


## Article

# *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,†</sup> ,  
Thiemo Zambanini <sup>1</sup> and Lars M. Blank <sup>1,\*</sup> 

<sup>1</sup> iAMB-Institute of Applied Microbiology, ABBt, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany; lena.ullmann@rwth-aachen.de (L.U.); philipp.kohl@rwth-aachen.de (P.K.); thiemo.zambanini@rwth-aachen.de (T.Z.)

<sup>2</sup> Unseen Biometrics ApS, DK-2800 Kgs. Lyngby, Denmark; christian.lieven@gmx.de

<sup>3</sup> Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

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

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

† Current address: Institute of Bio- and Geosciences IBG-5, Computational Metagenomics, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.

**Abstract:** *Ustilago maydis* is an important plant pathogen that causes corn smut disease and serves as 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, and the biomass equation was derived from literature values and from the codon composition. The final model contains over 25% annotated genes (6909) 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 decrease in glucose uptake rate with rising glucose concentration. A pangenome of four different *U. maydis* strains highlighted missing metabolic pathways in iUma22. The new model allows for studies of metabolic adaptations to different environmental niches as well as for biotechnological applications.

**Keywords:** *Ustilago maydis*; genome-scale metabolic model; constraint-based model; biotechnology; COBRA; FBA; metabolism; itaconate



**Citation:** Liebal, U.W.; Ullmann, L.; Lieven, C.; Kohl, P.; Wibberg, D.; Zambanini, T.; Blank, L.M. *Ustilago maydis* Metabolic Characterization and Growth Quantification with a Genome-Scale Metabolic Model. *J. Fungi* **2022**, *8*, 524. <https://doi.org/10.3390/jof8050524>

Academic Editors: Michael H. Perlin, Barry J. Saville and Jan Schirawski

Received: 24 March 2022

Accepted: 17 May 2022

Published: 20 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 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 the tumors are also used as food [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 its pathogenic mechanisms as well as metabolic competencies [6]. Annotated genomes can be used to construct genome-scale metabolic models (GSMM), which serve

as a knowledgebase 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 in order to identify exchange metabolites, community stability, and metabolic properties that mark transitions from mutualism to parasitism [9–11].

Here, we present the first high-quality genome-scale metabolic model for *U. maydis*, called iUma22. Growth phenotype assays based on BIOLOG with 190 substrates were conducted to reveal the metabolic versatility of *U. maydis* for more realistic model predictions in native habitats. Moreover, growth kinetics across a range of high glucose concentrations were performed, which allowed for improved metabolic characterization during biotechnological fermentations. To judge the metabolic completeness of iUma22, as well as metabolic capacities in comparison to the *Ustilago* genus, a pangenome of annotated enzymes of different *U. maydis* strains was constructed. The model is freely available online via the Biomodels database ID: MODEL2203250001 and GitHub: [https://github.com/iAMB-RWTH-Aachen/Ustilago\\_maydis-GEM](https://github.com/iAMB-RWTH-Aachen/Ustilago_maydis-GEM) (accessed on 24 March 2022). The quality of the model was assessed using the Memote evaluation tool, as well as FROG reports for reproducibility (see Supplementary Materials).

## 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 the MetaCyc reactions database, a new Pathway/Genome Database (PGDB) was created. During pathway cleaning, reactions from other taxa are pruned, unless there were enzymes matching all of the reactions. Additional metabolic activity was identified using the 'Pathway Hole Filler' function and the 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 nonmetabolic pathways, were discarded. While PEDANT and MUMDB are discontinued, information on the sequence and annotation for *U. maydis* can be accessed via EnsemblFungi: [https://fungi.ensembl.org/Ustilago\\_maydis/Info/Index](https://fungi.ensembl.org/Ustilago_maydis/Info/Index) (accessed 24 March 2022) [17], MycoCosm: [https://mycocosm.jgi.doe.gov/Ustma2\\_2/Ustma2\\_2.home.html](https://mycocosm.jgi.doe.gov/Ustma2_2/Ustma2_2.home.html) (accessed on 24 March 2022), NCBI (genome assembly ID: 225285), and Uniprot (Proteome ID: UP000000561).

### 2.2. Strains Sequenced, Pangenome, KEGG Pathway Enrichment

To identify metabolic differences within the *U. maydis* strain family, a pangenome consisting of five *Ustilago maydis* strains was assembled, including strains 198, 482, 485, and 512 [18]. 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 in [19]. Run control was based on MinKNOW (Oxford Nanopore Technologies, Oxford, UK) with the 48-h sequencing run protocol. Base-calling was performed offline

using a Bonito assembly with canu v2.1.1 [20], contigs were polished with Pilon [21] for ten iterative cycles, and for read mapping, BWA-MEM [22] and Bowtie2 v2.3.2 [23] were used in the first and second five iterations, respectively.

Genes were predicted using GeneMark-ES 4.6.2. [24] and functionally annotated using a modified version of the genome annotation platform GenDB 2.0 [25] for eukaryotic genomes [26]. Similarity searches were conducted against COG [27], KEGG [16] and SWISS-PROT [28]. Identification of putative tRNA genes was conducted with tRNAscan-SE [29]. Completeness, contamination, and strain heterogeneity were estimated with BUSCO (v3.0.2 [30]) 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) [18]. The pangenome of all available *U. maydis* strains was calculated by means of EDGAR 3.0 [31]. The KEGG pathway annotation was performed by comparison of E.C. numbers in the pangenome 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 pangenome and iUma (iUmaANDpan), and only present in pangenome (panNOTiUma). The panNOTiUma list was exported as a fasta-file, and KAAS [32] 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 [33].

### 2.3. Biomass Equation and Growth/Non-Growth Maintenance

The composition of proteins, RNA, and DNA was estimated based on the respective protein and genome sequences, whereas the composition of lipids and the cell wall were results-mined from scientific articles [34,35]. The exact biomass composition of *U. maydis* is not available; however, the specific elemental composition [36] (and the biomass composition for fungi in general [37]) was used as a starting point, and linear programming was applied to approximate the total biomass composition (Supplement). The composition values of each monomer were converted into stoichiometric values [38]. For example, to determine the amino acid (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) (Equation (1)). Normalizing each AA-protein molarity by the overall sum yields the weight fraction of each AA (WP<sub>AA</sub>) (Equation (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) provides the stoichiometric factor (SF<sub>AA</sub>) (Equation (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*) (Equation (2)).

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

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

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

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, and the corresponding C-mole content was calculated [36]. Phosphorous and Sulfur were added from the elemental composition of *Sacharomyces cerevisiae*. The optimization followed the formula:

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

$$\text{subject to: } x_{lbi} \leq x_i \leq x_{ubi} \quad (5)$$

The rows of matrix *A* correspond to elements *C*, *H*, *O*, *N*, *P*, and *S*, whereas the columns correspond to the macromolecule types (protein, DNA, RNA, lipids, cell wall). Vector *b* represents the measured elemental biomass, *C*, *H*, *O*, and *N*, supplemented by the ele-

mental content of the *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 Materials). The ATP-associated growth-maintenance (GAM) with ATP-dependent glucose uptake was calculated as 31 mmol/g<sub>CDW</sub> by optimization of the sum of the squared errors of growth experiments (see Section 3). The nongrowth-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 (see Section 3.3).

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

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 h. To prepare the precultures, 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 h (Ecotron Incubation shaker, Infors HT AG, Bottmingen, Switzerland). The inoculation fluid was prepared with IFY-0 (1.2×), 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% for up to 168 h (Multitron Incubation shaker, Infors HT AG, Bottmingen, Switzerland). Microbial growth was measured with the SynergyMX (BioTek Instruments, Winooski, VT, USA) with an optical density of 600 nm. The BIOLOG raw data is available in the Supplements.

The threshold for positive growth was determined by examining the optical density (OD) histograms for each plate. A normal distribution at low OD values represents the OD range below positive growth (see Section 3.2). The final growth threshold of 0.4 a.u. (absorbance units) 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 [39] were conducted with strains MB215, FB1 mating type a1b1, and literature data was used to estimate growth rates and glucose uptake rates. The OD measurements of the growth data were converted into g<sub>CDW</sub>/L using the empirical relation from yeast of 0.62 g<sub>CDW</sub>/L /OD (BNID 111182, [40]). The growth rates were identified using a nonlinear fit of the biomass to the Verhulst equation,

$$X(t) = X_0 * C / (X_0 + (C - X_0) * \exp(-\mu * t)), \quad (6)$$

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 [41].

### 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, as well as a comparison to the community yeast model ([42], version 8.5.0). Whereas the community yeast model was more comprehensive, iUma22 had a higher gene-to-reaction ratio as well as gene-protein-reaction relationships (GPR), as we aimed to include well-connected metabolic

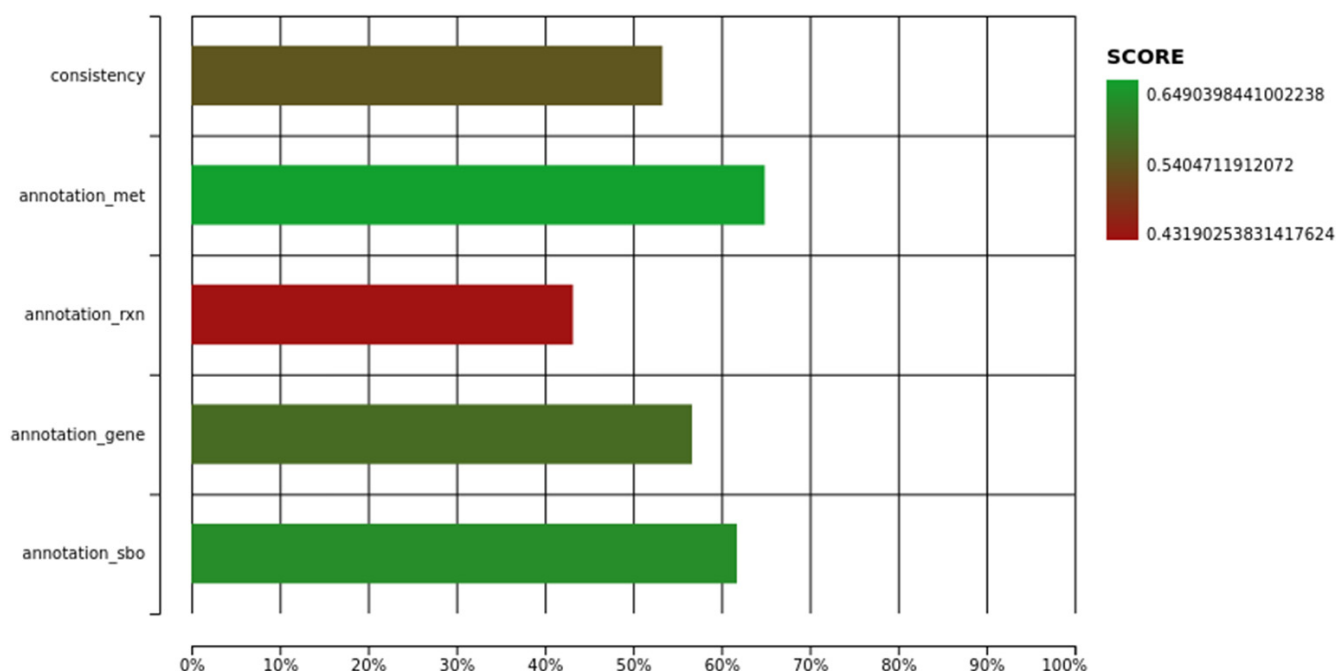
pathways with (predicted) annotated genes. *U. maydis* and *S. cerevisiae* have a similar number of predicted genes and, when assuming yeast 8.5.0 as a benchmark of metabolic representation, the iUma22 reached 70% of genes completeness. There are likely gaps in the secondary metabolism discussed in Ullmann et al. (2022) [18], as well as adaptations to the pathogenic lifestyle.

**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 [42]).

Component	iUma22	Yeast 7.6 <sup>1</sup>
Genes	814	1150
Metabolites	1233	2742
Reactions	1856	4058
Reactions with GPR	1434	2633
Predicted genes <sup>2</sup>	6909	6464

<sup>1</sup> <https://github.com/SysBioChalmers/yeast-GEM>, commit 24 June 2021. <sup>2</sup> <https://www.ncbi.nlm.nih.gov/datasets>, accessed on 4 February 2022.

The quality of iUma22 was tested with Memote with an overall performance of 57% [43] (Figure 1). Mass and charge balance, as well as metabolite connectivity, showed high quality, with scores of over 98%. Memote detected unbounded fluxes that reached 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 for the *S. cerevisiae* community model ([42], version 8.5.0), the stoichiometric consistency test also failed. Annotations for metabolites, reactions, and genes contained detailed and unique annotations. The community yeast model, developed for more than a decade, was evaluated by Memote with a total score of 65%.



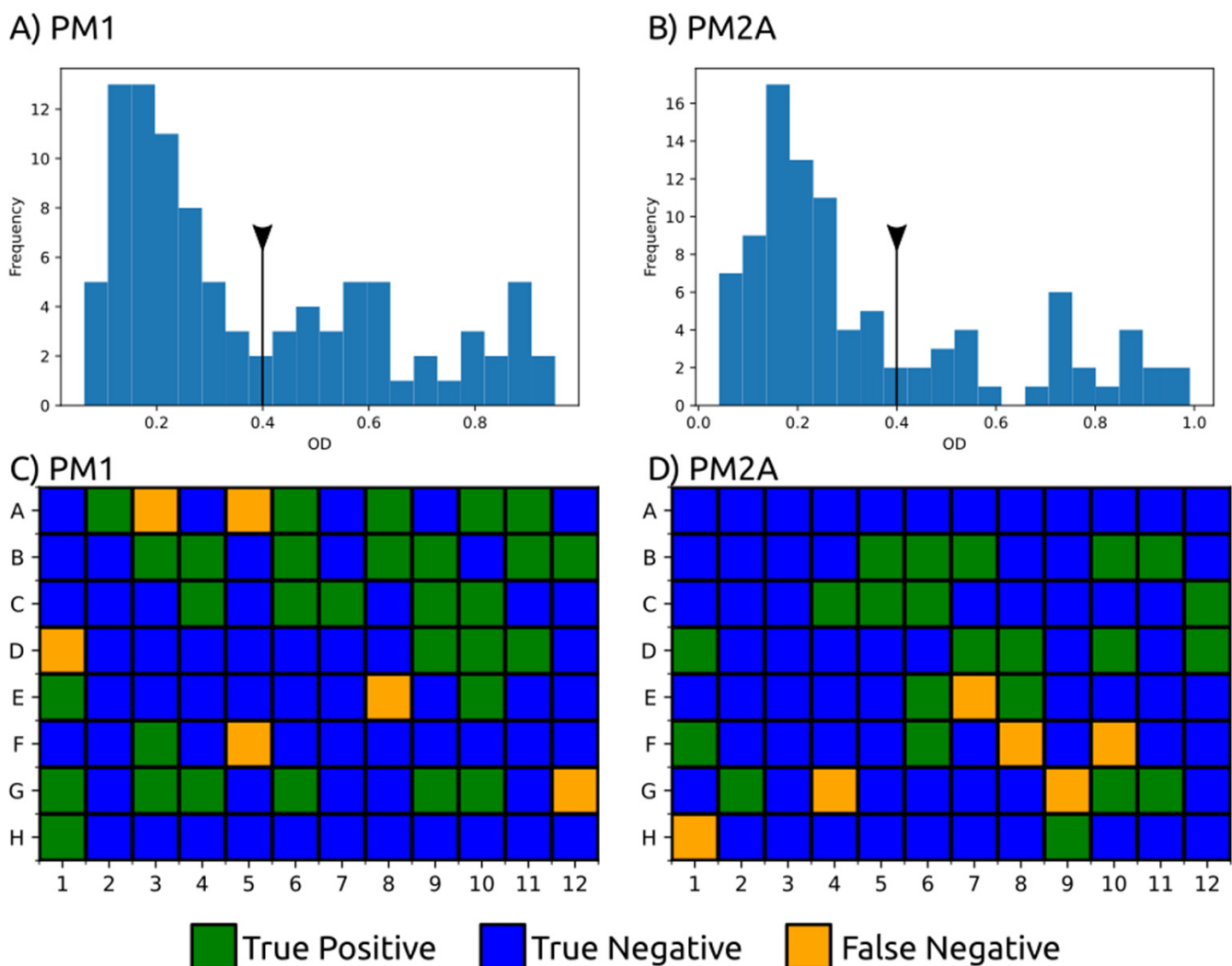
**Figure 1.** Memote quality report of iUma22 with total score of 57%. The full HTML report is provided as Supplement 3.

### 3.2. Carbon Substrate Tests with BIOLOG Phenotype Arrays

The model iUma22 correctly reproduced 96% of the growth phenotypes tested in BIOLOG carbon source assays. In these assays, each well of a 96-well plate was equipped



with a different substrate; we chose carbon substrate plates PM1 and PM2A from the manufacturer (an overview of the carbon source distribution in the wells is provided in the Supplements). Substrate utilization was tested photometrically, and we chose an OD threshold of 0.4 a.u. for growth, a value right after the apparent normal distribution of nongrowth for low final OD values (Figure 2). The threshold was a compromise to include growth for glycine dipeptides (PM1: E1, G1, G6, H1), but also included TCA cycle intermediates (succinate (PM1:A5), fumarate (PM1:F5), aspartate (PM1:A7), and 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 included di- and oligosaccharide metabolism and methylated central carbon metabolites (detailed list in the Supplements). Overall, growth took place in 52 wells (36 in PM1 and 28 in PM2A). iUma22 was manually adjusted to reproduce the majority of the growth phenotypes (Figure 2).



**Figure 2.** BIOLOG phenotype experiments with carbon sources from PM1 and PM2A. Growth was evaluated by OD 600 after 144 h for PM1 (A) and 288 h 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. Fifty-two substrates were correctly predicted to growth (true positive, green), and 128 were correctly assigned to nongrowth by iUma22 (true negative, yellow) in plates PM1 (C) and PM2A (D). Twelve substrates could not be balanced to enable growth in iUma22 (false negative). Results of PM1 and PM2A and an overview of the substrates on the plates are provided as Supplementary Materials.

While the majority of the substrates 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, such as lactate (PM1:B9), malate (PM1:G12), and succinate (PM2A:F10), were positive (Figure 2). The degradation of arginine (PM2A:G4), isoleucine (PM2A:G9), and ornithine (H1) was dysfunctional in iUma22, although proline had been predicted to enable growth. These metabolites are carbon intermediates 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. False positive growth predictions indicated exchange and transport reactions of metabolites in the model that could not actually be imported. Thus, growth predictions of the corresponding metabolites 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 (Table 3). Note that the model considers ATP-consuming glucose uptake via Hxt1 [44]. Experiments with high initial glucose concentrations of >50 g/L resulted in lower glucose uptake rates, which is not associated with osmotic stress, because the yield stayed high with increasing glucose levels. The correlation between substrate uptake rate and growth rate (Figure 3A) was strong,  $R^2 = 0.99$ , with a realistic growth-associated yield (slope) of  $0.47 \text{ g}_{\text{CDW}}/\text{g}_{\text{glc}}$ . While experiments '130v1' and '130v2' displayed disparate substrate uptake and growth rates despite similar initial glucose concentrations, their yield was comparable, indicating a similar 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  $0.2 \text{ mmol}/\text{g}_{\text{CDW}}/\text{h}$ , comparable to *S. cerevisiae* ( $0.2 \text{ mmol}/\text{g}_{\text{CDW}}/\text{h}$  [45]). The growth predictions of iUma22 are closely captured in the experimental results (Figure 3C).

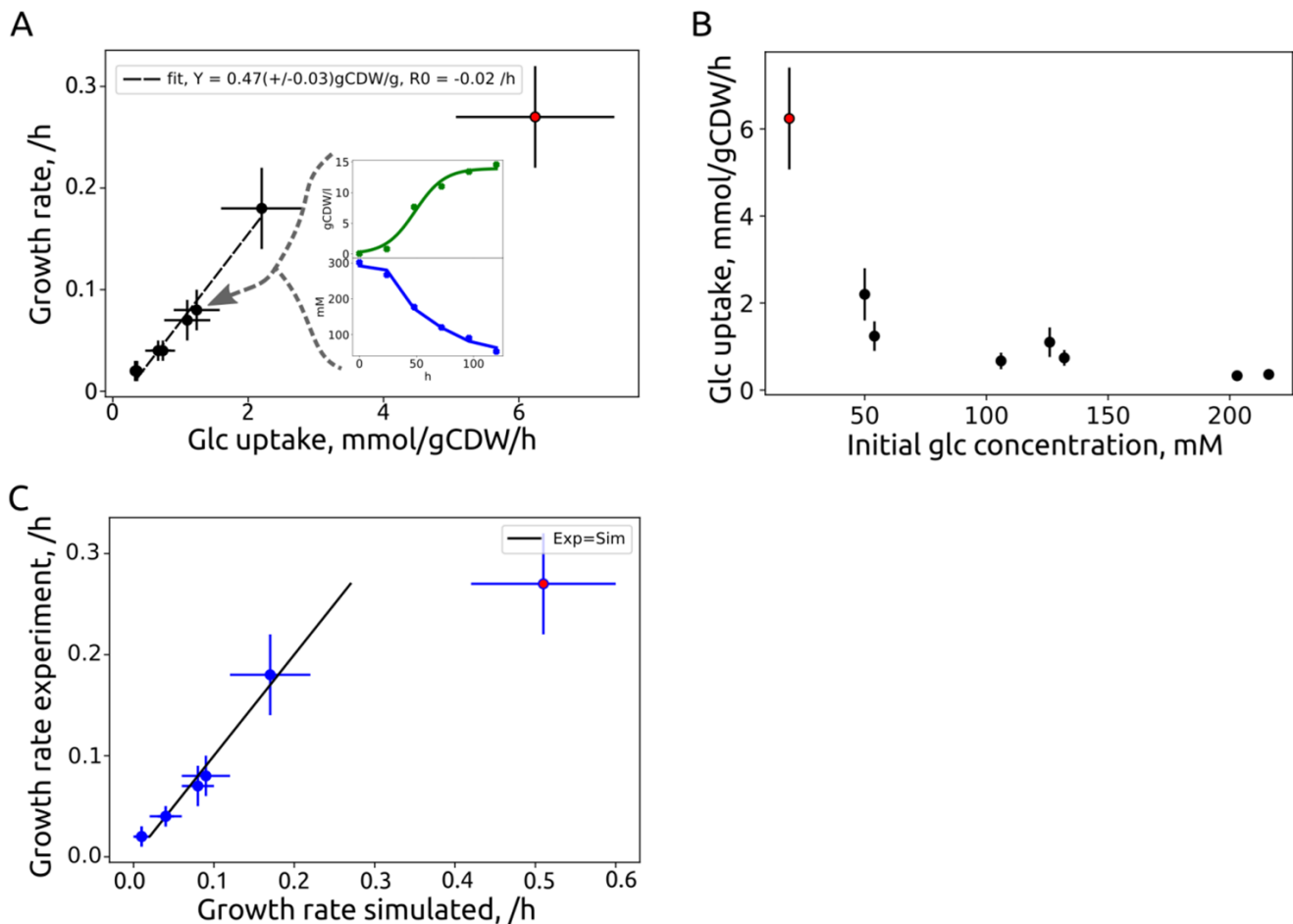
**Table 3.** Glucose batch growth experiments were performed and used from the literature [4]. The data provided growth and substrate uptake rates for testing iUma22 predictions. Growth results for each experiment is provided in the Supplements.

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	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

### 3.4. *U. maydis* Pangenome Comparison

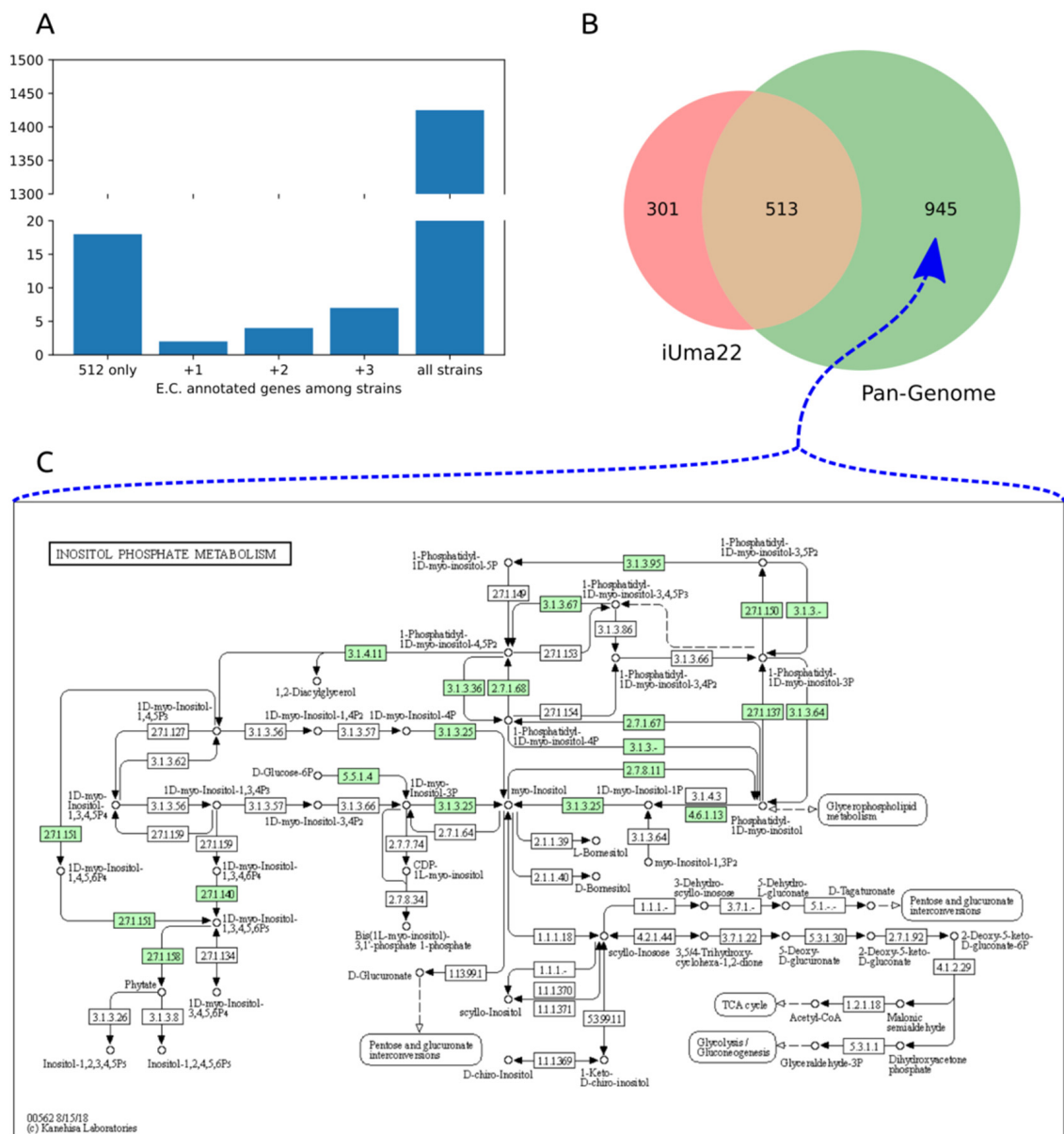
We used the available sequenced *U. maydis* strains [18] to compare the enzymatic gene inventory among strains and regarding iUma22. A pangenome of strains 198, 482, 485, and 512 was constructed by means of EDGAR 3.0, resulting in 7838 coding genes, of which 1458 are annotated with an E.C. number. We 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). An overwhelming number of genes is shared among all strains ('all strains' in Figure 4A). Strain 512 has a number of unique enzyme-coding genes that are more likely shared with other strains because genes with more selective distributions (shared among three strains to one other strain) are becoming less frequent (see also Ullmann et al. [18], 2022). We then evaluated how the

genetic composition of iUma22 differed with respect to the *U. maydis* pangenome 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 pangenome-unique genes with E.C. numbers. The majority of iUma22-unique genes belong to oxidative phosphorylation, and the unique genes in the strain pangenome 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 pangenome unique metabolic capacity but also because inositol was a growth-supporting substrate of the BIOLOG, which was manually added to the model.



**Figure 3.** Growth characteristics of *U. maydis* glucose batch cultures from Table 3 and similarity to iUma22 predictions. (A) Seven 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  $0.47 \pm 0.03 \text{ gCDW/g}_{\text{glc}}$ , and the interception of the x-axis provides the glucose maintenance uptake rate with  $0.2 \pm 0.01 \text{ mmol/gCDW/h}$ . 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 indicating an inverse correlation. (C) Simulated and experimental growth rates, with optimal predictions represented by the black line. The individual growth rate data is provided in the Supplements.





**Figure 4.** Comparison of enzymes in *U. maydis* strain pangene 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 pangene identified by KAAS. Table 4 show the top five pathways with the most association for iUma22-unique, pangene-unique, and their intersection. **(C)** The inositol phosphate metabolism contains 20 genes with the highest level of missing genes in iUma22 (Supplements).

**Table 4.** Top five metabolic pathways with the highest number of missing genes in iUma22 compared to the strain pangene. The annotation is based on KAAS, considering only KEGG pathways (KAAS outputs for iUma22-unique, -shared, and pangene-unique provided as Supplementary Materials).

iUma22-Unique	Shared	Pan-Unique
Oxidative phosphor. (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)

#### 4. 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 decrease in glucose uptake at high concentrations indicates potential for biotechnological optimization. The model can be further used to identify the biotechnological potential of metabolite overproduction 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 into additional metabolic pathways, which could be used to generate a pangenome-scale metabolic model of *U. maydis*.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jof8050524/s1>. 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: Zip file with FROG report. S5: Excel file with the results of BIOLOG phenotype arrays PM1. S6: Excel file with the results of BIOLOG phenotype arrays, PM2A. S7: PDF with BIOLOG substrate overview on plates PM1 and 2A. S8: Excel file with reactions added on the basis of the BIOLOG experiments. S9: Excel file with the growth experiments. S10: HTML output of KAAS for pathways unique to iUma22. S11: HTML output of KAAS for pathways shared among iUma22 and strain pan genome. S12: HTML output of KAAS for pathways unique to strain pan genome. S13: 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.; pangenome 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. and 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".

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data and the model are available on GitHub: [https://github.com/iAMB-RWTH-Aachen/Ustilago\\_maydis-GEM](https://github.com/iAMB-RWTH-Aachen/Ustilago_maydis-GEM) (accessed on 24 March 2022).

**Acknowledgments:** We appreciate the support of Jennifer Scheuplein and Brigida Fabry regarding pathway visualization and BIOLOG analysis; 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. de León, V.C.; 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. [[CrossRef](#)] [[PubMed](#)]
2. Tehrani, H.H.; Becker, J.; Bator, I.; Saur, K.; Meyer, S.; Lóia, A.C.R.; 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*, 263. [[CrossRef](#)] [[PubMed](#)]
3. Becker, J.; Tehrani, H.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. [[CrossRef](#)] [[PubMed](#)]
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* **2020**, *7*, 20. [[CrossRef](#)]

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**, *7*, 149. [\[CrossRef\]](#)
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*. *Nature* **2006**, *444*, 97–101. [\[CrossRef\]](#)
7. Fang, X.; Lloyd, C.J.; Palsson, B.O. Reconstructing organisms in silico: Genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* **2020**, *18*, 731–743. [\[CrossRef\]](#)
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. [\[CrossRef\]](#)
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. [\[CrossRef\]](#)
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. [\[CrossRef\]](#)
11. Sambamoorthy, G.; Raman, K. Deciphering the evolution of microbial interactions: In silico studies of two-member microbial communities. *bioRxiv* **2022**, *14*, 476316. [\[CrossRef\]](#)
12. Tatusova, T.; Ciufu, S.; Fedorov, B.; O'Neill, K.; Tolstoy, I. RefSeq microbial genomes database: New representation and annotation strategy. *Nucleic Acids Res.* **2014**, *42*, D553–D559. [\[CrossRef\]](#) [\[PubMed\]](#)
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. [\[CrossRef\]](#)
14. Karp, P.D.; Latendresse, M.; Caspi, R. The Pathway Tools Pathway Prediction Algorithm. *Stand. Genom. Sci.* **2011**, *5*, 424–429. [\[CrossRef\]](#)
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. [\[CrossRef\]](#) [\[PubMed\]](#)
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. [\[CrossRef\]](#)
17. Yates, A.D.; Allen, J.; Amode, R.M.; Azov, A.G.; Barba, M.; Becerra, A.; Bhai, J.; Campbell, L.I.; Martinez, M.C.; Chakiachvili, M.; et al. Ensembl Genomes 2022: An expanding genome resource for non-vertebrates. *Nucleic Acids Res.* **2022**, *50*, D996–D1003. [\[CrossRef\]](#)
18. 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. [\[CrossRef\]](#)
19. 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* AG<sub>2-2</sub>IIIB revealed high numbers in secreted proteins and cell wall degrading enzymes. *BMC Genom.* **2016**, *17*, 245. [\[CrossRef\]](#)
20. 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. [\[CrossRef\]](#)
21. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
23. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [\[CrossRef\]](#) [\[PubMed\]](#)
24. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
25. 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. [\[CrossRef\]](#)
26. Takahara, T.; Minamoto, T.; Yamanaka, H.; Doi, H.; Kawabata, Z. Estimation of Fish Biomass Using Environmental DNA. *PLoS ONE* **2012**, *7*, e35868. [\[CrossRef\]](#)
27. 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 Bioinform.* **2003**, *4*, 41. [\[CrossRef\]](#)
28. 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. [\[CrossRef\]](#)
29. 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. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#)
31. 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–W192. [\[CrossRef\]](#) [\[PubMed\]](#)

32. 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. [[CrossRef](#)] [[PubMed](#)]
33. 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. [[CrossRef](#)] [[PubMed](#)]
34. 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. [[CrossRef](#)]
35. 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. [[CrossRef](#)]
36. Klement, T.; Milker, S.; Jäger, G.; Grande, P.M.; de María, P.D.; Büchs, J. Biomass pretreatment affects *Ustilago maydis* in producing itaconic acid. *Microb. Cell Fact.* **2012**, *11*, 43. [[CrossRef](#)]
37. 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. [[CrossRef](#)]
38. Thiele, I.; Palsson, B.O. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **2010**, *5*, 93–121. [[CrossRef](#)]
39. Geiser, E.; Przybilla, S.K.; Engel, M.; Kleineberg, W.; Büttner, L.; Sarikaya, E.; den Hartog, T.; Klankermayer, J.; Leitner, W.; Böcker, M.; et al. Genetic and biochemical insights into the itaconate pathway of *Ustilago maydis* enable enhanced production. *Metab. Eng.* **2016**, *38*, 427–435. [[CrossRef](#)]
40. 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–D753. [[CrossRef](#)]
41. 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. [[CrossRef](#)]
42. Lu, H.; Li, F.; Sánchez, B.J.; Zhu, Z.; Li, G.; Domenzain, I.; Marčišauskas, 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*, 3586. [[CrossRef](#)] [[PubMed](#)]
43. 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. [[CrossRef](#)] [[PubMed](#)]
44. Schirawski, J. Invasion is sweet. *New Phytol.* **2015**, *206*, 892–894. [[CrossRef](#)]
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. [[CrossRef](#)]