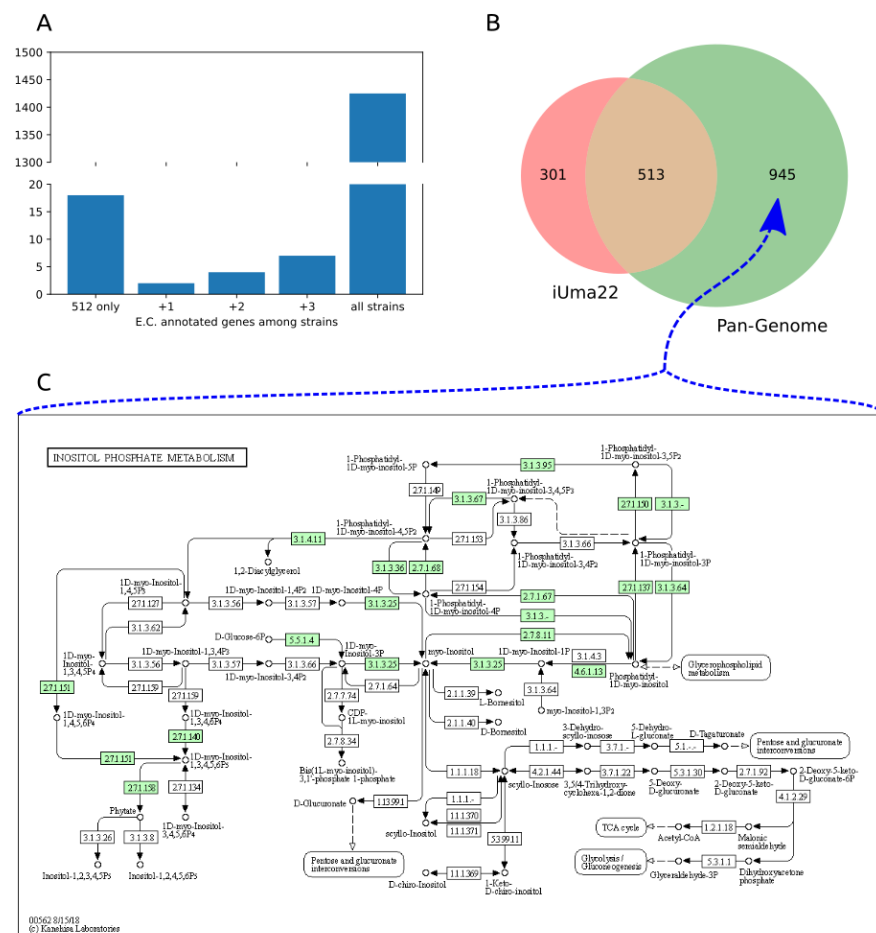
### 3.4. *U. maydis* Pan-Genome comparison

We used the available sequenced *U. maydis* strains [17] to compare the enzymatic gene inventory among strains and regarding iUma22. A pan-genome of strains 198, 482, 485 and 512 was constructed by means of EDGAR 3.0 resulting in 7838 coding genes of which 1458 are annotated with an E.C. number. We explored how many genes are shared among all strains and used strain 512 as a reference to identify unique enzyme coding genes and the proportion of shared genes to other strains (Figure 4A). The overwhelming number of genes is shared among all strains ('all strains' in Figure 4A), strain 512 has a number of unique enzyme coding genes are more likely shared with other strains because genes with more selective distributions (shared among 3 strains to 1 other strain), are getting less frequent (see also Figure 5 in Ullmann et al., 2022). We then evaluated how the genetic composition of iUma22 differed with respect to the *U. maydis* pan-genome of E.C. annotated genes (Figure 4B). Table 4 shows the top five pathways with the most enzyme annotations for iUma22 unique genes, shared genes and pan-genome unique genes with E.C. numbers. The majority of iUma22 unique genes belong to oxidative phosphorylation, the unique genes in the strain pan-genome belong to diverse central carbon metabolic pathways (Table 4). Particularly noteworthy is the inositol phosphate pathway (Figure 4C), not only because of the highest number of pan-genome unique metabolic capacity but also because inositol was a growth supporting substrate of the Biolog which was manually added to the model.



**Table 4.** Top five metabolic pathways with the highest number of missing genes in iUma22 compared to the strain Pan genome. The annotation is based on KAAS considering only KEGG pathways (KAAS outputs for iUma22-Unique, Shared and Pan-Unique provided as Supplementary 7–9).

iUma22-Unique	Shared	Pan-Unique
Oxidative phosphoryl. (42)	Purine (29)	Inositol phosphate (20)
TCA cycle (2)	Pyruvate metabolism (27)	Purine (12)
C5-branched metabolism (1)	Glycolysis (25)	N-Glycan biosynth. (11)
Nitrogen metabolism (1)	Gly, Ser, Thr metab. (24)	(GPI)-anchor biosynth. (11)
Starch and sucrose (1)	Val, Leu, Iso metab. (24)	Starch and sucrose (10)



**Figure 4.** Comparison of enzymes in *U. maydis* strain pan-genome and iUma22. A) E.C. annotated genes in strain 512 that are unique to 512 or shared with the other strains. B) Coverage of the genes in iUma22 of E.C. annotated genes in the pan-genome identified by KAAS. Table 4 show the top five pathways with most association for iUma22 unique, pan-genome unique and their intersection. C) The inositol phosphate metabolism contains with 20 genes the highest level of missing genes in iUma22 (Supplementary 9).

## 5. Conclusions

Here, we present iUma22, a genome scale metabolic model of *U. maydis*, which correctly simulates a large number of substrate phenotypes as well as glucose-based growth

rates. The pronounced metabolic reconfiguration at high glucose uptake rates and the decrease of glucose uptake at high concentrations indicates potential for biotechnological optimization. The model can be further used to identify biotechnological potential of metabolite over-production and to optimize metabolic engineering strategies. It can also be used to study metabolic shifts in different life cycles of the fungus during plant infection. While the reconstruction was performed based on model strain 521, the genome sequencing of additional *U. maydis* strains provided insight to additional metabolic pathways, which could be used to generate a pan-genome scale metabolic model of *U. maydis*.

#### Supplementary Materials:

The supplementary material is available on the GitHub page:

[https://github.com/iAMB-RWTH-Aachen/Ustilago\\_maydis-GEM](https://github.com/iAMB-RWTH-Aachen/Ustilago_maydis-GEM)

S1: Excel file for calculations of the *U. maydis* elemental composition.

S2: Matlab workspace file with variables and results of the optimization for the growth equation.

S3: Html output of full SBML quality scan with Memote.

S4: Excel file with the results of Biolog phenotype arrays PM1.

S5: Excel file with the results of Biolog phenotype arrays PM2A.

S6: Excel file with the growth experiments.

S7: Html output of KAAS for pathways unique to iUma22.

S8: Html output of KAAS for pathways shared among iUma22 and strain pan genome.

S9: Html output of KAAS for pathways unique to strain pan genome.

S10: iUma22 genome scale model SBML file.

**Author Contributions:** Conceptualization, L.M.B., C.L., T.Z., and U.W.L.; methodology, C.L. and U.W.L.; experiments L.U. and P.K.; pan-genome construction D.W.; validation, U.W.L. and C.L.; resources, L.M.B.; data curation, U.W.L. and C.L.; writing—original draft preparation, U.W.L.; writing—review and editing, U.W.L., L.U.; visualization, U.W.L.; supervision, T.Z. and U.W.L.; project administration, L.M.B.; funding acquisition, L.M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the Cluster of Excellence FSC 2186 "The Fuel Science Center".

**Data Availability Statement:** The data and the model is available on GitHub: [https://github.com/iAMB-RWTH-Aachen/Ustilago\\_maydis-GEM](https://github.com/iAMB-RWTH-Aachen/Ustilago_maydis-GEM)

#### Acknowledgments:

We appreciate support by Jennifer Scheuplein and Brigida Fabry regarding pathway visualization and Daniel Kaplan with the Biolog experiments.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

#### References

1. Castañeda de León, V.; Martínez-Carrera, D.; Morales, P.; Sobal, M.; Gil-Muñoz, A.; Severiano-Pérez, P.; Leal-Lara, H. Productivity and flavor of diverse genotypes of *Ustilago maydis* "cuitlacoche" for human consumption. *Fungal Biol.* **2019**, *123*, 481–488, doi:10.1016/J.FUNBIO.2019.04.003.
2. Hosseinpour Tehrani, H.; Becker, J.; Bator, I.; Saur, K.; Meyer, S.; Rodrigues Lóia, A.C.; Blank, L.M.; Wierckx, N. Integrated strain- And process design enable production of 220 g L<sup>-1</sup> itaconic acid with *Ustilago maydis*. *Biotechnol. Biofuels* **2019**, *12*, 1–11, doi:10.1186/S13068-019-1605-6/TABLES/1.
3. Becker, J.; Hosseinpour Tehrani, H.; Gauert, M.; Mampel, J.; Blank, L.M.; Wierckx, N. An *Ustilago maydis* chassis for itaconic acid production without by-products. *Microb. Biotechnol.* **2020**, *13*, 350–362, doi:10.1111/1751-7915.13525.
4. Becker, J.; Tehrani, H.H.; Ernst, P.; Blank, L.M.; Wierckx, N. An Optimized *Ustilago maydis* for Itaconic Acid Production at Maximal Theoretical Yield. *J. Fungi* **2021**, Vol. 7, Page 20 **2020**, *7*, 20, doi:10.3390/JOF7010020.

5. Wege, S.M.; Gejer, K.; Becker, F.; Bölker, M.; Freitag, J.; Sandrock, B. Versatile CRISPR/Cas9 Systems for Genome Editing in *Ustilago maydis*. *J. Fungi* **2021**, Vol. 7, Page 149 **2021**, 7, 149, doi:10.3390/JOF7020149.
6. Kämper, J.; Kahmann, R.; Bölker, M.; Ma, L.J.; Brefort, T.; Saville, B.J.; Banuett, F.; Kronstad, J.W.; Gold, S.E.; Müller, O.; et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nat.* **2006**, *444*, 97–101, doi:10.1038/nature05248.
7. Fang, X.; Lloyd, C.J.; Pálsson, B.O. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* **2020**, 1–13, doi:10.1038/s41579-020-00440-4.
8. Liebal, U.W.; Blank, L.M.; Ebert, B.E. CO<sub>2</sub> to succinic acid – Estimating the potential of biocatalytic routes. *Metab. Eng. Commun.* **2018**, 7, e00075, doi:10.1016/j.mec.2018.e00075.
9. Chan, S.H.J.; Simons, M.N.; Maranas, C.D. SteadyCom: Predicting microbial abundances while ensuring community stability. *PLOS Comput. Biol.* **2017**, 13, e1005539, doi:10.1371/JOURNAL.PCBI.1005539.
10. Ravikrishnan, A.; Blank, L.M.; Srivastava, S.; Raman, K. Investigating metabolic interactions in a microbial co-culture through integrated modelling and experiments. *Comput. Struct. Biotechnol. J.* **2020**, 18, 1249–1258, doi:10.1016/J.CSBJ.2020.03.019.
11. Sambamoorthy, G.; Raman, K. Deciphering the evolution of microbial interactions: in silico studies of two-member microbial communities. *bioRxiv* **2022**, 2022.01.14.476316, doi:10.1101/2022.01.14.476316.
12. Tatusova, T.; Ciufo, S.; Fedorov, B.; O'Neill, K.; Tolstoy, I. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* **2014**, 42, doi:10.1093/NAR/GKT1274.
13. Frishman, D.; Albermann, K.; Hani, J.; Heumann, K.; Metanowski, A.; Zollner, A.; Mewes, H.W. Functional and structural genomics using PEDANT. *Bioinformatics* **2001**, 17, 44–57, doi:10.1093/BIOINFORMATICS/17.1.44.
14. Karp, P.D.; Latendresse, M.; Caspi, R. The Pathway Tools Pathway Prediction Algorithm. *Stand. Genomic Sci.* **2011**, 5, 424, doi:10.4056/SIGS.1794338.
15. Caspi, R.; Altman, T.; Billington, R.; Dreher, K.; Foerster, H.; Fulcher, C.A.; Holland, T.A.; Keseler, I.M.; Kothari, A.; Kubo, A.; et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2014**, 42, D459–D471, doi:10.1093/nar/gkt1103.
16. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, 32, D277–D280, doi:10.1093/NAR/GKH063.
17. Ullmann, L.; Wibberg, D.; Busche, T.; Rückert, C.; Müsgens, A.; Kalinowski, J.; Blank, L.M. Seventeen Ustilaginaceae High-Quality Genome Sequences Allow Phylogenomic Analysis and Provide Insights into Secondary Metabolite Synthesis. *J. Fungi* **2022**, 8, 269, doi:10.3390/jof8030269.
18. Wibberg, D.; Andersson, L.; Tzelepis, G.; Rupp, O.; Blom, J.; Jelonek, L.; Pühler, A.; Fogelqvist, J.; Varrelmann, M.; Schlüter, A.; et al. Genome analysis of the sugar beet pathogen *Rhizoctonia solani* AG2-2IIIB revealed high numbers in secreted proteins and cell wall degrading enzymes. *BMC Genomics* **2016**, 17, 1–12, doi:10.1186/S12864-016-2561-1/FIGURES/5.
19. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **2017**, 27, 722–736, doi:10.1101/GR.215087.116.
20. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **2014**, 9, e112963, doi:10.1371/JOURNAL.PONE.0112963.
21. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **2013**.
22. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, 9, 357–359, doi:10.1038/nmeth.1923.
23. Ter-Hovhannisyan, V.; Lomsadze, A.; Chernoff, Y.O.; Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **2008**, 18, 1979–1990, doi:10.1101/GR.081612.108.

24. Meyer, F.; Goesmann, A.; McHardy, A.C.; Bartels, D.; Bekel, T.; Clausen, J.; Kalinowski, J.; Linke, B.; Rupp, O.; Giegerich, R.; et al. GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* **2003**, *31*, 2187–2195, doi:10.1093/NAR/GKG312.
25. Takahara, T.; Minamoto, T.; Yamanaka, H.; Doi, H.; Kawabata, Z. Estimation of Fish Biomass Using Environmental DNA. *PLoS One* **2012**, *7*, e35868, doi:10.1371/JOURNAL.PONE.0035868.
26. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E. V.; Krylov, D.M.; Mazumder, R.; Smirnov, S.; Nikolskaya, A.N.; et al. The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* **2003**, *4*, 1–14, doi:10.1186/1471-2105-4-41/FIGURES/5.
27. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370, doi:10.1093/NAR/GKG095.
28. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964, doi:10.1093/NAR/25.5.955.
29. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E. V.; Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212, doi:10.1093/BIOINFORMATICS/BTV351.
30. Dieckmann, M.A.; Beyvers, S.; Nkouamedjo-Fankep, R.C.; Hanel, P.H.G.; Jelonek, L.; Blom, J.; Goesmann, A. EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure. *Nucleic Acids Res.* **2021**, *49*, W185, doi:10.1093/NAR/GKAB341.
31. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185, doi:10.1093/NAR/GKM321.
32. Cokelaer, T.; Pultz, D.; Harder, L.M.; Serra-Musach, J.; Saez-Rodriguez, J.; Valencia, A. BioServices: a common Python package to access biological Web Services programmatically. *Bioinformatics* **2013**, *29*, 3241–3242, doi:10.1093/BIOINFORMATICS/BTT547.
33. Hernández, A.; Cooke, D.T.; Lewis, M.; Clarkson, D.T. Fungicides and sterol-deficient mutants of *Ustilago maydis*: plasma membrane physico-chemical characteristics do not explain growth inhibition. *Microbiology* **1997**, *143*, 3165–3174, doi:10.1099/00221287-143-10-3165.
34. Ruiz-Herrera, J.; Leon, C.G.; Carabez-Trejo, A.; Reyes-Salinas, E. Structure and chemical composition of the cell walls from the haploid yeast and mycelial forms of *Ustilago maydis*. *Fungal Genet. Biol.* **1996**, *20*, 133–142, doi:10.1006/FGBI.1996.0028.
35. Klement, T.; Milker, S.; Jäger, G.; Grande, P.M.; Domínguez de María, P.; Büchs, J. Biomass pretreatment affects *Ustilago maydis* in producing itaconic acid. *Microb. Cell Fact.* **2012**, *11*, 1–13, doi:10.1186/1475-2859-11-43/FIGURES/6.
36. Bushell, M.E. Fungal physiology (2nd Edition). By David H. Griffin, J. Wiley & Sons, Inc., New York, 1994, x + 458 pp., price £66.00, US\$92.50. ISBN 0 471 59586 1. *J. Chem. Technol. Biotechnol.* **1995**, *64*, 105–105, doi:10.1002/JCTB.280640118.
37. Thiele, I.; Palsson, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **2010**, *5*, 93–121, doi:10.1038/nprot.2009.203.
38. Geiser, E.; Przybilla, S.K.; Engel, M.; Kleineberg, W.; Büttner, L.; Sarikaya, E.; Hartog, T. den; Klankermayer, J.; Leitner, W.; Bötker, M.; et al. Genetic and biochemical insights into the itaconate pathway of *Ustilago maydis* enable enhanced production. *Metab. Eng.* **2016**, *38*, 427–435, doi:10.1016/J.YMBEN.2016.10.006.
39. Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **2010**, *38*, D750, doi:10.1093/NAR/GKP889.
40. Sato, T.K.; Tremaine, M.; Parreiras, L.S.; Hebert, A.S.; Myers, K.S.; Higbee, A.J.; Sardi, M.; McIlwain, S.J.; Ong, I.M.; Breuer, R.J.; et al. Directed Evolution Reveals Unexpected Epistatic Interactions That Alter Metabolic Regulation and Enable

- Anaerobic Xylose Use by *Saccharomyces cerevisiae*. *PLOS Genet.* **2016**, *12*, e1006372, doi:10.1371/JOURNAL.PGEN.1006372.
41. Lu, H.; Li, F.; Sánchez, B.J.; Zhu, Z.; Li, G.; Domenzain, I.; Marcišauskas, S.; Anton, P.M.; Lappa, D.; Lieven, C.; et al. A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **2019**, *10*, 1–13, doi:10.1038/s41467-019-11581-3.
42. 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.; et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **2020**, *38*, 272–276, doi:10.1038/s41587-020-0446-y.
43. Schirawski, J. Invasion is sweet. *New Phytol.* **2015**, *206*, 892–894, doi:10.1111/nph.13397.
44. Uribe Larrea, J. -L.; Winter, J.; Goma, G.; Pareilleux, A. Determination of maintenance coefficients of *Saccharomyces cerevisiae* cultures with cell recycle by cross-flow membrane filtration. *Biotechnol. Bioeng.* **1990**, *35*, 201–206, doi:10.1002/BIT.260350211.
45. Verduyn, C.; Stouthamer, A.H.; Scheffers, W.A.; van Dijken, J.P. A theoretical evaluation of growth yields of yeasts. *Antonie van Leeuwenhoek* **1991**, *59*, 49–63, doi:10.1007/BF00582119.
46. Voll, A.; Klement, T.; Gerhards, G.; Büchs, J.; Marquardt, W. Metabolic Modelling of Itaconic Acid Fermentation with *Ustilago Maydis*. *Chem. Eng. Trans.* **2012**, *27*, 367–372, doi:10.3303/CET1227062.
47. Schuler, D.; Wahl, R.; Wippel, K.; Vranes, M.; Münsterkötter, M.; Sauer, N.; Kämper, J. Hxt1, a monosaccharide transporter and sensor required for virulence of the maize pathogen *Ustilago maydis*. *New Phytol.* **2015**, *206*, 1086–1100, doi:10.1111/nph.13314.