



pAo: A Consensus Genome-Scale Metabolic Model for *Aspergillus oryzae* Capturing Intraspecies Diversity

Downloaded from: <https://research.chalmers.se>, 2026-09-14 17:59 UTC

Citation for the original published paper (version of record):

Gilis, J., van der Luijt, C., Feller, M. et al (2026). pAo: A Consensus Genome-Scale Metabolic Model for *Aspergillus oryzae* Capturing Intraspecies Diversity. Computational and Structural Biotechnology Journal, 35(1).
<http://dx.doi.org/10.34133/csbj.0173>

N.B. When citing this work, cite the original published paper.

RESEARCH ARTICLE

pAo: A Consensus Genome-Scale Metabolic Model for *Aspergillus oryzae* Capturing Intraspecies Diversity

Jeroen Gilis^{1†}, Casper Robert Balten van der Luijt^{2,3,4†}, Marilena Feller³, Candela Sánchez-Girón Barba³, Morten Otto Alexander Sommer³, Leonie Johanna Jahn^{3*}, and Eduard Johannes Kerkhoven^{1,3,5*}

¹Department of Life Sciences, Chalmers University of Technology, 412 96 Gothenburg, Sweden. ²Department of Food Science, Faculty of Science, University of Copenhagen, 1958 Frederiksberg, Denmark. ³The Novo Nordisk Foundation Biotechnology Research Institute for the Green Transition, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark. ⁴National Food Institute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark. ⁵Science for Life Laboratory (SciLifeLab), 412 96 Gothenburg, Sweden.

*Address correspondence to: lejj@dtu.dk (L.J.J.); eduardk@chalmers.se (E.J.K.)

†These authors contributed equally to this work.

Aspergillus oryzae (koji mold) is a key microorganism in traditional food fermentations including soy sauce, sake, and miso and is important in novel culinary applications and modern biotechnology, such as sustainable meat alternatives and enzyme production. Despite its industrial importance, until recently, the most recent genome-scale metabolic model (GEM) for *A. oryzae* dated back to 2008 and was limited to a single strain (RIB40). Here, we present *pAo*, a pan-GEM for *A. oryzae*, integrating genomic data from 187 strains to capture species-wide metabolic diversity. Our model comprises 2,025 reactions, representing a 52% increase in metabolic coverage over the RIB40-based model and includes previously overlooked pathways, such as cytochrome P450-mediated xenobiotic metabolism and extended amino acid metabolism. Using this pan-GEM, we derived strain-specific GEMs and validated 8 of them through high-throughput phenotypic screening on 285 substrates. Growth experiments on 4 industrially relevant carbon sources revealed substantial interstrain metabolic diversity, although flux balance analysis indicated that this variability originates at the regulatory rather than network-structural level. This resource provides a foundation for informed strain selection for biotechnological applications and future metabolic engineering in *A. oryzae*.

Introduction

Aspergillus oryzae, commonly known as koji mold, is a cornerstone of traditional food fermentations, such as soy sauce, sake, and miso. Because of its long history of safe use in the food industry, *A. oryzae* was granted the status of a Generally Recognized as Safe organism by the US Food and Drug Administration and the World Health Organization [1–3]. This has enabled its expansion into modern biotechnology as a leading cell factory. For decades, *A. oryzae* has been renowned for its high secretion of native enzymes such as amylases and proteases [4–8] and its excellent capacity for recombinant protein production [9–11]. It is also widely used for producing organic acids, such as malic acid, kojic acid, and gluconic acid [12–16]. More recently, microbial foods using *A. oryzae* as a source of edible mycelium for sustainable meat alternatives have emerged [17–22], and its potential as a prebiotic and a probiotic in human diets and animal feed was demonstrated [23–27].

This versatility has established *A. oryzae* as a pivotal microorganism in both food science and industrial biotechnology.

However, fundamental research into the molecular components that govern its metabolism—genes, proteins, and metabolic pathways—is key to further optimizing its biotechnological use and novel culinary applications [28,29]. In this context, genome-scale metabolic models (GEMs) have proven particularly powerful [30,31]. A GEM is a knowledge base constructed from data on an organism's genome, the links between genes, proteins, and reactions, as well as the stoichiometry of these reactions. By incorporating biological constraints, such as nutrient availability and protein localization within the cell, GEMs can be used to perform flux balance analysis (FBA), which predicts the flux through each metabolic reaction in the organism [32]. These predicted fluxes provide a computational model of cellular metabolism at the genome scale, enabling the prediction of phenotypes such as gene essentiality, growth rate, and metabolite production rates. In addition, GEMs can inform metabolic engineering strategies that modify the organism's genome to enhance the production titer, rate, and yield of a metabolite of interest [33,34].

Citation: Gilis J, van der Luijt CRB, Feller M, Barba CS-G, Sommer MOA, Jahn LJ, Kerkhoven EJ. *pAo*: A Consensus Genome-Scale Metabolic Model for *Aspergillus oryzae* Capturing Intraspecies Diversity. *Comput. Struct. Biotechnol. J.* 2026;35(1):Article 0173. <https://doi.org/10.34133/csbj.0173>

Submitted 19 March 2026

Revised 1 July 2026

Accepted 1 July 2026

Published 19 August 2026

Copyright © 2026 Jeroen Gilis et al. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License (CC BY 4.0).

With continuous improvements in molecular biology and whole-genome sequencing, the information required for constructing GEMs has become increasingly abundant and more accurate. For model organisms such as *Escherichia coli* [35,36], *Saccharomyces cerevisiae* [37,38], *Caenorhabditis elegans* [39,40], and *Homo sapiens* [41–43], dedicated efforts have been made to continuously update and refine GEMs with the latest data. Beyond these canonical models, GEMs are increasingly being developed for nonmodel and industrially relevant organisms, including several filamentous fungi [44–48]. For *A. oryzae*, only 2 GEMs have been reported to date. The first, and most widely used, was published in 2008 [49]. More recently, an updated GEM was developed [50]. However, this model was specifically designed to provide a detailed representation of cordycepin metabolism. It was constructed using the GEM of the entomopathogenic fungus *Cordyceps militaris* as a template, supplemented with additional reactions from the 2008 *A. oryzae* model. Consequently, this updated GEM is not ideally suited to function as a comprehensive, general purpose knowledge base for *A. oryzae* metabolism.

Moreover, strain-specific [51–53] and cell-type-specific [54–57] GEMs have become available for model organisms, enabling the capture of intraspecies diversity in metabolism. However, GEMs developed for *A. oryzae* remain limited to only 2 strains: the laboratory strain RIB40 and the engineered strain AoCordy-T1 [49,50].

The first objective of this manuscript is to build a knowledge base for the metabolism of *A. oryzae* at the species level, which we refer to as the pan-GEM for *A. oryzae* (*pAo*). Secondly, we integrate genotype information from 187 strains to derive strain-specific GEMs based on *pAo*. We then perform a high-throughput phenotype characterization and growth experiment to validate these strain-specific models and to assess intraspecies metabolic diversity. Finally, we argue that this extensive set of strain-specific GEMs provides an ideal foundation for making informed decisions about which strain offers the most potential for cell factory development and food fermentation processes.

Results

Mapping the metabolic potential of *A. oryzae* with a pan-GEM

We constructed a pan-GEM for *A. oryzae*, *pAo*, to capture the metabolic potential of the species. This model expands substantially upon the previous *A. oryzae* GEM (*oryzae2008*), which was limited to a single laboratory strain, RIB40 [49]. Our reconstruction pipeline integrated genomic data from 187 *A. oryzae* strains (Fig. 1A and Table S1). We predicted protein sequences and inferred their metabolic functions through homology searches against 3 curated *Aspergillus* GEMs (from *A. oryzae* RIB40, *Aspergillus niger*, and *A. fumigatus*) [46,49,58].

This analysis identified 11,949 orthologous protein sequence clusters, classified as strict core (present in all 187 strains), soft-core (>95%), shell (5% to 95%), cloud (<5% but >1 strain), or strain specific (Fig. S1), comprising 6,959, 1,882, 1,687, 611, and 810 clusters, respectively. For clusters with orthologs in the established GEMs, we assigned the corresponding metabolic reactions. For the remaining clusters, we performed a BLAST [59] search against the Kyoto Encyclopedia of Genes and Genomes (KEGG) [60] reaction database to identify potential new metabolic functions.

The resulting model, *pAo*, comprises 2,025 reactions, with 1,426 obtained from *oryzae2008*, 86 incorporated from other *Aspergillus* GEMs, and 454 from KEGG (Fig. 1B). An additional 59 reactions were included as gap-filling based on experimental validation to ensure growth on specific substrates (Table S2). This curated network, catalyzed by 1,397 genes and encompassing 1,824 metabolites, serves as the *pAo* and substantially expands on the *oryzae2008* model (Fig. 1C) [49]. Most notably, the number of biochemical reactions (this excludes transport and exchange reactions) has increased from 1,069 in *oryzae2008* to 1,652 in *pAo*.

To assess the coverage of *pAo* relative to the previous *A. oryzae* GEM, we examined to what extent *oryzae2008* genes are represented in the *pAo* core. Of the 1,160 genes in *oryzae2008*, 69 are noncore (neither strict core nor soft core) in *pAo*; these genes catalyze 124 reactions in the template model, 30 of which relate to sterigmatocystin/aflatoxin biosynthesis, a pathway known to be strain variable and therefore not expected to form part of the *pAo* core [3,61,62]. Furthermore, although *pAo* is a superset of *oryzae2008*, *oryzae2008* lacks 49 genes that form part of the *pAo* core, which together catalyze 124 reactions, 26 of which relate to fatty acid biosynthesis (Fig. S2).

To facilitate downstream pathway analyses, each reaction was assigned a single pathway annotation according to the subsystem information from the KEGG database [60]. Comparing these annotations to *oryzae2008* revealed additional metabolic pathways in *pAo* that were likely overlooked in the earlier model (Fig. 1D). The most notable expansions were in fatty acid biosynthesis, comprising 38 reactions encoded by 4 genes, and in metabolism of xenobiotics (foreign compounds) by cytochromes P450—a superfamily of heme-containing monooxygenases [63]—comprising 34 novel reactions catalyzed by 7 genes. Although thoroughly documented in *A. oryzae* [64], this class of enzymes was absent in *oryzae2008*. We also observed 4 pathways involved in amino acid metabolism. While all enzymes that comprise the core of amino acid metabolism were already included in *oryzae2008*, our approach has identified 54 additional genes that extend amino acid metabolism in *A. oryzae*, for example, opening additional pathways for direct-arginine metabolism toward butanoate or glyoxylate metabolism (Fig. S3).

Strain-specific genome-scale models for *A. oryzae* reveal substantial intraspecies metabolic diversity

We used our pan-GEM to construct strain-specific GEMs for all 187 *A. oryzae* strains. Each model represents a subset of the reactions of the pan-GEM for which the corresponding genes are present in that specific strain. Spontaneous reactions and those with unknown genetic basis were included in all models. Analysis revealed a highly conserved metabolic core, with 1,152 genes (82.5%) and 1,118 gene–protein–reaction rule (GPR)-associated reactions (71.1%) present in more than 95% of strains (Fig. 2A). Despite this, we observed substantial strain-specific (90 GPR-associated reactions) metabolic content. Reduced-dimensional representation of the reactions present in each strain-specific GEM shows that *A. oryzae* strains cluster primarily according to their taxonomic clades (Fig. 2B), indicating that evolutionary relationships are reflected in metabolic network topology. This is corroborated by higher within-clade than between-clade reaction content similarity and by pairwise Jaccard similarity of reaction content (Figs. S4 and S5).

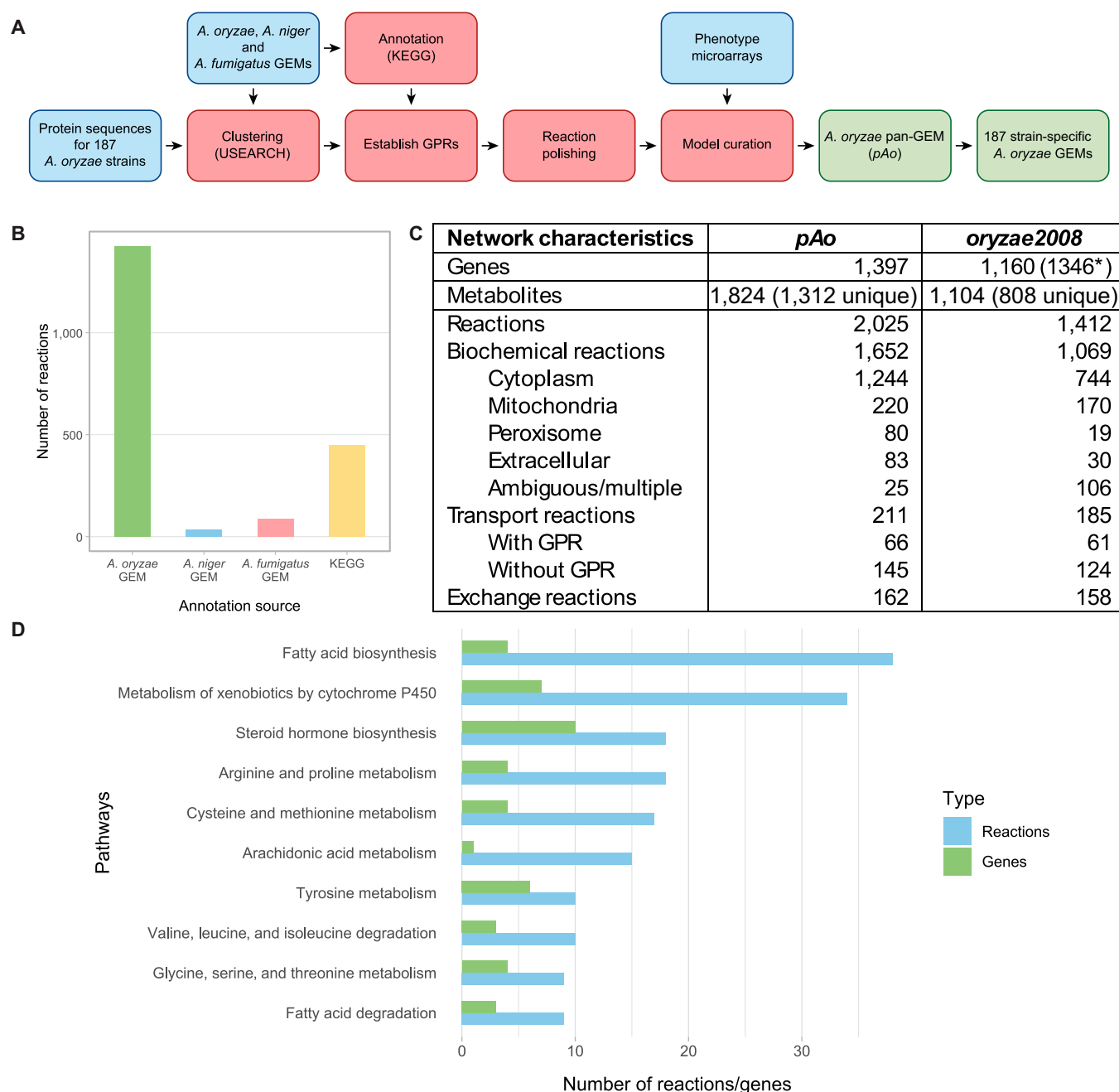


Fig. 1. Reconstruction of a pan-genome-scale metabolic model (GEM) for *Aspergillus oryzae* (*pAo*). (A) Workflow for pan-GEM reconstruction. Protein sequences from 187 *A. oryzae* strains and 3 *Aspergillus* species with existing GEMs were clustered using USEARCH. Metabolic reactions were assigned on the basis of homology to known models and Kyoto Encyclopedia of Genes and Genomes (KEGG), followed by manual polishing of compartments, directionality, and metabolite names. The model was curated using phenotype microarray data, resulting in a pan-GEM. Strain-specific models were derived by subsetting the pan-GEM based on individual strain's gene content. (B) Bar chart indicating the number of reactions coming from each of the different reaction databases used to construct the pan-GEM. (C) Network characteristics comparing the pan-GEM and *oryzae2008*, demonstrating the expanded coverage of genes, metabolites, and reactions in the pan-GEM. Genes are counted as the number of unique orthologous gene clusters as obtained from the USEARCH algorithm, while the value with the asterisks indicates the number of individual genes for *oryzae2008*. Metabolites present in multiple compartments are either counted separately or as one (unique). GPR, Gene-protein-reaction rule. (D) Bar chart displaying the number of reactions (blue) and genes (green) involved in the top 10 enriched pathways in *pAo* compared to the *oryzae2008* model.

High-throughput phenotyping drives model curation and reveals consistent substrate utilization patterns

To assess phenotypic diversity, we performed high-throughput phenotyping of 8 phylogenetically diverse *A. oryzae* strains on 285 substrates (190 carbon and 95 nitrogen sources) on phenotypic microarrays (PMs). Clustering of the samples on the

PMs indicated that biological replicates cluster together, confirming that interstrain differences exceed experimental variation (Fig. 3 and Fig. S6). Clustering all 285 substrates (Fig. S6) additionally revealed several coherent substrate clusters, particularly among nitrogen sources. For instance, one cluster contained key metabolites in arginine metabolism (putrescine,

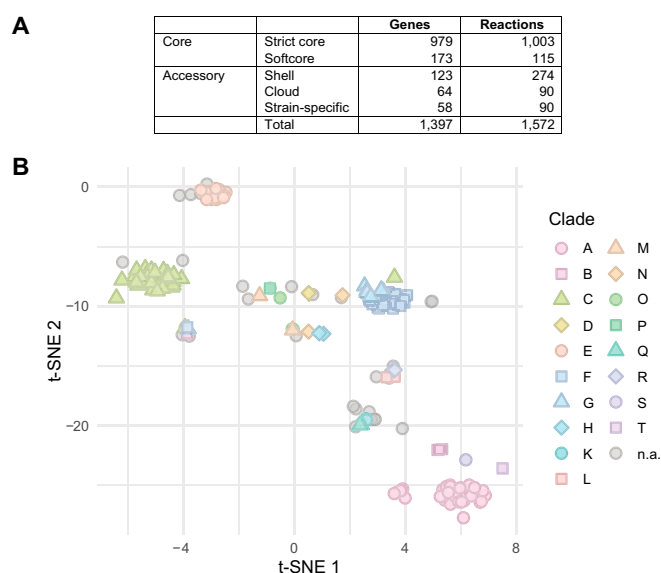


Fig. 2. Analysis of 187 strain-specific *Aspergillus oryzae* genome-scale metabolic models (GEMs). (A) Conservation of gene–protein–reaction rule (GPR)-associated reactions and genes across the strain-specific models. Reactions and genes were classified by the percentage of strains in which they are present: strict core (100% of strains), softcore (>95% but not all strains), shell (5% to 95% of strains), cloud (<5% but more than one strain), and strain-specific (present in only one strain). (B) t-Distributed stochastic neighbor embedding (t-SNE) of the 187 strains based on the reaction pathway information. Strains are colored according to phylogenetic clade.

agmatine, and ornithine), and another group contained most of the tested dipeptides.

We then used the PM results to curate the 8 strain-specific GEMs. We selected 45 substrates that unambiguously supported metabolic activity across most replicates and performed FBA, using adenosine triphosphate (ATP) hydrolysis as a proxy for metabolic activity to specify the cellular objective function. For substrates on which FBA failed to predict metabolic activity, we systematically (a) assessed whether including a transport reaction for the substrates allowed for supporting metabolic activity, if appropriate; (b) assessed whether metabolic activity could be restored through gap-filling of reactions from the *A. oryzae* pan-GEM, and (c) manually inspected pathways involved in metabolizing the substrate to identify missing reactions and perform gap-filling. The latter was achieved by performing a manual literature search on the metabolism of the target compound in related fungi, further aided by a manual inspection of the metabolic pathway maps provided by KEGG. This led to the addition of 59 reactions to *pAo* (Table S2), ensuring that all 8 curated models accurately predict growth on the 45 validation substrates.

Strain-specific substrate preferences guide optimal carbon source selection

To identify the optimal carbon source for each strain, we experimentally assessed the growth of the 8 phylogenetically diverse *A. oryzae* strains on 4 industrially relevant carbon sources: glucose, maltose, glycerol, and xylose (Table S3). Comparing biomass production and maximum growth rates for each strain across these substrates revealed distinct substrate preferences (Fig. 4 and Figs. S7 and S8). For instance, strain NRRL 35890 showed exceptionally high growth rates on xylose but slower growth on glycerol, a pattern also observed for NRRL 471. In

contrast, strains NRRL 3488, NRRL 5592, and RIB40 appeared to favor glycerol as a substrate, achieving higher growth rates. These results provide a practical guide for selecting the best carbon source to maximize growth for a given strain, which is crucial for industrial applications where strain identity is fixed.

Substrate-specific strain performance informs selection of optimal strains for industrial substrates

Next, we evaluated which strain performs best on each carbon source to inform strain selection for industrial applications. We observed substantial interstrain variation in maximum growth rates across all carbon sources (Fig. 5). Strain NRRL 5589 showed consistently slow growth rates, while NRRL 35890 was consistently among the fastest. However, performance was highly substrate dependent. For example, on xylose, NRRL 35890 outperformed other strains, whereas on glycerol, NRRL 2217, NRRL 3488, NRRL 5592, and RIB40 were top performers alongside NRRL 35890. This highlights that strain selection can be tailored to the intended substrate to optimize utilization and growth rates.

To place the observations from these growth experiments in context, we compared our data with that used by Vongsangnak *et al.* [49]. It should be noted that although their model was developed for strain RIB40, the underlying experimental data originated from a study on another strain (A1560), and the reported xylose yield was inferred from a study on *A. niger* and *Aspergillus nidulans* [49,65,66]. Accordingly, for comparative purposes, we used overall mean values of maximum growth rate and biomass across all strains to represent the species as a whole.

In their study, biomass yields show minimal variation after carbon normalization, whereas our data indicated comparable maximum biomasses across all substrates except glycerol, which supported lower biomass production (Fig. S9). Furthermore, while Vongsangnak *et al.* [49] reported that *A. oryzae* achieved the highest maximum growth rate on glucose, followed by maltose, glycerol, and xylose, our experiments did not reveal a clear hierarchy among carbon sources (Fig. S10). Altogether, these results emphasize the need for substrate-specific strain evaluation.

Flux analysis reveals a conserved metabolic solution space across the strain-specific models

The growth experiments revealed substantial interstrain variation in growth across the 4 carbon sources (Fig. 5). To assess whether the strain-specific models reproduced this phenotypic diversity, we used them to simulate growth by FBA. The models predicted identical biomass yields for all 8 strains and thus did not capture the experimentally observed variation (Fig. 6A). To establish whether this reflected a genuine equivalence of the reconstructed networks, rather than merely the limited discriminatory power of a single optimal solution, we characterized the metabolic solution space of the 8 curated models in detail.

Parsimonious FBA (pFBA) showed that, under each carbon source, the 8 models converged on an essentially identical optimal flux state. The predicted biomass yield per carbon mole, the ATP turnover per unit biomass, and the total flux required per unit growth (Fig. 6B) each varied by less than 0.3% across strains, whereas the differences between substrates were an order of magnitude larger and consistent across every strain (Fig. 6B). The carbon source, rather than strain identity, therefore determined the simulated metabolic operating point.

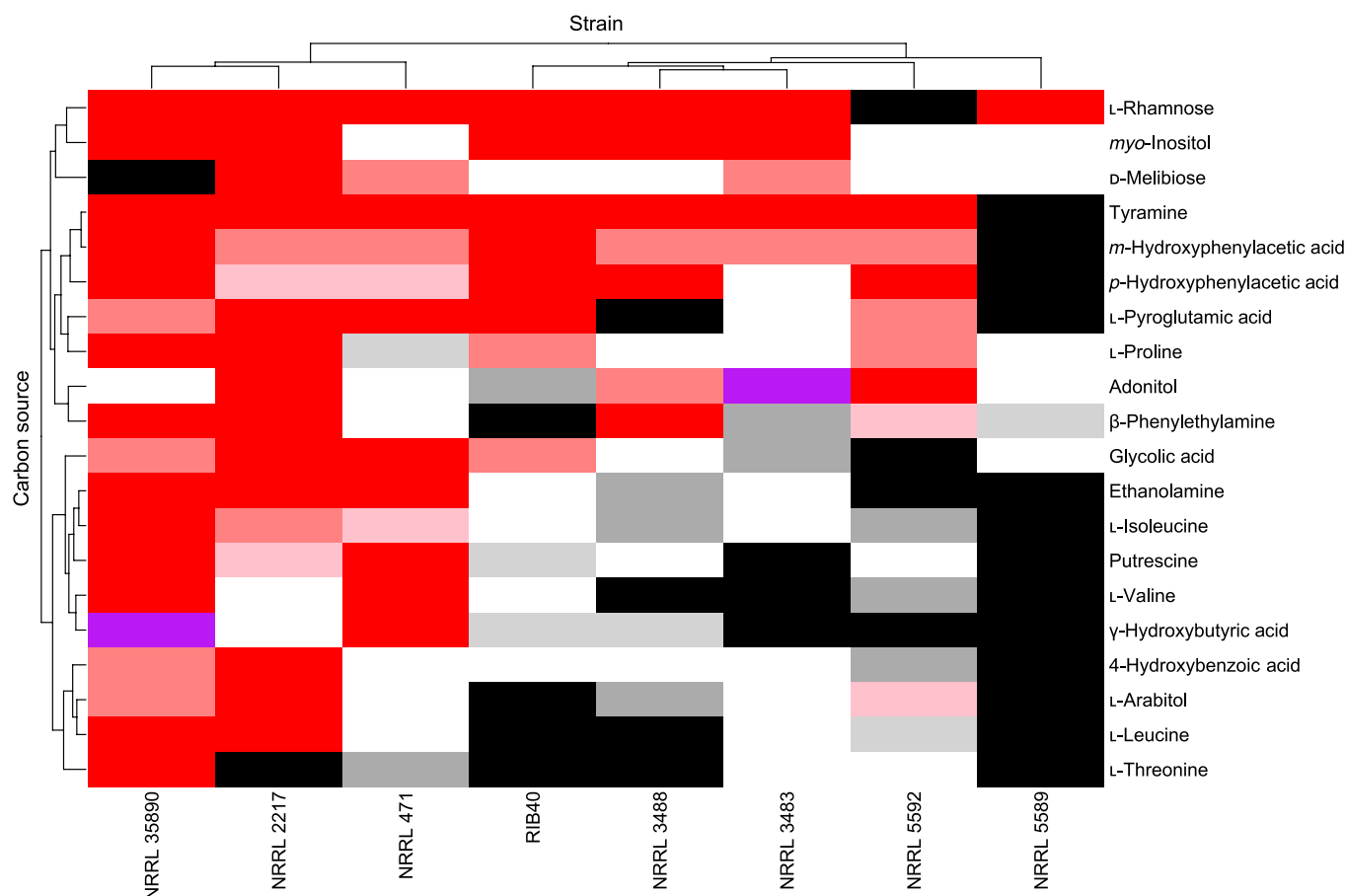


Fig. 3. Heatmap for the metabolic activity measurements for 8 *Aspergillus oryzae* strains on the 20 carbon sources of the phenotype microarray experiment that display the largest intraspecies variability. Clustering was performed both on the strains and the substrates. Experiments were performed with 3 biological replicates, which have been collapsed into a single consensus call per strain–substrate combination. Coloring reflects both the consensus activity status and the agreement among replicates: Red indicates activity across all replicates; black indicates no activity across all replicates; light red indicates probable activity (2 active replicates); dark gray indicates probable no activity (one ambiguous or one active replicate among otherwise inactive replicates); pink or light gray indicates a possible but nonconclusive result; white indicates an ambiguous outcome; and purple flags contradictory replicate calls.

Systematic single-reaction deletion identified 169 reactions whose removal abolished growth, all belonging to the conserved core; in contrast, none of the 634 accessory reactions, deleted individually in each strain that encodes it, reduced the predicted growth rate by more than 0.06% (Fig. 6C). Loopless flux variability analysis (FVA) showed that, of the 827 reactions able to carry flux, 650 (79%) had an attainable flux range that was identical across strains and the comparatively few reactions that did differ were predominantly core reactions whose feasible direction or magnitude shifted. These variable range reactions concentrated in the functionally redundant fatty acid synthase cycle, peripheral sugar catabolism, and a small number of intercompartmental transporters, rather than accessory genes (Fig. S11). Because these reactions are not used at the parsimonious optimum, the interstrain variability resided in unused network capacity and did not affect the growth-supporting flux.

Finally, the substrate-specific rewiring of metabolism was itself consistent across strains. The change in flux distributed to each metabolic subsystem when switching from glucose to glycerol, maltose, or xylose was highly reproducible between the 8 strains (maximum interstrain standard deviation ~3 percentage points; Fig. 6D), confirming that the conserved

core, and not strain-specific gene content, governed the predicted response to each carbon source. Together, these analyses show that the strain-specific models do not differ appreciably in the solution space probed by the tested growth conditions, which both explains the limited interstrain resolution of FBA and localizes the experimentally observed phenotypic diversity to regulatory rather than network-structural differences.

Discussion

In this study, we deliver 3 key resources to advance the metabolic modeling of the industrially important mold *A. oryzae*: first, a comprehensive pan-GEM (*pAo*), representing the full metabolic potential of 187 *A. oryzae* strains; second, a corresponding collection of 187 strain-specific GEMs derived from this pan-GEM; and, third, a set of 8 highly curated models for phylogenetically diverse strains, validated against high-throughput PM data. The *pAo* model serves as a state-of-the-art catalog that captures a considerably higher number of reactions and metabolites than the previous *oryzae2008* model. Its primary use lies in its function as a template, which streamlines the generation of strain-specific models by enabling automated

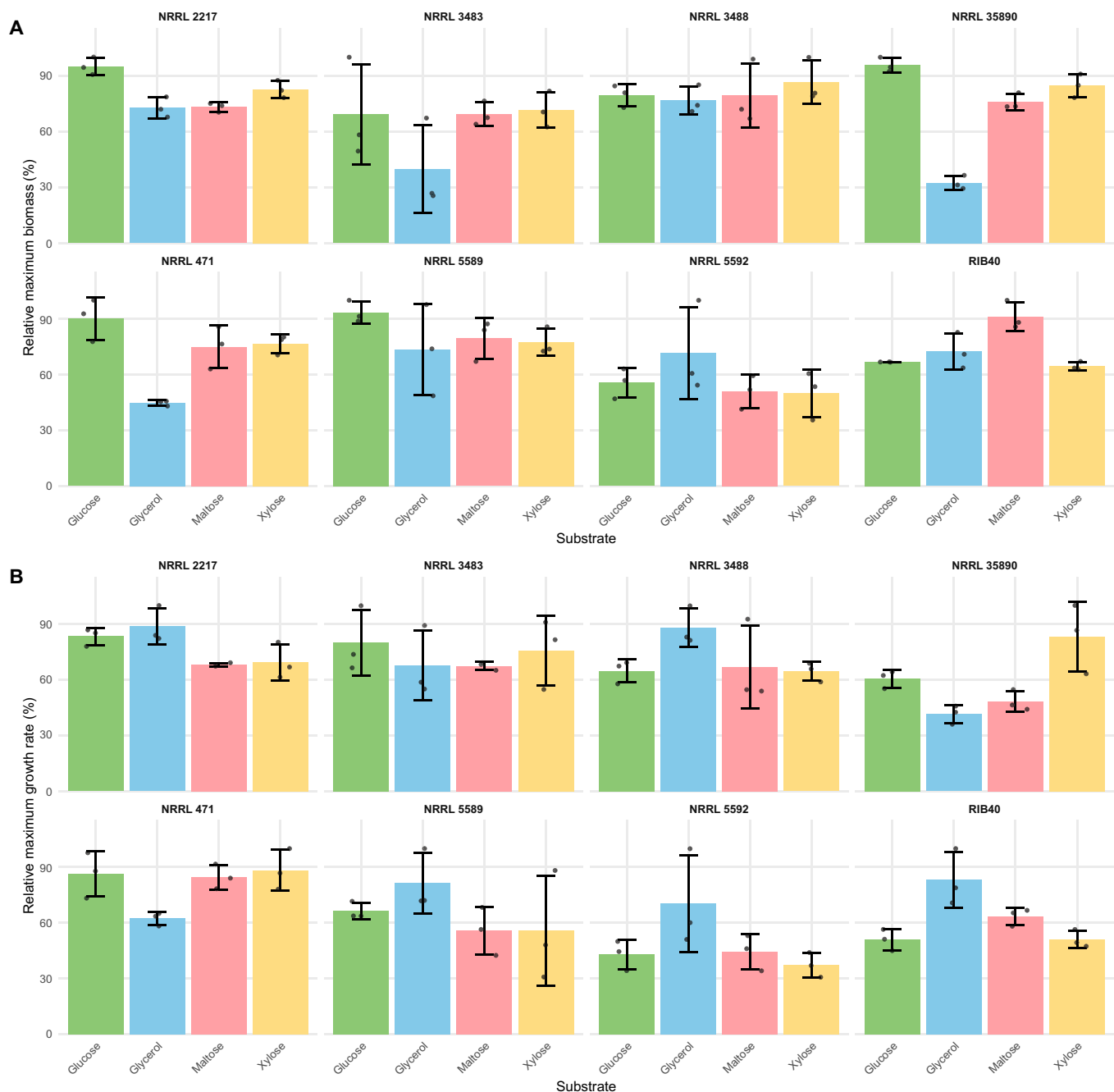


Fig. 4. Comparison of growth between different substrates for 8 *Aspergillus oryzae* strains. (A) Biomass measurements. (B) Measurements of the maximum specific growth rate. Results are means $\pm$ SD of 3 biological replicates, normalized to the maximum value obtained for each strain.

gap-filling and allowing for the rapid annotation of new strains via straightforward homology searches. Furthermore, the pan-GEM can be used to assess the overall metabolic capacity of *A. oryzae* for a specific biotechnological application and subsequently for the identification of the most suitable strain. We envision this resource platform will facilitate the consistent metabolic annotation of novel *A. oryzae* strains and enable iterative updates as new genomic data becomes available.

To validate our approach, we selected 8 strain-specific GEMs from diverse phylogenetic clades for experimental curation. These strains were chosen to span the major phylogenetic diversity within *A. oryzae*, which we show is reflected in substantial differences in metabolic network topology, enabling

representative validation across evolutionary lineages while maintaining experimental feasibility given the inherently lower throughput of filamentous fungal systems compared to bacterial or yeast models. High-throughput phenotyping on 285 substrates identified 45 that unambiguously supported metabolic activity across most of the 8 selected strains. We curated the models to ensure that they predicted growth on these substrates. A conscious decision was made not to curate models based on a lack of metabolic activity, as this phenotype cannot be unambiguously translated into reaction-level constraints. While removal of specific individual reactions may mimic the observed nongrowth phenotype *in silico*, it would not be possible to reliably identify true absent reactions due to baseline

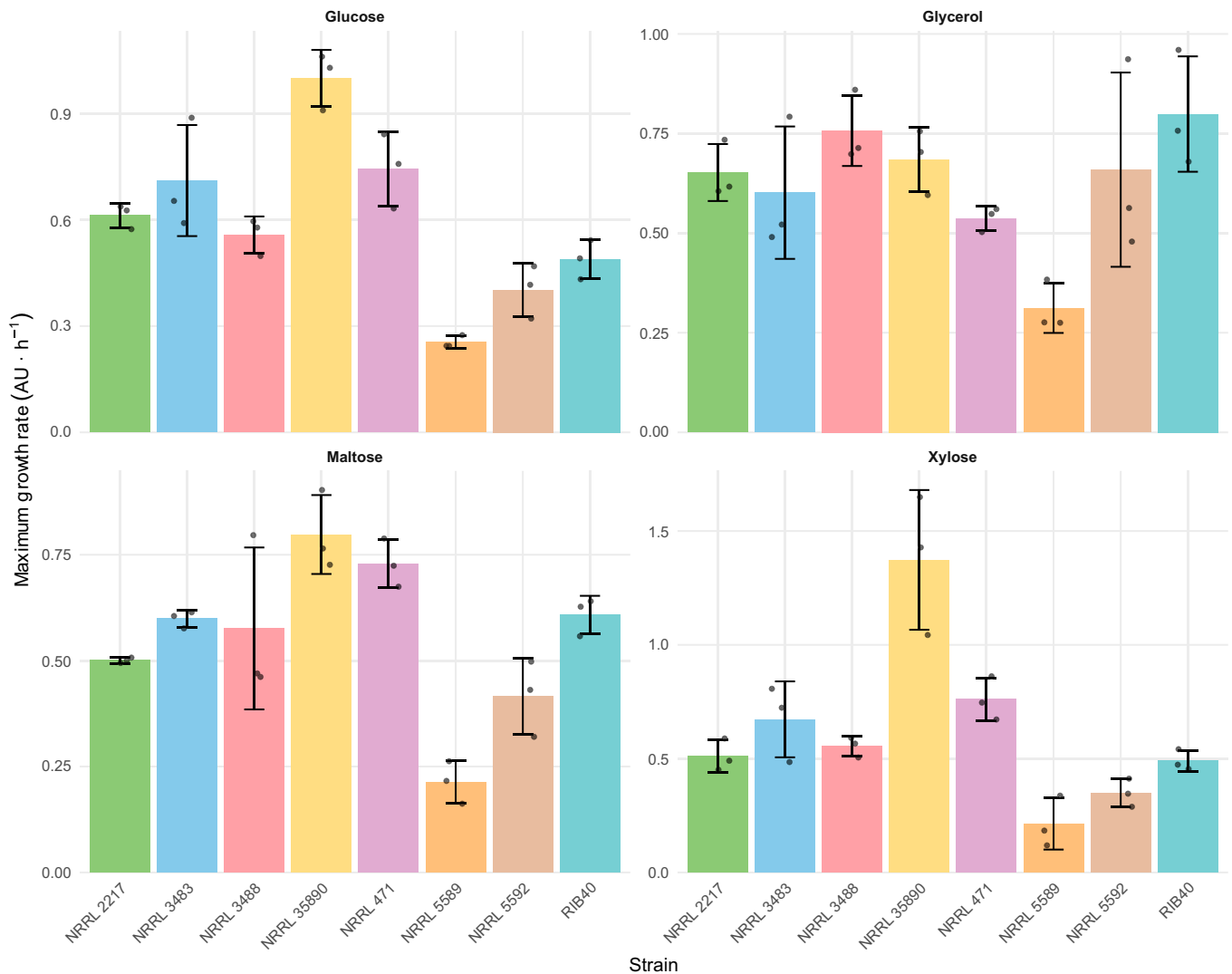


Fig. 5. Bar chart for comparing the growth rates of 8 *Aspergillus oryzae* strains on different carbon sources. Measurements were obtained for 3 biological replicates on glucose, maltose, glycerol, and xylose. Results are means $\pm$ SD of 3 biological replicates, expressed in arbitrary units (AU).

signal that impedes nongrowth identification, regulatory effects that may be alternative sources of nongrowth, and metabolic redundancy. We also encountered a challenge with substrates showing ambiguous or strain-dependent activity. For these, defining a universal cutoff to separate true metabolic capability from noise was problematic; future work could use a sliding window analysis across multiple activity thresholds to resolve these cases, which would further enhance model heterogeneity and predictive accuracy.

A key finding from this validation highlighted a fundamental challenge in metabolic modeling: While our experiments revealed strain-specific differences in biomass production on 4 carbon sources, pFBA and FVA failed to predict this variation. The model simulations rather showed that 8 strain-specific models occupy essentially the same flux solution space under the tested conditions, with strain-to-strain differences in predicted yield and energetic cost below 0.3% and with no accessory reaction individually affecting the growth optimum. The reactions that distinguish the strains lie predominantly in peripheral, functionally redundant capacity that is not carried at the optimum. The conserved metabolic core dictates the

simulated phenotype, and the substantial growth differences measured experimentally must originate at the regulatory or kinetic level rather than in metabolic network structure. This reframes the limited interstrain resolution of FBA not as a shortcoming of the reconstructions but as a structural property of the species' metabolism, and it reinforces the case for enzyme- or proteome-constrained extensions [67] to capture the regulatory layer. This limitation is further compounded by the absence of systematically curated gene essentiality datasets for *A. oryzae*. Existing knockout studies are fragmented and inconsistently annotated, limiting their direct integration into large-scale model validation workflows. While genome-wide deletion collections are available for *S. cerevisiae* [68] and *Neurospora crassa* [69,70] and have been used for curation of GEMs [52,71,72], comparable resources are not yet available for *A. oryzae*. Their development would provide a valuable foundation for future constraint-based model validation and refinement.

When comparing our work to previous studies, methodological differences are crucial to consider. For instance, discrepancies between our growth data for the RIB40 strain and

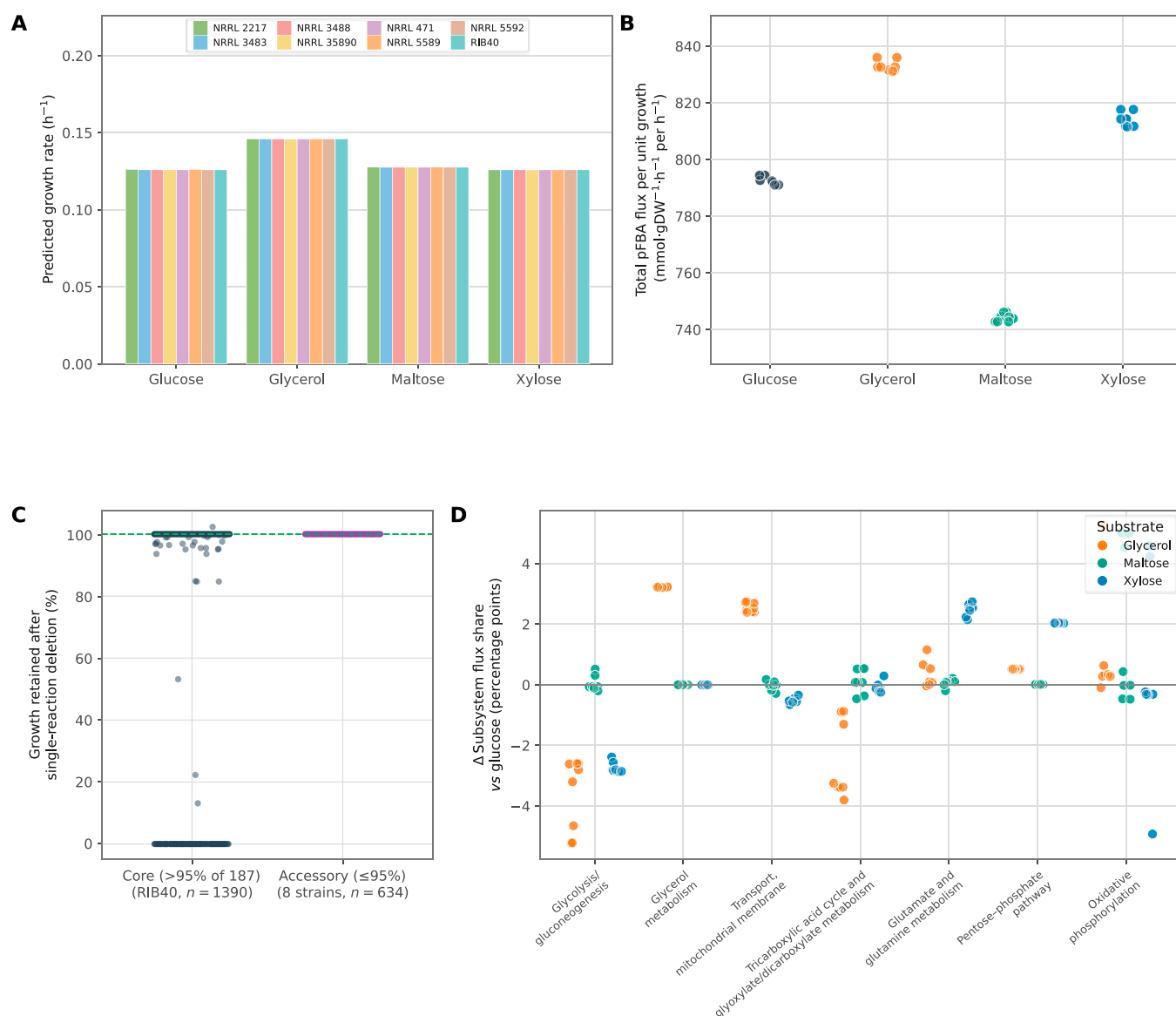


Fig. 6. Comparative flux analysis of 8 strain-specific *Aspergillus oryzae* genome-scale metabolic models (GEMs) across 4 carbon sources. (A) Flux balance analysis (FBA)-predicted maximum growth rate per strain, with carbon uptake normalized to an equal carbon mole supply. (B) Parsimonious FBA (pFBA) costs, as total flux per unit growth, per strain. (C) Growth retained after single-reaction deletion of core and accessory reactions. The core simulations were done in the RIB40 GEM, and the accessory reactions were done in the relevant strain-specific GEM. (D) Change in subsystem flux share relative to glucose for the 7 most-affected subsystems; each point is one strain, colored by substrate.

the study by Vongsangnak *et al.* (which presented the *oryzae2008* model) [49] likely stem from differing experimental setups. Our biomass production and growth rates were determined in 24-well plates, based on arbitrary units, whereas their data are compiled from various experiments on strain A1560 rather than RIB40: growth rates from batch fermentations in 1.2 l of bioreactors and biomass yields from 1.5 l of chemostats [65]. Furthermore, their reported biomass yield on xylose was inferred from a study on *A. nidulans* and *A. niger* [66], not *A. oryzae*.

Furthermore, a single biomass equation was used for all models generated in this study, adopted directly from *oryzae2008* [49], which, in turn, was based on experimental data from various studies [73–76]. The use of a single biomass composition across strains is common practice in fungal genome-scale metabolic modeling [45,49,77,78], although biomass compositions are known to (partially) vary with growth conditions and

genetic background [74,78]. While strain-specific biomass composition measured under a defined growth condition could refine model accuracy, characterizing every strain represented by a pan-GEM such as *pAo* would require substantial additional experimental work, even if many of the resulting compositions are anticipated to be comparable across phylogenetically related strains. This limitation may particularly be relevant for biomass yield predictions [78], where FBA was not able to fully capture the experimentally observed strain-specific variations in this study. However, this discrepancy likely arises not only from the use of a fixed biomass composition but also from the more general limitations of using conventional GEMs, which do not account for regulatory effects or kinetic constraints.

Concurrently with our study, another GEM, *iNR1684*, was developed for *A. oryzae* AoCordy-T1, a derivative of the strain

BCC7051 engineered for cordycepin production [50]. This model was based on homology to a model for *C. militaris* (*iNR1329*), incorporated *oryzae2008*, and was gap-filled using GEMs from various other fungi. This approach resulted in a model highly specialized for a specific strain and biosynthetic pathway. In contrast, our pan-GEM provides a comprehensive, species-wide foundation that enables the consistent generation and comparison of models across *A. oryzae* strains, making it a broadly applicable resource for the community. Notably, compared to the *iNR1684* model, *pAo* contains 274 additional biochemical reactions and 153 more metabolites, highlighting the advantage of a pan-genome-based model in capturing the metabolic capabilities of the species (Table S4).

In conclusion, we present a collection of curated GEMs for *A. oryzae* based on a pan-GEM, substantially expanding on previous models, but with limited representation of phenotypic intraspecies variability. A promising strategy to overcome this limitation is incorporating additional layers of information into the FBA simulation workflow. We anticipate that in the future, our pan-GEM will be extended to include more *A. oryzae* strains and enhanced with such data, following the same trend observed for model organisms such as *S. cerevisiae*. Specific implementations could include imposing enzyme constraints [79], developing a proteome-constrained genome-scale protein secretory model [80], or expanding the compartmental resolution of *pAo* beyond the current 4-compartment structure. Such enhancements would directly address the current limitations in predicting phenotypic expression and would substantially expand the utility of our pan-GEM for the research community, given *A. oryzae*'s widespread use as a host for recombinant protein production.

Methods

Strain selection

The comprehensive genomic dataset used in this study is described in detail in van der Luijt *et al.* (in preparation). Briefly, we combined genome assemblies from newly sequenced *A. oryzae* NRRL strains and A1560 with publicly available genomes from National Center for Biotechnology Information (NCBI)'s Genome database [81] and the comprehensive *A. oryzae* genome database across the clades [82]. In addition, we included *de novo* assemblies from raw sequencing data obtained from NCBI's Sequence Read Archive [83]. All genomes were assessed for completeness using BUSCO (v5.8.2) [84,85], and gene prediction was performed with AUGUSTUS (v3.4.0) [86]. To ensure high-quality data, we retained only genomes with a BUSCO completeness score of >94%. Clade assignments are also described in detail in van der Luijt *et al.* (in preparation).

Constructing the *A. oryzae* pan-GEM

GPRs were established to link each protein sequence to the metabolic reactions it catalyzes. This is accomplished by associating the novel protein sequences with those of known GPRs through orthology. Specifically, we use protein sequences from 3 well-curated GEMs: the *oryzae2008* GEM of *A. oryzae* RIB40 [49] and 2 GEMs from the closely related species *A. niger* [58] and *Aspergillus fumigatus* [46].

To group homologous sequences into clusters of orthologous genes, ensuring that duplicated or highly similar genes are represented as a single gene rather than counted multiple times, protein sequence clustering is performed using BPGA [87], which uses the

clust_fast function from the USEARCH algorithm [88]. Sequences were grouped into clusters of orthologous genes using an identity threshold of 0.5. Importantly, this clustering step does not itself determine gene function; GPRs are subsequently assigned to novel sequences through guilt by association with the established GEMs.

This threshold of 0.5 was chosen because it is (a) permissive enough to link a larger fraction of novel sequences to those in the 3 established GEMs, while (b) stringent enough to ensure that most sequences from the established GEMs cluster separately. For instance, of the 1,346 proteins annotated with a GPR in *oryzae2008*, 1,160 sequence clusters were identified using these settings. In total, this approach resulted in the identification of 11,949 protein sequence clusters.

For our pan-GEM, we adopt the same 4 cellular compartments described in the *oryzae2008* GEM: cytoplasm, mitochondria, peroxisome, and extracellular space. In contrast, the *A. niger* and *A. fumigatus* GEMs also assign reactions to additional compartments, including the nucleus, Golgi apparatus, endoplasmic reticulum, vacuole, and lipid particles. To ensure consistency across template models, reactions from these compartments were all assigned to the cytoplasm in our pan-GEM. Compartmentalization matters in GEMs, but the reactions most influential for phenotypic prediction belong to energy metabolism and (branched chain) amino acid metabolism [89], both localized to the retained mitochondrion. The collapsed organelles are dominated by lipids and are already connected by transport reactions representing vesicle shuttling, so collapsing these few compartments should not detrimentally affect predictions. In addition, we note that the 3 established GEMs use different nomenclatures for metabolites. To address this, we consolidated the original metabolite names using the KEGG [60], ChEBI [90], and PubChem [91] databases. Metabolite identifiers were standardized to KEGG identifiers when available, ChEBI identifiers if no KEGG identifier could be obtained, and metabolite names when neither identifier was found.

For sequences without orthologs in any of the 3 established GEMs, we perform a BLAST [59] homology search to the KEGG [60] reaction database using the *getKEGGModelForOrganism* function from the RAVEN toolbox [92]. This function, designed to reconstruct a GEM based on protein homology to orthologs of all eukaryotic organisms cataloged in KEGG (KEGG release 105), allows us to retrieve the GPRs for sequences lacking orthologs. Subsequently, we remove any additional reactions that were included by *getKEGGModelForOrganism* to report a standalone GEM. The cellular localization of the remaining reactions is then predicted using DeepLoc [93]. Finally, since the KEGG reaction database typically lacks accurate information on reaction reversibility, we update the reversibility of reactions using the online equilibrators tool [94] (<https://equilibrator.weizmann.ac.il>). Sequences for which no GPRs could be established using any of these methods were excluded from the pan-GEM.

The biomass equation was adopted directly from *oryzae2008* (Table S5) [49]. The absence of erroneous energy-generating cycles was evaluated by testing whether dissipation reactions of the following energy carriers were zero: ATP, cytidine triphosphate, guanosine triphosphate, uridine triphosphate, inosine triphosphate, reduced nicotinamide adenine dinucleotide (NADH), reduced nicotinamide adenine dinucleotide phosphate (NADPH), reduced flavin adenine dinucleotide (FADH₂), reduced flavin mononucleotide (FMNH₂), acetyl-coenzyme A, L-glutamate, and ubiquinol-8, in accordance to the approach described by Fritzemeier *et al* [95]. Three such cycles initially present in the pan-GEM, arising from reactions adopted from

other *Aspergillus* species, were eliminated by constraining 4 reactions (OtherAsp_R04962, OtherAsp_R01708, r1736, and r1737) to their thermodynamically feasible direction.

The model quality was assessed by *memote* [96]. The consensus *pAo* model scored 57.7% overall, with all reactions charge-balanced and ~80% mass-balanced (limited by ~430 reactions with generic/polymeric pseudometabolites) that cannot carry a unique formula. Full reports are available at <https://doi.org/10.5281/zenodo.18598684>.

Model similarity analysis for strain-specific GEMs

We used the *compareMultipleModels* function from RAVEN toolbox [92] to compare the strain-specific GEMs at the reaction level. Briefly, a binary reaction existence matrix was collected based on which pairwise Hamming distances between strains are calculated. These distances were used as input for *t*-distributed stochastic neighbor embedding (t-SNE) to visualize the strain-specific GEMs in reduced dimension space.

Strains and media

Eight *A. oryzae* strains (RIB40, NRRL 471, NRRL 2217, NRRL 3483, NRRL 3488, NRRL 5589, NRRL 5592 and NRRL 35890) were used for phenotypic characterization in 96- and 24-well plates.

Potato dextrose agar (glucose [20 g/l], potato infusion [4 g/l], and agar [15 g/l]) plates and slants were used to initially revive strains from conidial suspension stored as cryostocks in 20% glycerol at -70°C and harvest conidia.

For PM assays, filamentous fungi inoculating fluid (Biolog Inc., Hayward, CA, USA) was used as base medium, supplemented differently depending on the PM plate type: no supplements for PM1 and PM2A plates and 100 mM D-glucose, 5 mM potassium phosphate monobasic anhydrous, and 2 mM sodium sulfate for PM3B plates.

For 24-well plate growth studies, a Czapek–Dox medium (NaNO_3 [3 g/l], KCl [2 g/l], KH_2PO_4 [1 g/l], $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ [0.5 g/l], and FeSO_4 [0.02 g/l] [pH 5.5]) was prepared. This medium was supplemented with 20 g/l of one of 4 carbon sources (glucose, maltose, glycerol, or xylose), 0.01% (v/v) Tween 80 as a surfactant, and chloramphenicol [25 $\mu\text{g/ml}$] to prevent bacterial contamination.

Phenotypic microarrays

PM experiments were conducted using PM1, PM2A, and PM3B plates (Biolog Inc.). Each well was inoculated with 100 μl of a fresh conidial suspension prepared in filamentous fungi inoculating fluid with the appropriate supplements, dispensed using an Opentrons OT-2 automated pipette (Opentrons Labworks Inc., Long Island City, NY, USA). Redox Dye D (Biolog Inc.), a tetrazolium-based dye that indicates metabolic activity via NAD(P)H-dependent reduction, was added to each well. Plates were incubated at 30°C in an OmniLog multiple plate reader (Biolog Inc.) for 96 h, with metabolic activity monitored kinetically. All experiments were performed in biological triplicates.

Analysis of phenotypic microarray data

Raw data from the PMs, stored as D5E files, was imported in Data Analysis 1.7.1.58 (Biolog Inc.) and subsequently exported as CSV files. The CSV files were imported into R 3.5.1 and processed with the *opm* package (v1.3.77), using the custom R script *opm_pipeline*.

To construct the heatmaps of Fig. 3 and Fig. S6, we used the raw data from the PM experiment, i.e., metabolic activity measurements along 96 h taken at 1-h time intervals for 8 strains in biological triplicate. As a measure for metabolic activity for a specific replicate of a strain, we take the metabolic activity at the last time point and subtract the average of the 10 media with the lowest metabolic activity on the same plate at the last time point. While a negative control sample is included on each plate for the purpose of normalization, we found that, in a few cases, these were noisy, and, therefore, we instead rely on the assumption that each plate contains at least 10 media for which truly no metabolic activity is observed. Finally, the heatmaps were constructed in R using the *complexHeatmap* package (v2.24.1) [97].

Growth characterization

For growth characterization, 2.5 ml of modified Czapek–Dox medium (prepared as described above) was dispensed into each well of a 24-well plate (CR1424dg, EnzyScreen B.V., Heemstede, The Netherlands). Fresh conidial suspensions were used to inoculate the wells, and plates were incubated at 30°C with shaking at 200 rpm in a Growth Profiler 960 (EnzyScreen B.V.) for 72 h. Growth was monitored over time, and each condition was tested in biological triplicates.

Analysis of 24-well plate cultivation data

Images captured by the Growth Profiler 960 were analyzed using GP960Viewer software (EnzyScreen B.V.). The 24Wells_Square profile was applied with default multiwell plate dimensions, alongside the following custom parameters: spot size in X and Y directions (14 mm each), geometric correction (0.5 mm), and exclusion of the top 5% and bottom 75% of pixel intensity values to minimize medium and edge artifacts.

For each well, green intensity values were exported to a CSV file. Time points were converted to hours, and raw intensity values were blank corrected by subtracting the average of the 5 minimum values observed in the respective well. The blanked data were then analyzed using a custom R pipeline to fit splines using function *smooth.spline*, with smoothing set to $\text{spar} = 0.7$, to estimate (pseudo) growth rates from time-dependent changes in green value, taking the steepest section as maximum growth rate (Figs. S12 to S15).

Simulating metabolism using FBA

To validate the 8 strain-specific GEMs, we used metabolic activity measurements on various carbon and nitrogen sources from the PM experiment. Each model was curated using strain-specific FBA [98], applied to metabolites where the PM data indicated strong metabolic activity. To run FBA on multiple models simultaneously in an efficient manner, we used the Python package *Medusa* [99]. Media were selected for curation based on the following criteria. First, we estimated background metabolic activity for each strain replicate by averaging the ten metabolites with the lowest activity on the corresponding plate at the final time point. A metabolite was included for downstream analysis if it met one of 2 conditions: (a) At least 3 strains had at least 2 replicates with an activity of ≥ 150 units above background, or (b) at least 2 strains had at least 2 replicates with an activity of ≥ 150 units above background, and at least 4 additional strains had at least 2 replicates with an activity of ≥ 100 units above background. With these criteria, 45 metabolites were retained for running FBA, and 240 were excluded. If FBA predicted no

growth for a strain-specific GEM for a metabolite that clearly supported metabolic activity in the PM experiment, we curate the GEM by (a) automatically gap-filling reactions from the pan-GEM, (b) adding a transport reaction for the target metabolite if appropriate, or (c) by manual gap-filling. After this round of model validation, strain-specific FBA growth predictions correspond perfectly with the observed metabolic activity measurements of the PM experiment for all 45 selected metabolites.

To compare with experimental growth data, FBA and FVA were performed with each carbon source supplied at an equal carbon flux of $6.72 \text{ mmol C}\cdot\text{gDW}^{-1}\cdot\text{h}^{-1}$, corresponding to the measured specific glucose uptake rate ($q_s = 1.12 \text{ mmol glucose}\cdot\text{gDW}^{-1}\cdot\text{h}^{-1}$; Table S5), with a nongrowth associated ATP maintenance of $1 \text{ mmol}\cdot\text{gDW}^{-1}\cdot\text{h}^{-1}$. FVA was run at 90% of the maximal growth rate under a loopless constraint. Reactions differing most between strains were identified by ranking them by the variance, across the 8 strains of their loopless flux range on glucose.

Acknowledgments

Funding: J.G., C.R.B.v.d.L., M.O.A.S., L.J.J., and E.J.K. were supported by the Novo Nordisk Foundation grant NNF20CC0035580. C.R.B.v.d.L. was additionally supported by the Novo Nordisk Foundation grant NNF21OC0065495, and M.F., C.S.-G.B., and M.O.A.S. were additionally supported by funding from the Bezos Earth Fund Center for Alternative Proteins. We would like to thank M. N. Lund and P. E. Jensen at the Department of Food Science, University of Copenhagen for funding acquisition and guidance; T. Adkins at the ARS Culture Collection (NRRL) and J. B. Hoof at the IBT Culture Collection for providing strains for this study; L. Munro for access to the OpenTrons OT-2 at NNF BRIGHT; and C. H. Bang-Berthelsen for granting access to the BioLog OmniLog at the National Food Institute (DTU).

Author contributions: J.G.: Conceptualization, methodology, software, formal analysis, investigation, writing—original draft, and visualization. C.R.B.v.d.L.: Conceptualization, methodology, software, formal analysis, investigation, writing—original draft, and visualization. M.F.: Investigation and writing—review. C.S.-G.B.: Investigation and writing—review. M.O.A.S.: Funding acquisition and writing—review. L.J.J.: Conceptualization, supervision, writing—review and editing, and funding acquisition. E.J.K.: Conceptualization, methodology, software, formal analysis, supervision, writing—review and editing, visualization, and funding acquisition.

Competing interests: M.O.A.S. is a board member of Novonesis. M.O.A.S. and L.J.J. are cofounders of MATR Foods ApS. The other authors declare that they have no competing interests.

Data Availability

The code for reproducing all analyses and figures and the final GEM for *A. oryzae* are shared on GitHub (<https://github.com/SysBioChalmers/panAsp-GEM>). All code, models, and data used and generated in this project can be retrieved from Zenodo (10.5281/zenodo.18598684).

Supplementary Materials

Figs. S1 to S15
Tables S1 to S5
References [100–120]

References

- Barbesgaard P, Heldt-Hansen HP, Diderichsen B. On the safety of *Aspergillus oryzae*: A review. *Appl Microbiol Biotechnol.* 1992;36(5):569–572.
- Machida M, Yamada O, Gomi K. Genomics of *Aspergillus oryzae*: Learning from the history of koji mold and exploration of its future. *DNA Res.* 2008;15(4):173–183.
- Frisvad JC, Møller LLH, Larsen TO, Kumar R, Arnau J. Safety of the fungal workhorses of industrial biotechnology: Update on the mycotoxin and secondary metabolite potential of *Aspergillus niger*, *Aspergillus oryzae*, and *Trichoderma reesei*. *Appl Microbiol Biotechnol.* 2018;102(22):9481–9515.
- Daba GM, Mostafa FA, Elkhateeb WA. The ancient koji mold (*Aspergillus oryzae*) as a modern biotechnological tool. *Bioresour Bioprocess.* 2021;8(1):52.
- Watarai N, Yamamoto N, Sawada K, Yamada T. Evolution of *Aspergillus oryzae* before and after domestication inferred by large-scale comparative genomic analysis. *DNA Res.* 2019;26(6):465–472.
- Balakrishnan M, Jeevarathinam G, Kumar SKS, Muniraj I, Uthandi S. Optimization and scale-up of α -amylase production by *Aspergillus oryzae* using solid-state fermentation of edible oil cakes. *BMC Biotechnol.* 2021;21(1):33.
- Murthy PS, Kusumoto KI. Acid protease production by *Aspergillus oryzae* on potato pulp powder with emphasis on glycine releasing activity: A benefit to the food industry. *Food Bioprod Process.* 2015;96:180–188.
- Østergaard LH, Olsen HS. Industrial applications of fungal enzymes. In: Hofrichter M, editor. *Industrial applications*. Berlin (Heidelberg): Springer; 2011. p. 269–290.
- Feng J, Hauser M, Cox RJ, Skellam E. Engineering *Aspergillus oryzae* for the heterologous expression of a bacterial modular polyketide synthase. *J Fungi.* 2021;7(12):1085.
- Meng X, Fang Y, Ding M, Zhang Y, Jia K, Li Z, Collemare J, Liu W. Developing fungal heterologous expression platforms to explore and improve the production of natural products from fungal biodiversity. *Biotechnol Adv.* 2022;54: Article 107866.
- Ntana F, Mortensen UH, Sarazin C, Figge R. *Aspergillus*: A powerful protein production platform. *Catalysts.* 2020;10(9):1064.
- Raksha S, Srinivasan S, Prasant G, Prabu R. Over-expression of gluconic acid in *Aspergillus oryzae* RP-21 mutants generated by a random mutagenesis approach. *3 Biotech.* 2012;2(3):219–223.
- Ogawa A, Wakisaka Y, Tanaka T, Sakiyama T, Nakanishi K. Production of kojic acid by membrane-surface liquid culture of *Aspergillus oryzae* NRRL484. *J Ferment Bioeng.* 1995;80(1):41–45.
- Brown SH, Bashkirova L, Berka R, Chandler T, Doty T, McCall K, McCulloch M, McFarland S, Thompson S, Yaver D, et al. Metabolic engineering of *Aspergillus oryzae* NRRL 3488 for increased production of l-malic acid. *Appl Microbiol Biotechnol.* 2013;97(20):8903–8912.
- Yang L, Lübeck M, Lübeck PS. *Aspergillus* as a versatile cell factory for organic acid production. *Fungal Biol Rev.* 2017;31(1):33–49.
- Dai Z, Zhou H, Zhang S, Gu H, Yang Q, Zhang W, Dong W, Ma J, Fang Y, Jiang M, et al. Current advance in biological production of malic acid using wild type and metabolic engineered strains. *Bioresour Technol.* 2018;258:345–353.

17. Wiebe MG. Myco-protein from *Fusarium venenatum*: A well-established product for human consumption. *Appl Microbiol Biotechnol*. 2002;58(4):421–427.
18. Maini Rekdal V, Van Der Luijt CRB, Chen Y, Kakumanu R, Baidoo EE, Petzold CJ, Cruz-Morales P, Keasling JD. Edible mycelium bioengineered for enhanced nutritional value and sensory appeal using a modular synthetic biology toolkit. *Nat Commun*. 2024;15(1):2099.
19. Devanthi PVP, Pratama F, Pramanda IT, Bani MD, Kadar AD, Kho K. Exploring the potential of *Aspergillus oryzae* for sustainable mycoprotein production using okara and soy whey as cost-effective substrates. *J Fungi*. 2024;10(8):555.
20. Uwineza C, Mahboubi A, Atmowidjojo A, Ramadhani A, Wainaina S, Millati R, Wikandari R, Niklasson C, Taherzadeh MJ. Cultivation of edible filamentous fungus *Aspergillus oryzae* on volatile fatty acids derived from anaerobic digestion of food waste and cow manure. *Bioresour Technol*. 2021;337:Article 125410.
21. Hashempour-Baltork F, Khosravi-Darani K, Hosseini H, Farshi P, Reihani SFS. Mycoproteins as safe meat substitutes. *J Clean Prod*. 2020;253:Article 119958.
22. Souza Filho PF, Nair RB, Andersson D, Lennartsson PR, Taherzadeh MJ. Vegan-mycoprotein concentrate from pea-processing industry byproduct using edible filamentous fungi. *Fungal Biol Biotechnol*. 2018;5(1):5.
23. Lee K, Lee SK, Lee BD. *Aspergillus oryzae* as probiotic in poultry - A review. *Int J Poult Sci*. 2006;5(1):1–3.
24. Yohe TT, O'Diam KM, Daniels KM. Growth, ruminal measurements, and health characteristics of Holstein bull calves fed an *Aspergillus oryzae* fermentation extract. *J Dairy Sci*. 2015;98(9):6163–6175.
25. Hamajima H, Matsunaga H, Fujikawa A, Sato T, Mitsutake S, Yanagita T, Nagao K, Nakayama J, Kitagaki H. Japanese traditional dietary fungus koji *Aspergillus oryzae* functions as a prebiotic for *Blautia coccoides* through glycosylceramide: Japanese dietary fungus koji is a new prebiotic. *Springerplus*. 2016;5(1):1321.
26. Iwashita MKP, Nakandakare IB, Terhune JS, Wood T, Ranzani-Paiva MJT. Dietary supplementation with *Bacillus subtilis*, *Saccharomyces cerevisiae* and *Aspergillus oryzae* enhance immunity and disease resistance against *Aeromonas hydrophila* and *Streptococcus iniae* infection in juvenile tilapia *Oreochromis niloticus*. *Fish Shellfish Immunol*. 2015;43(1):60–66.
27. Konishi H, Isozaki S, Kashima S, Moriichi K, Ichikawa S, Yamamoto K, Yamamura C, Ando K, Ueno N, Akutsu H, et al. Probiotic *Aspergillus oryzae* produces anti-tumor mediator and exerts anti-tumor effects in pancreatic cancer through the p38 MAPK signaling pathway. *Sci Rep*. 2021;11(1):11070.
28. Arrigan D, Kothe CI, Oliverio A, Evans JD, Wolfe BE. Novel fermentations integrate traditional practice and rational design of fermented-food microbiomes. *Curr Biol*. 2024;34(21):R1094–R1108.
29. Rodríguez Valerón N, Mak T, Jahn LJ, Arboleya JC, Sørensen PM. Derivation of kokumi γ -glutamyl peptides and volatile aroma compounds from fermented cereal processing by-products for reducing bitterness of plant-based ingredients. *Foods*. 2023;12(23):4297.
30. Somerville V, Grigaitis P, Battjes J, Moro F, Teusink B. Use and limitations of genome-scale metabolic models in food microbiology. *Curr Opin Food Sci*. 2022;43:225–231.
31. Scott WT, Rockx S, Mariën Q, Regueira A, Candry P, Ganigué R, Koehorst JJ, Schaap PJ. Implementation of a *Clostridium luticellarii* genome-scale model for upgrading syngas fermentations. *Comput Struct Biotechnol J*. 2025;27:649–660.
32. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010;28(3):245–248.
33. Oyetunde T, Liu D, Martin HG, Tang YJ. Machine learning framework for assessment of microbial factory performance. *PLoS One*. 2019;14(1):Article e0210558.
34. Banerjee D, Eng T, Lau AK, Sasaki Y, Wang B, Chen Y, Prah JP, Singan VR, Herbert RA, Liu Y, et al. Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale. *Nat Commun*. 2020;11(1):5385.
35. Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem*. 1999;274(25):17410–17416.
36. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol*. 2017;35(10):904–908.
37. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*. 2003;13(2):244–253.
38. Zhang C, Sánchez BJ, Li F, Eiden CW, Scott WT, Liebal UW, Blank LM, Mengers HG, Anton M, Rangel AT, et al. Yeast9: A consensus genome-scale metabolic model for *S. cerevisiae* curated by the community. *Mol Syst Biol*. 2024;20(10):1134–1150.
39. Yilmaz LS, Walhout AJM. A *Caenorhabditis elegans* genome-scale metabolic network model. *Cell Syst*. 2016;2(5):297–311.
40. Witting M, Hastings J, Rodriguez N, Joshi CJ, Hattwell JP, Ebert PR, Van Weeghel M, Gao AW, Wakelam MJ, Houtkooper RH, et al. Modeling meets metabolomics—The WormJam consensus model as basis for metabolic studies in the model organism *Caenorhabditis elegans*. *Front Mol Biosci*. 2018;5:96.
41. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA*. 2007;104(6):1777–1782.
42. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, Gatto F, Nilsson A, Preciat Gonzalez GA, Aurich MK, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*. 2018;36(3):272–281.
43. Mardinoglu A, Palsson BØ. Genome-scale models in human metabologenomics. *Nat Rev Genet*. 2025;26(2):123–140.
44. Viana R, Couceiro D, Newton W, Coutinho L, Dias O, Coelho C, Teixeira MC. Unveiling new features of the human pathogen *Cryptococcus neoformans* through the reconstruction and exploitation of a dedicated genome-scale metabolic model. *Comput Struct Biotechnol J*. 2025;27:2336–2346.
45. Dziurzyński M, Nowak ME, Furman M, Okrańska A, Pawłowska J, Fondi M. Insights into optimization of oleaginous fungi – Genome-scale metabolic reconstruction and analysis of *Umbelopsis* sp. WA50703. *Comput Struct Biotechnol J*. 2025;27:1431–1439.
46. Mirhakkak MH, Chen X, Ni Y, Heinekamp T, Sae-Ong T, Xu LL, Kurzai O, Barber AE, Brakhage AA, Boutin S, et al.

- Genome-scale metabolic modeling of *Aspergillus fumigatus* strains reveals growth dependencies on the lung microbiome. *Nat Commun.* 2023;14(1):4369.
47. Kaushal M, Upton DJ, Gupta JK, Wood AJ, Srivastava S. Reconstruction of a genome-scale metabolic model and in-silico flux analysis of *Aspergillus tubingensis*: A non-mycotoxinogenic citric acid-producing fungus. *Biotechnol Biofuels Bioprod.* 2024;17(1):70.
48. Han Y, Rangel AT, Pomraning KR, Kerkhoven EJ, Kim J. Advances in genome-scale metabolic models of industrially important fungi. *Curr Opin Biotechnol.* 2023;84:Article 103005.
49. Vongsangnak W, Olsen P, Hansen K, Krogsgaard S, Nielsen J. Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. *BMC Genomics.* 2008;9(1):245.
50. Raethong N, Jeennor S, Anantayanon J, Wannawilai S, Vongsangnak W, Laoteng K. Reconstruction of a genome-scale metabolic model for *Aspergillus oryzae* engineered strain: A potent computational tool for enhancing cordycepin production. *Int J Mol Sci.* 2025;26(14):6906.
51. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BØ. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci USA.* 2013;110(50):20338–20343.
52. Lu H, Li F, Sánchez BJ, Zhu Z, Li G, Domenzain I, Marčišauskas S, Anton PM, 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):3586.
53. Neal M, Brakewood W, Betenbaugh M, Zengler K. Pan-genome-scale metabolic modeling of *Bacillus subtilis* reveals functionally distinct groups. *mSystems.* 2024;9(11):e00923–e00924.
54. Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol.* 2012;6(1):153.
55. Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol.* 2012;8(5):Article e1002518.
56. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol.* 2008;4(5):Article e1000082.
57. Gustafsson J, Anton M, Roshanzamir F, Jörnsten R, Kerkhoven EJ, Robinson JL, Nielsen J. Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data. *Proc Natl Acad Sci USA.* 2023;120(6):Article e2217868120.
58. Brandl J, Aguilar-Pontes MV, Schäpe P, Noerregaard A, Arvas M, Ram AF, Meyer V, Tsang A, de Vries RP, Andersen MR. A community-driven reconstruction of the *Aspergillus niger* metabolic network. *Fungal Biol Biotechnol.* 2018;5(1):16.
59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–410.
60. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
61. Lehmbeck J, Andersen B, Sáez-Sáez J, Frisvad JC, Arnau J. Mycotoxin-free *Aspergillus oryzae* strain lineage for alternative and novel protein production at industrial scale. *Appl Microbiol Biotechnol.* 2025;109(1):94.
62. Kiyota T, Hamada R, Sakamoto K, Iwashita K, Yamada O, Mikami S. Aflatoxin non-productivity of *Aspergillus oryzae* caused by loss of function in the aflJ gene product. *J Biosci Bioeng.* 2011;111(5):512–517.
63. Črešnar B, Petrič Š. Cytochrome P450 enzymes in the fungal kingdom. *Biochim Biophys Acta.* 2011;1814(1):29–35.
64. Nazir KHMNH, Ichinose H, Wariishi H. Construction and application of a functional library of cytochrome P450 monooxygenases from the filamentous fungus *Aspergillus oryzae*. *Appl Environ Microbiol.* 2011;77(9):3147–3150.
65. Carlsen M, Nielsen J. Influence of carbon source on α -amylase production by *Aspergillus oryzae*. *Appl Microbiol Biotechnol.* 2001;57(3):346–349.
66. Prathumpai W, Gabelgaard JB, Wanchanthuek P, Van de Vondervoort PJ, De Groot MJ, McIntyre M, Nielsen J. Metabolic control analysis of xylose catabolism in *Aspergillus*. *Biotechnol Prog.* 2003;19(4):1136–1141.
67. Borodina I, Nielsen J. From genomes to *in silico* cells via metabolic networks. *Curr Opin Biotechnol.* 2005;16(3):350–355.
68. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature.* 2002;418(6896):387–391.
69. Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC. A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc Natl Acad Sci USA.* 2006;103(27):10352–10357.
70. Dunlap JC, Borkovich KA, Henn MR, Turner GE, Sachs MS, Glass NL, McCluskey K, Plamann M, Galagan JE, Birren BW, et al. Enabling a community to dissect an organism: Overview of the *Neurospora* functional genomics project. *Adv Genet.* 2007;57:49–96.
71. Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol.* 2008;26(10):1155–1160.
72. Dreyfuss JM, Zucker JD, Hood HM, Ocasio LR, Sachs MS, Galagan JE. Reconstruction and validation of a genome-scale metabolic model for the filamentous fungus *Neurospora crassa* using FARM. *PLoS Comput Biol.* 2013;9(7):Article e1003126.
73. Pedersen H. Metabolic flux in *Aspergillus oryzae* [master thesis]. Lundtofte (Denmark): Technical University of Denmark; 1996.
74. Pedersen H, Carlsen M, Nielsen J. Identification of enzymes and quantification of metabolic fluxes in the wild type and in a recombinant *Aspergillus oryzae* strain. *Appl Environ Microbiol.* 1999;65(1):11–19.
75. Debois M, Heydorn A, Obel N. Macromolecule analysis of *Aspergillus oryzae* [master thesis]. Lundtofte (Denmark): Technical University of Denmark; 1995.
76. Sakuradani E, Kobayashi M, Shimizu S. Δ^9 -fatty acid desaturase from arachidonic acid-producing fungus: Unique gene sequence and its heterologous expression in a fungus, *Aspergillus*. *Eur J Biochem.* 1999;260(1):208–216.
77. Andersen MR, Nielsen ML, Nielsen J. Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. *Mol Syst Biol.* 2008;4(1):178.
78. Bernstein DB, Sulheim S, Almaas E, Segrè D. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biol.* 2021;22(1):64.

79. Chen Y, Gustafsson J, Tafur Rangel A, Anton M, Domenzain I, Kittikunapong C, Li F, Yuan L, Nielsen J, Kerkhoven EJ. Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO Toolbox 3.0. *Nat Protoc.* 2024;19(3):629–667.
80. Li F, Chen Y, Qi Q, Wang Y, Yuan L, Huang M, Elsemman IE, Feizi A, Kerkhoven EJ, Nielsen J. Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints. *Nat Commun.* 2022;13(1):2969.
81. National Center for Biotechnology Information. Genome. National Center for Biotechnology Information. 2004. [Accessed 30 Apr 2025] <https://www.ncbi.nlm.nih.gov/datasets/genome/>
82. National Research Institute of Brewing (NRIB). Comprehensive *Aspergillus oryzae* genome database across the clades (CAoGDX). NRIB. 31 Mar 2025. [Accessed 1 May 2025] <https://nribf21.nrib.go.jp/CAoGDX/>
83. Leinonen R, Sugawara H, Shumway M. International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* 2011;39(Suppl_1):D19–D21.
84. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 2021;38(10):4647–4654.
85. Tegenfeldt F, Kuznetsov D, Manni M, Berkeley M, Zdobnov EM, Kriventseva EV. OrthoDB and BUSCO update: Annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res.* 2025;53(D1):D516–D522.
86. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 2008;24(5):637–644.
87. Chaudhari NM, Gupta VK, Dutta C. BPGA—An ultra-fast pan-genome analysis pipeline. *Sci Rep.* 2016;6(1):24373.
88. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26(19):2460–2461.
89. Klitgord N, Segre D. The importance of compartmentalization in metabolic flux models: Yeast as an ecosystem of organelles. *Genome Inform.* 2010;22:41–55.
90. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* 2016;44(D1):D1214–D1219.
91. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, et al. PubChem 2025 update. *Nucleic Acids Res.* 2025;53(D1):D1516–D1525.
92. Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, Nielsen J, Kerkhoven EJ. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLOS Comput Biol.* 2018;14(10):Article e1006541.
93. Thumhuri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. DeepLoc 2.0: Multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* 2022;50(W1):W228–W234.
94. Noor E, Haraldsdóttir HS, Milo R, Fleming RMT. Consistent estimation of Gibbs energy using component contributions. *PLoS Comput Biol.* 2013;9(7):Article e1003098.
95. Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Comput Biol.* 2017;13(4):Article e1005494.
96. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell JA, Blank LM, Chauhan S, Correia K, et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol.* 2020;38(3):272–276.
97. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847–2849.
98. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc.* 2010;5(1):93–121.
99. Medlock GL, Moutinho TJ, Papin JA. Medusa: Software to build and analyze ensembles of genome-scale metabolic network reconstructions. *PLoS Comput Biol.* 2020;16(4):Article e1007847.
100. Zhao G, Yao Y, Qi W, Wang C, Hou L, Zeng B, Cao X. Draft genome sequence of *Aspergillus oryzae* strain 3.042. *Eukaryot Cell.* 2012;11(9):1178.
101. Chacón-Vargas K, McCarthy CO, Choi D, Wang L, Yu JH, Gibbons JG. Comparison of two *Aspergillus oryzae* genomes from different clades reveals independent evolution of alpha-amylase duplication, variation in secondary metabolism genes, and differences in primary metabolism. *Front Microbiol.* 2021;12:Article 691296.
102. Zhao G, Yao Y, Hou L, Wang C, Cao X. Draft genome sequence of *Aspergillus oryzae* 100-8, an increased acid protease production strain. *Genome Announc.* 2014;2(3):Article e00548-14.
103. National Institute of Technology and Evaluation. NBRC online catalogue. National Institute of Technology and Evaluation. 2025. [Accessed 3 May 2025] <https://www.nite.go.jp/nbrc/catalogue/>
104. Zhang L, Kang L, Xu Y. Phenotypic, genomic, and transcriptomic comparison of industrial *Aspergillus oryzae* used in Chinese and Japanese soy sauce: Analysis of key proteolytic enzymes produced by koji molds. *Microbiol Spectr.* 2023;11(2):Article e0083622.
105. Thammarongtham C, Nookaew I, Vorapreeda T, Srisuk T, Land ML, Jeennor S, Laoteng K. Genome characterization of oleaginous *Aspergillus oryzae* BCC7051: A potential fungal-based platform for lipid production. *Curr Microbiol.* 2018;75(1):57–70.
106. Jeon J, Kim JA, Park S-Y, Kim GW, Park CS, Kim C, Park HY, Yeo JH, Lee YH, Kim S. Draft genome sequence of *Aspergillus oryzae* BP2-1, isolated from traditional malted rice in South Korea. *Microbiol Resour Announc.* 2020;9(1):Article 01405-19.
107. Coblenz M, Evans JD, Kothe CI, Mak T, Valerón NR, Chwalek P, Wejendorp K, Garg S, Pless L, Mak S, et al. Food fermentation in space: Opportunities and challenges. *iScience.* 2025;28(4):Article 112189.
108. Westerdijk Fungal Biodiversity Institute. 2025. [Accessed 3 May 2025] <https://wi.knaw.nl/>
109. Jeong E, Seo J-A. Comparative genome analyses of *Aspergillus oryzae* and *Aspergillus flavus* originated from a Korean fermentation starter, nuruk. *Food Microbiol.* 2025;131:Article 104807.
110. Alshannaq AF, Gibbons JG, Lee MK, Han KH, Hong SB, Yu JH. Controlling aflatoxin contamination and propagation of *Aspergillus flavus* by a soy-fermenting *Aspergillus oryzae* strain. *Sci Rep.* 2018;8(1):16871.

111. Han DM, Baek JH, Choi DG, Jeon CO. Transcriptomic analysis of the metabolic characteristics of *Aspergillus oryzae* during the fermentation of doenjang-meju, a traditional Korean fermented soybean brick. *LWT*. 2024;208: Article 116740.
112. Wicklow DT, McAlpin CE, Peterson SW. Common genotypes (RFLP) within a diverse collection of yellow-green aspergilli used to produce traditional oriental fermented foods. *Mycoscience*. 2002;43(4):289–297.
113. US Department of Agriculture. ARS culture collection (NRRL). US Department of Agriculture. 2025. [Accessed 3 May 2025] <https://nrnl.ncaur.usda.gov/>
114. Wicklow DT, McAlpin CE, Yeoh QL. Diversity of *Aspergillus oryzae* genotypes (RFLP) isolated from traditional soy sauce production within Malaysia and Southeast Asia. *Mycoscience*. 2007;48(6):373–380.
115. Chaves FC, Gianfagna TJ, Aneja M, Posada F, Peterson SW, Vega FE. *Aspergillus oryzae* NRRL 35191 from coffee, a non-toxicogenic endophyte with the ability to synthesize kojic acid. *Mycol Prog*. 2012;11(1):263–267.
116. Microbe Division (JCM). JCM catalogue. Microbe Division (JCM). 2025. [Accessed 3 May 2025] <https://jcm.brc.riken.jp/en/>
117. Okabe T, Katayama T, Mo T, Mori N, Jin FJ, Fujii I, Iwashita K, Kitamoto K, Maruyama JI. BiFC-based visualisation system reveals cell fusion morphology and heterokaryon incompatibility in the filamentous fungus *Aspergillus oryzae*. *Sci Rep*. 2018;8(1):2922.
118. Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, Klich MA, Tabb DL, McDonald WH, Rokas A. The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. *Curr Biol*. 2012;22(15):1403–1409.
119. National Center for Biotechnology Information (NCBI). Genome assembly ASM221495v1 (GCA_002214955.1) of *Aspergillus oryzae* strain SRCM101975. National Center for Biotechnology Information. 5 Jul 2017. [Accessed 3 May 2025] https://www.ncbi.nlm.nih.gov/assembly/GCA_002214955.1
120. National Center for Biotechnology Information (NCBI). Genome assembly ASM221496v1 (GCA_002214965.1) of *Aspergillus oryzae* strain SRCM101989. National Center for Biotechnology Information. 5 Jul 2017. [Accessed 3 May 2025] https://www.ncbi.nlm.nih.gov/assembly/GCA_002214965.1

***pAo*: A Consensus Genome-Scale Metabolic Model for *Aspergillus oryzae* Capturing Intraspecies Diversity**

Jeroen Gilis, Casper Robert Balten van der Luijt, Marilena Feller, Candela Sánchez-Girón Barba, Morten Otto Alexander Sommer, Leonie Johanna Jahn, and Eduard Johannes Kerkhoven

Citation: Gilis J, van der Luijt C, Feller M, Barba C, Sommer M, Jahn L, Kerkhoven E. *pAo*: A Consensus Genome-Scale Metabolic Model for *Aspergillus oryzae* Capturing Intraspecies Diversity. *Comput Struct Biotechnol J*. 2026;**35** (1):0173. DOI: 10.34133/csbj.0173

Aspergillus oryzae (koji mold) is a key microorganism in traditional food fermentations including soy sauce, sake, and miso and is important in novel culinary applications and modern biotechnology, such as sustainable meat alternatives and enzyme production. Despite its industrial importance, until recently, the most recent genome-scale metabolic model (GEM) for *A. oryzae* dated back to 2008 and was limited to a single strain (RIB40). Here, we present *pAo*, a pan-GEM for *A. oryzae*, integrating genomic data from 187 strains to capture species-wide metabolic diversity. Our model comprises 2,025 reactions, representing a 52% increase in metabolic coverage over the RIB40-based model and includes previously overlooked pathways, such as cytochrome P450-mediated xenobiotic metabolism and extended amino acid metabolism. Using this pan-GEM, we derived strain-specific GEMs and validated 8 of them through high-throughput phenotypic screening on 285 substrates. Growth experiments on 4 industrially relevant carbon sources revealed substantial interstrain metabolic diversity, although flux balance analysis indicated that this variability originates at the regulatory rather than network-structural level. This resource provides a foundation for informed strain selection for biotechnological applications and future metabolic engineering in *A. oryzae*.

Image

View the article online

<https://spj.science.org/doi/10.34133/csbj.0173>

Use of this article is subject to the [Terms of Service](#)

Computational and Structural Biotechnology Journal (ISSN 2001-0370) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Copyright © 2026 Jeroen Gilis et al.

No claim to original U.S. Government Works. Distributed under a [Creative Commons Attribution License \(CC BY 4.0\)](#).