AlphaGEM enables precise genome-scale metabolic modelling by integrating protein structure alignment with deep-learning-based dark

nature communications

Article

https://doi.org/10.1038/s41467-026-75549-w

AlphaGEM enables precise genome-scale metabolic modelling by integrating protein structure alignment with deep-learning-based dark metabolism mining

Received: 10 November 2025

Accepted: 30 June 2026

Published online: 16 July 2026

Check for updates

Weishang Han[1,10], Luchi Xiao[1,10], Haocheng Sun[1,10], Guangming Xiang[1], Qianxi Jia[1,2], Haoyu Wang[1,3,4], Boyang Ji[5], Cheng Zhang[6,7], Eduard J. Kerkhoven[8,9], Jens Nielsen[5,9] & Hongzhong Lu[1]

Constructing high-quality genome-scale metabolic models (GEMs) for non-model organisms remains challenging. To address this, we developed AlphaGEM, a versatile toolbox leveraging proteome-scale structural alignment, protein language models (PLMSearch), and deep-learning-based predictions for efficient genomic mining to generate GEMs ready for applications. AlphaGEM enhances homologous relationship identification compared to traditional sequence-based methods. Crucially, it employs an ensemble procedure empowered by multiple deep learning toolboxes to effectively mine dark metabolic functions encoded by nonhomologous proteins, thereby expanding species-specific networks. We validated AlphaGEM across prokaryotes (*Klebsiella pneumoniae*, *Bacillus subtilis*), eukaryotes (*Rhodosporidium toruloides*, *Pichia pastoris*), and complex mammals (*Mus musculus*, *Cricetulus griseus*), achieving predictions comparable to manually curated models while outperforming existing tools. Furthermore, we demonstrated its scalability by automatically reconstructing high-fidelity GEMs for 332 distinct yeast species. In summary, AlphaGEM enables precise, rapid GEM construction across diverse domains, providing a solid foundation for universal functional analysis of organisms having genome sequences available.

Genome-scale metabolic models (GEMs) are powerful computational frameworks for characterizing cellular metabolism, enabling the prediction of physiological responses under various genetic and environmental perturbations[1–4]. GEMs have diverse applications, such as optimizing microbial strains for biochemical production in metabolic engineering and identifying potential therapeutic targets in drug discovery for precision medicine[4–7]. Furthermore, GEMs allow for the investigation of metabolic interactions within microbial communities,

[1]State Key Laboratory of Microbial Metabolism, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. [2]CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [3]Division of Biotechnology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China. [4]University of Chinese Academy of Sciences, Beijing, China. [5]BioInnovation Institute, Ole Måløes Vej, Copenhagen N, Denmark. [6]Science for Life Laboratory, KTH-Royal Institute of Technology, Stockholm, Sweden. [7]Laboratory of Advanced Drug Preparation Technologies, School of Pharmaceutical Sciences, Ministry of Education, Zhengzhou University, Zhengzhou, China. [8]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. [9]Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden. [10]These authors contributed equally: Weishang Han, Luchi Xiao, Haocheng Sun. e-mail: hongzhonglu@sjtu.edu.cn

providing insights into ecological dynamics and adaptive evolution[8]. Integrating omics data with advanced algorithms further enhances the predictive capabilities of GEMs, expanding their applications in personalized medicine and synthetic biology[9–12].

Currently, the increasing accumulation of large-scale genomic data from high-throughput sequencing experiments has created a great demand for high-quality GEMs for computational analysis of cellular metabolism[13]. Note that various GEMs for model organisms that are widely used have been going through a combination of semi-automated and manual curation over many years (i.e., yeast-GEM[14], *i*ML1515[15], human-GEM[16]), where the quality of the model is improved with each iteration. However, the labour-intensive approach used for model organisms is not feasible for large numbers of newly sequenced species. Meanwhile, the complexity and variety in metabolic networks across species seriously hinder the reconstruction of high-quality, organism-specific GEMs for many less characterized species[17]. Therefore, the development of next-generation computational toolboxes capable of producing ready-to-use GEMs is of great value for the community.

Until now, several automated modelling tools, including CarveMe[18], ModelSEED[10], gapseq[19], *merlin*[20], MetaDraft[21], and AuReMe[22] have been developed to facilitate the reconstruction of GEMs. However, these tools often produce models with suboptimal quality and accuracy, requiring manual corrections before they are ready for use[23]. CarveMe, for example, generates metabolic models for the target organism by manually building universal metabolic network templates[18]. Although the metabolic models it generates have more reactions without gene annotations and can be used directly for flux balance analysis (FBA), those GEMs frequently have low specificity, with high similarity between models from different species. Other template-based tools, including MetaDraft and AuReMe, build models from predefined template models. These tools have the same issues as CarveMe, and they perform poorly when annotating proteins that do not have homologous proteins in the template model's species, which may not reflect species-specific metabolic characteristics. ModelSEED, *merlin*, and gapseq, on the other hand, begin with the genome, then utilize various bioinformatics tools to annotate genes, potentially improving the completeness of metabolic models. However, these toolboxes may suffer from poor protein annotation, which, to some extent, lowers the model quality[23]. Addressing these issues is highly critical to promote the usability of GEMs by the wider scientific community.

In this work, to increase the quality of GEMs and accurately characterize the metabolic diversity of numerous newly sequenced species, we create AlphaGEM, a versatile computational toolbox that integrates protein homology inference and dark metabolism mining to address the bottleneck existing in current toolboxes for automated GEM reconstruction. The performance of AlphaGEM is thoroughly assessed, and multiple case studies are used to demonstrate its accuracy and efficiency in creating GEMs for various organisms from simple bacterial organisms to complex mammals. Overall, AlphaGEM enables high-quality and high-throughput GEMs reconstruction through protein structure alignment and deep learning, which will certainly accelerate model-based simulation and analysis for a wide range of organisms.

Results

Overview of AlphaGEM framework

Recent advancements in deep learning offer unprecedented opportunities to determine gene and protein function with high resolution. We introduce AlphaGEM, a versatile framework for reconstructing high-quality GEMs for sequenced organisms through protein function inference derived from protein 3D structure comparison and protein language model prediction. Different from previous studies[10,18], which select mature universal models to rebuild GEMs for non-model organisms, we utilize curated high-quality GEMs from model organisms as references—employing *i*ML1515[15] as the template model for prokaryotes, Yeast9[14] for unicellular eukaryotes and fungi, and Human1[16] for mammalian cells (Fig. 1). Within AlphaGEM's initial module, draft-GEMs could be generated in an automated manner from the aforementioned reference models based on the inferred protein homologous relationships between the model and non-model organisms. Departing from conventional approaches that depend solely on sequence similarity to establish homology, AlphaGEM implements two complementary approaches: (1) analyzing genome-scale homologous relationships by combining sequence similarity and protein structure alignment when proteome-scale 3D structures are available, and (2) leveraging protein-language-model-based approaches, i.e., PLMSearch[24], to infer homologous relationships when proteome-scale structures are unavailable (Fig. 1, "Methods"), both of which are detailed in the following sections.

Although reference models were utilized to generate draft GEMs, there still exist a number of species-specific metabolic reactions and pathways in the target organisms. To enhance metabolic coverage, we implemented a bottom-up strategy by annotating nonhomologous proteins from target organisms. Traditional sequence-based annotation tools often exhibit limited accuracy[25], whereas recent advances in deep learning have provided more robust and scalable solutions. Here, we developed an ensemble pipeline with the aid of several advanced computational tools, such as CLEAN[26], DeepECtransformer[27], and eggNOG-mapper[28], to annotate nonhomologous proteins encoding metabolic proteins within the genomes from the input organisms (Fig. 1), thereby expanding the scope of GEMs for non-model organisms. Lastly, to address the potential gaps in metabolic networks resulting from inaccuracies existing in genome sequencing or assembly, we performed gap-filling with reference models as templates using iterative gap-filling and evaluation[17]. Collectively, AlphaGEM seamlessly integrates a suite of advanced modules to provide a reliable and effective solution for modelling the metabolism of newly sequenced organisms.

Cross-species inference of homologous relationships by integrating sequence and structure information with AlphaGEM

The quality of initial draft-GEMs critically depends on the reliable annotation of protein homologous relationships between input and reference proteins. Traditional approaches typically rely on sequence alignment tools like HMMER[29], blastp[30], and OrthoFinder[31]. However, using sequence alignment to establish protein homologous relationships between distinct species is frequently insufficient, as certain proteins with similar functions may not be identified due to limited sequence similarity[32,33]. To address this limitation, we incorporated proteome-scale structural alignment to refine the inference of homologous relationships beyond what is possible with sequence alignment alone ("Methods", Supplementary Fig. 1a).

To validate the reliability of the homologous relationships obtained through the above procedures, we evaluated the performance of AlphaGEM in the inference of homologous relationships using proteomes from two representative fungal species, *Candida albicans* and *Saccharomyces cerevisiae*, the latter of which was set as the reference organism. At first glance, we observed that protein pairs exhibiting high sequence similarity generally align with those showing high structural similarity (Fig. 2a). However, a non-negligible fraction of protein pairs with high structural similarity have relatively low sequence similarity (Fig. 2b). Therefore, focusing on GEMs, we compared the homologous groups of metabolic proteins inferred by different steps encompassed within AlphaGEM: (1) OrthoFinder based grouping, (2) US-align based filtering, and (3) Foldseek-based expansion ("Methods", Fig. 2c). Compared with the first step, the US-align-based structural clustering approach could successfully decrease false homologs from 63 to 46 and improve the accuracy from 94.91% to 95.82% (Supplementary Fig. 2f). This shows that structure-based comparison can help to identify additional homologous

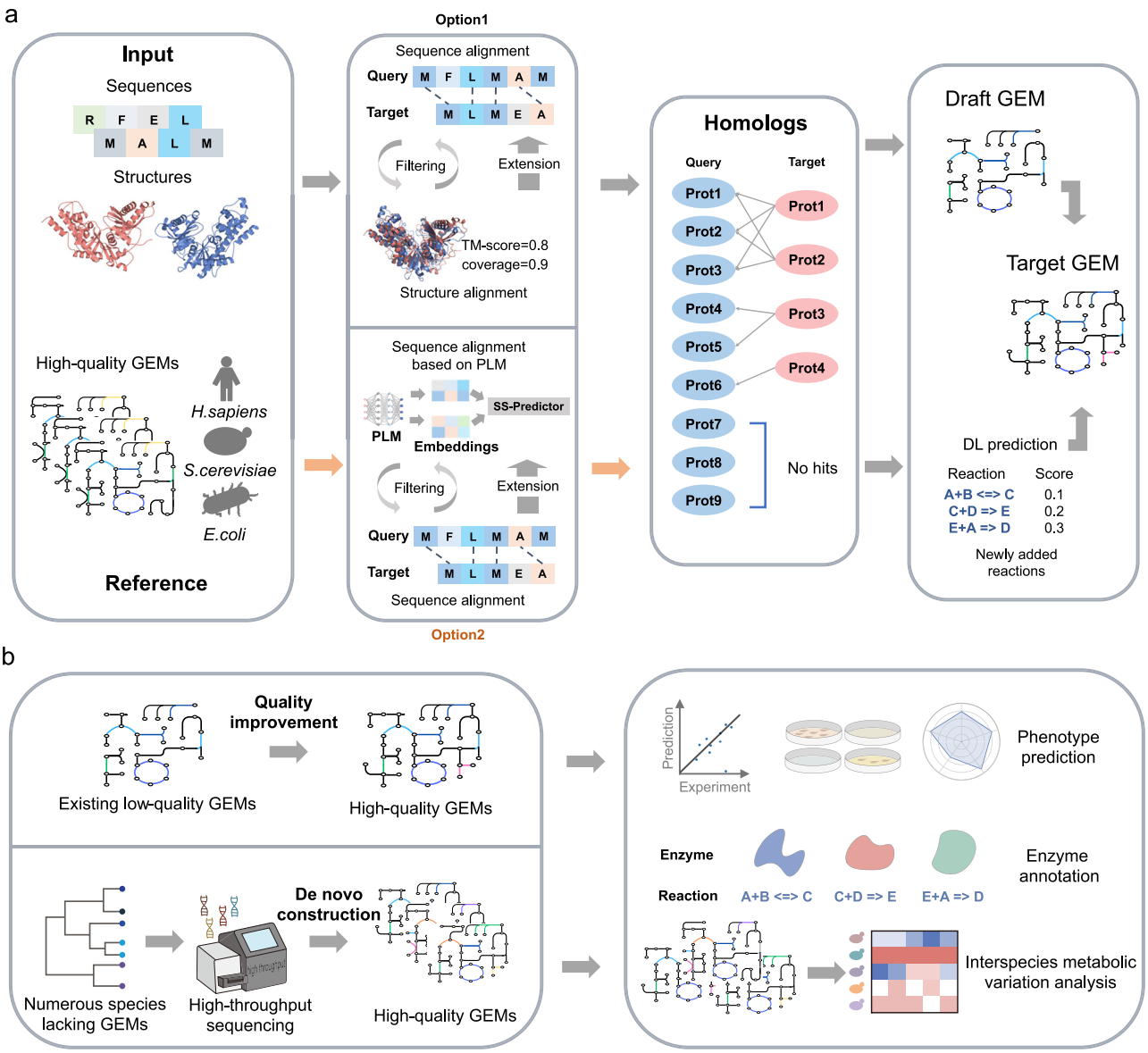

Fig. 1 | Framework used by AlphaGEM for constructing high-quality GEMs for non-model organisms. With model organisms' high-quality GEMs as references, AlphaGEM integrates biological sequence and protein structural information to identify homologous relationships via combined sequence/structure alignment (or PLMSearch when structure is unavailable). This homologous information drives the construction of a draft GEM. Subsequently, nonhomologous proteins undergo functional annotation, scoring, and filtering using multiple deep learning (DL) tools, and are integrated into the draft GEM. Finally, the above outputs are combined to obtain ready-to-use GEMs for specific organisms (**a**). Typical applications of AlphaGEM. For example, AlphaGEM enables optimization of low-quality GEMs or de novo reconstruction of GEMs for non-model organisms. GEMs generated by AlphaGEM could be used to perform systematic, quantitative, and predictive analyses of cellular metabolism, mine enzyme functions, and probe the interaction of distinct strains within a community (**b**).

relationships as some of the homologous pairs are more conserved in protein 3D structure than in protein sequence (Fig. 2d, e). Importantly, it is found that sequence-based homologous protein pairs with high percent identity (identity) but low TM-score could have distinct metabolic functions (Fig. 2f, g), showcasing the necessity of filtering out low-quality orthologous protein relationships using structural alignment. Overall, our whole pipeline encompassing the above three steps identified more homologous pairs (an increase of 9.01%) than the sequence-based clustering approach due to the extra Foldseek-based structure comparison at the proteome scale. Moreover, the increased number of homologous pairs resulted in an increase in accuracy, from 94.91% up to 95.36% (Fig. 2c).

Notably, for more distantly related species and complex organisms such as mammals, increasing the threshold used in protein structure alignment based on Foldseek could improve accuracy in identifying protein homologous relationships, while still keeping a large number of homologous pairs between two organisms ("Methods", Supplementary Fig. 2e and Supplementary Table 1), thus the structure-based protein function inference can serve as a better approach than the sequence-based procedure in refining the protein homologous relationships between two chosen species.

Protein language model accelerates the inference of homologous relationships between proteins across species without protein structure comparison

To ensure the broad applicability of AlphaGEM under limited computational resources, particularly for large-scale analyses involving hundreds of species, we developed a novel procedure for inferring homologous relationships with the aid of PLMSearch[24] ("Methods", Fig. 1a, Supplementary Fig. 1b, and Supplementary Note 1). This

https://doi.org/10.1038/s41467-026-75549-w

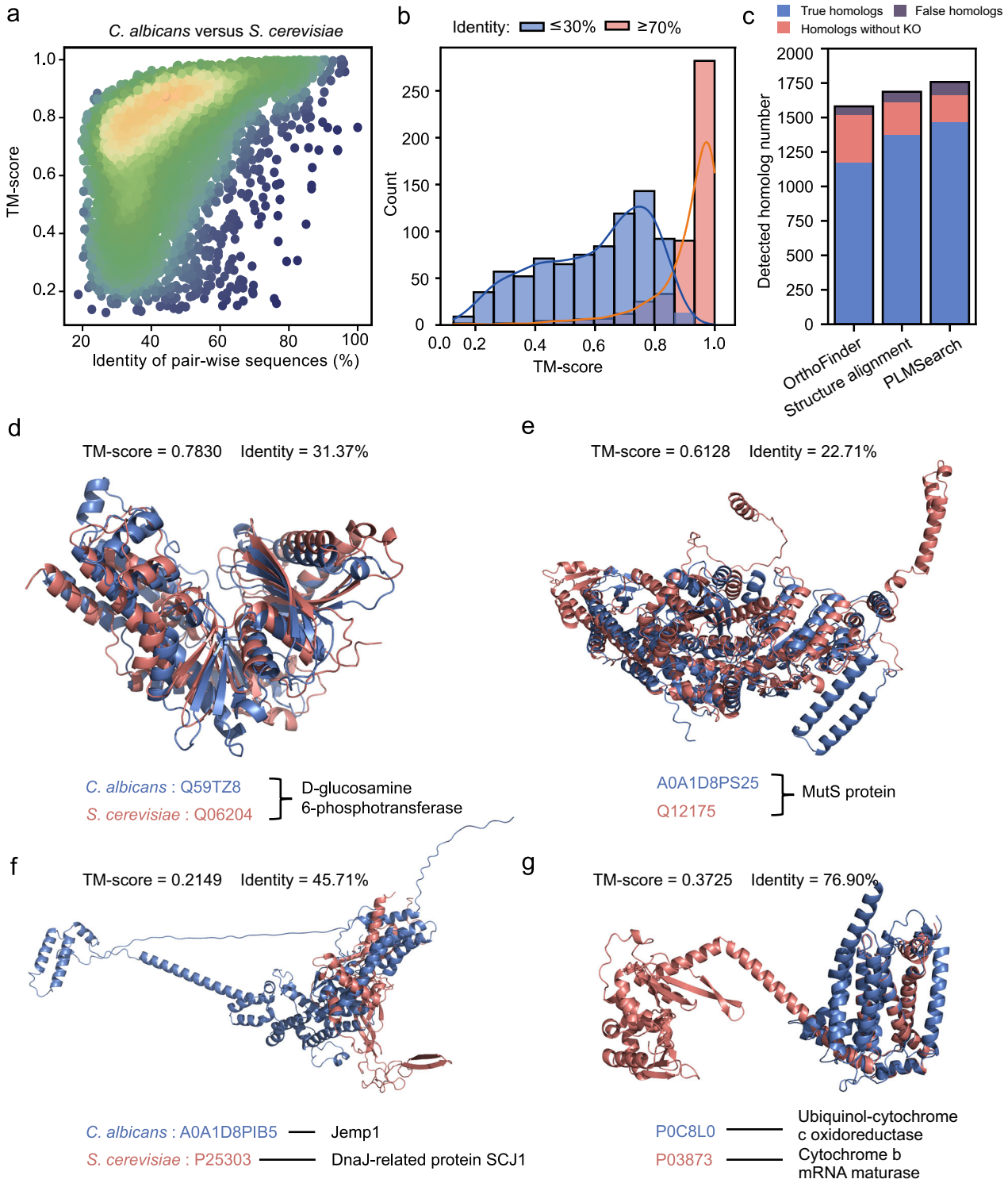

Fig. 2 | AlphaGEM enhances the inference of protein homologous relationships between *C. albicans* and *S. cerevisiae* by combining sequence alignment and structure alignment. Distribution plots of sequence identity from sequence alignment and TM-scores from structural alignment (**a**). Distribution of TM-scores for gene pairs with high (≥ 70%) and low (≤ 30%) sequence identity (**b**). Accuracy in inference of protein homology using three procedures: the alignment based on sequence blast, the alignment integrating both sequence and structural information and the alignment based on PLMSearch. Here, the homologous relationships annotated by KEGG Orthology (KO) groups were used as the reference (**c**). Protein pairs with low sequence identity (OrthoFinder non-clustered) but high structural similarity (TM-score > 0.6) could still have the same catalytic functions (**d**, **e**). Protein pairs with high sequence identity but low structural similarity (TM-score <0.4) show divergent functions (**f**, **g**). Source data are provided as a Source Data file.

alternative strategy mainly uses protein language models to detect homologous protein pairs without relying on structural data. Taking *C. albicans* as an example (Fig. 2c and Supplementary Fig. 3a), this strategy achieved accuracy comparable to the above integrative approach, accounting for both sequence and structure similarity (93.92% vs. 95.36%) while identifying additional homologous pairs (14.61%) missed by the latter.

To improve the reliability of the new strategy, we introduced a post-filtering step after the PLMSearch-based inference of homologous relationships, in which the sequence coverage (> 0.6 in *C. albicans* and >0.7 in mouse) and identity (> 30% in *C. albicans* vs. >45% in mouse) acquired with blastp[30] were used to refine the output of PLMSearch ("Methods", Supplementary Fig. 1b) to balance sensitivity and specificity (Supplementary Fig. 3) in identifying homologs. For instance, excessively stringent identity thresholds (> 50% in *C. albicans*) led to a sharp decline in valid homologous pairs, while overly lenient thresholds increased false positives (FP). What's more, we found that a few homologous relationships missed by PLMSearch can be recovered with the classical bidirectional best hits (BBHs) identified through global BLAST alignment. Notably, species-specific parameter setting was critical; for example, the sequence divergence in mammalian genomes (e.g., *M. musculus*) required higher identity thresholds to be consistent with structural alignment outcomes (Supplementary Fig. 3b). Taken together, alternative inference of homologous relationships using PLMSearch could be resource-efficient for high-throughput GEM reconstruction, whereas structure-based methods remain preferable for precision-focused tasks when computational resources permit.

Improved annotation for nonhomologous proteins using an ensemble pipeline

Accurately annotating nonhomologous proteins in the input organism is an important step in identifying species-specific metabolic pathways in order to refine GEMs. Traditional approaches, such as eggNOG-mapper, are extensively used for annotating unknown proteins, but suffer from poor precision[26]. Relying on breakthroughs in deep learning, several methods have shown much better prediction performance than traditional approaches[27]. However, using a single procedure for protein function prediction frequently generates unreliable results, which may lower the quality of the species-specific GEMs. To address this, AlphaGEM incorporates an ensemble pipeline integrating CLEAN, DeepECtransformer, eggNOG-mapper, and PLMSearch for comprehensive protein annotation to identify potential catalytic enzymes (Fig. 3a, "Methods"). To ensure robustness, we allocated weight scores to annotation results based on each tool's reliability and specificity. Tools with high sensitivity but low accuracy were given a lower weight score, whereas those with higher specificity received a higher weight score. For instance, PLMSearch, while effective in identifying correct homologous relationships between two organisms, might introduce an elevated false-positive rate when used for protein function annotation, resulting in lower F1-scores (Fig. 3b, c). Thus, a lower weight score was assigned to the output of PLMSearch when annotating the protein function with our ensemble pipeline. The sum of weight scores calculated from those four methods was used to rank the annotation results for each enzyme. We evaluated the annotation results using UniProt reference datasets for *C. albicans* and *Schizosaccharomyces pombe*, which have been manually curated and partially experimentally validated. Accuracy analysis of protein function annotations revealed that the F1-scores of our ensemble pipeline (0.379, 0.709) for two yeast species were both higher than those of individual tools (Fig. 3b, c): CLEAN (0.326, 0.654), DeepECtransformer (0.189, 0.423), PLMSearch (0.097, 0.105), and eggNOG-mapper (0.094, 0.137). It demonstrates that our integrative pipeline could improve the enzyme function prediction in both F1-score and precision when compared to any single procedure, demonstrating the reliability of our pipeline. Subsequent evaluation shows that the annotation F1-score increased with the number of integrated tools (Fig. 3d, e). Note that the selected optimal threshold in total weight score used in AlphaGEM could be adjusted to enhance the annotation performance of non-homologous proteins according to the input organisms (Fig. 3f, g).

Furthermore, our pipeline facilitates the discovery of "dark metabolism" in non-model organisms (Supplementary Note 2). A notable example is the *Rhodosporidium toruloides* protein M7XU53 (RHTO_06802). While it lacks a definitive reaction annotation in UniProt and is absent in current manually curated GEMs[34], AlphaGEM robustly identified it as a salicylate 1-monooxygenase (EC 1.14.13.1). This reaction involves the oxidative decarboxylation of salicylate, consuming one molecule of NADH and O_2 to yield catechol, CO_2, and H_2O. This assignment is supported by high structural similarity to the known protein NHG1_PSEPU (US-align TM-score = 0.84). Given that *R. toruloides* is experimentally known to utilize various aromatic compounds, this prediction represents a biologically consistent discovery supported by genomic context, rather than a database-driven FP[35,36]. However, the functional assignments reported here should be considered computational predictions warranting experimental validation, rather than confirmed biochemical activities.

Comprehensive evaluation of AlphaGEM in generating ready-to-use GEMs

To systematically evaluate the performance of AlphaGEM in generating ready-to-use GEMs, we first constructed GEMs for *C. albicans* and *S. pombe*, two widely used model yeast species. We assessed the prediction capabilities of these models for predicting carbon source utilization and essential genes, respectively.

In order to probe how each step within the AlphaGEM framework influences model output, we built models for *C. albicans* and *S. pombe* using homologous pairs generated at different stages: (1) OrthoFinder-based orthology mapping, (2) US-align structural filtering, and (3) Foldseek-based structural expansion ("Methods"). Our analysis revealed that US-align structural filtering could reduce the coverage of proteins in the GEMs of *C. albicans* and *S. pombe*. The subsequent Foldseek-based structural expansion partially recovered the coverage of proteins. Crucially, while US-align filtering minimally affected the counts of metabolic reactions in two GEMs, the Foldseek-based structural expansion enhanced metabolic reaction coverage. Finally, our ensemble pipeline enabled high-throughput protein function prediction, through which additional genes, reactions, and metabolites could be merged into GEMs (Supplementary Figs. 4a, b and 5c). Beyond merely expanding network coverage, the ablation study ("Methods") demonstrated that these structure-guided functional annotations are biologically critical; incorporating the structural alignment module improved the model's accuracy in predicting metabolic capabilities, such as carbon source utilization, indicating that the structurally-informed models are more complete and more accurate (Supplementary Fig. 5).

Using this integrated framework, the final GEMs for two yeast species were generated and benchmarked against both manually curated models[37,38] and those reconstructed using the automated tool CarveFungi[39]. Our models demonstrated increased coverage at the reaction and metabolite levels and gene coverage comparable to manually curated models (Fig. 4a–c). By further comparison, we found that *C. albicans* GEMs reconstructed by AlphaGEM performed well in predicting the single-carbon-source utilization (accuracy = 78.26%) and essential genes (accuracy = 79.90%) (Fig. 4d), yielding performance comparable to that of a manually curated model[37] (97.44 and 73.04%, respectively). When systematically benchmarked against CarveFungi, AlphaGEM also exhibited a substantial advantage; the CarveFungi-generated model for *C. albicans* only achieved accuracies of 27.27% for carbon source utilization and 62.18% for gene essentiality prediction. Notably, AlphaGEM outperforms the manually

https://doi.org/10.1038/s41467-026-75549-w

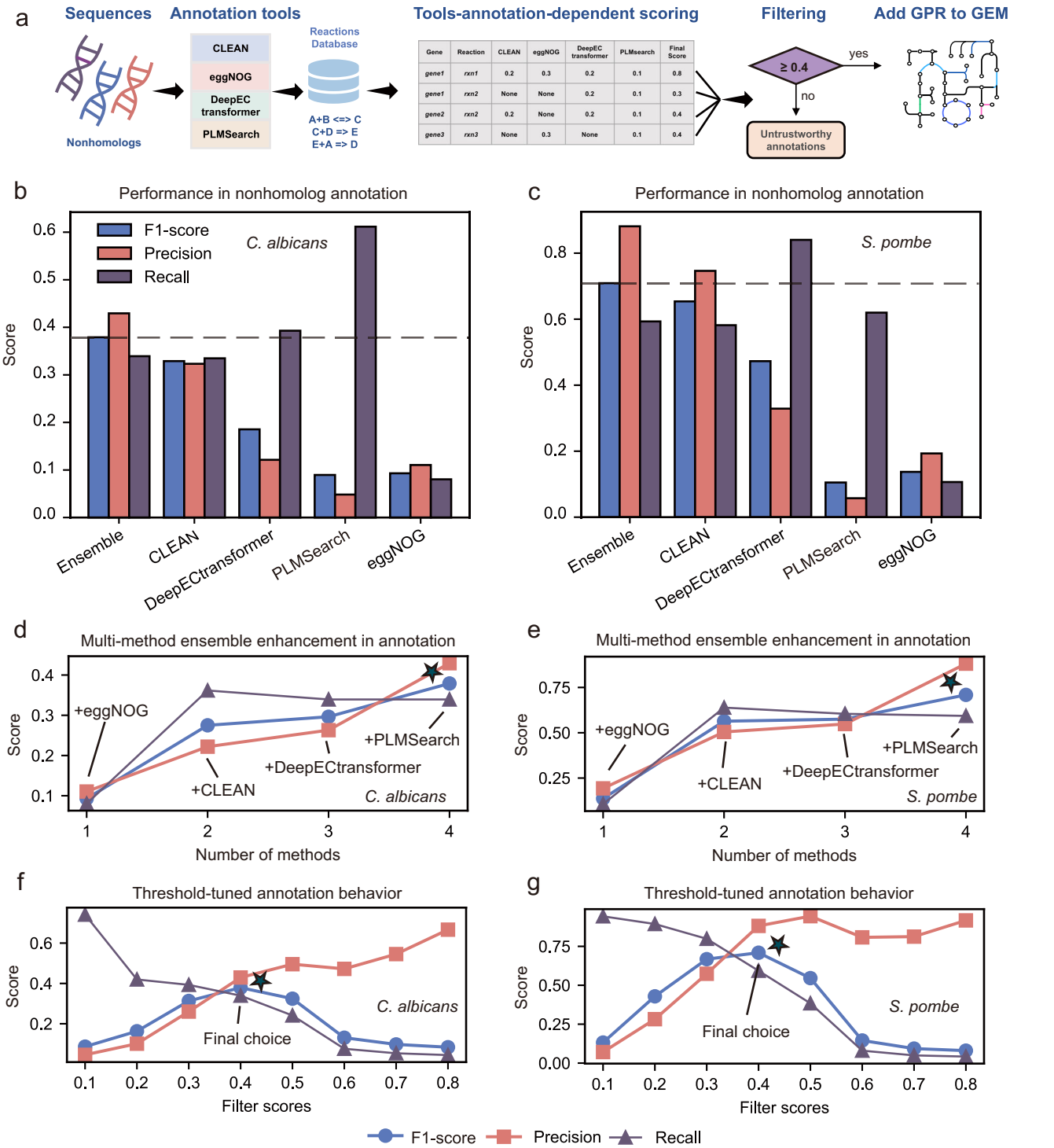

Fig. 3 | Systematic annotation of nonhomologous proteins in AlphaGEM. AlphaGEM executes parallel functional annotation using CLEAN (EC numbers), DeepECtransformer (EC numbers), PLMSearch-based function search (Rhea reactions), and eggNOG-mapper (KO). All annotations are mapped to Rhea reactions, and each protein-reaction pair receives a cumulative confidence score based on tool-specific weights: CLEAN (0.2), DeepECtransformer (0.2), PLMSearch (0.1), and eggNOG-mapper (0.3). Pairs with score ≥ 0.4 are retained; their gene-protein-reaction (GPR) rules are reconstructed and integrated into the draft GEM (**a**). Performance of individual annotation tools and the ensemble method was compared for *C. albicans* (**b**) and *S. pombe* (**c**). Profiles of annotation performance when incrementally adding tools into the ensemble method for the nonhomologous protein annotation in *C. albicans* (**d**) and *S. pombe* (**e**). Effects of filtering score on performance of nonhomologous protein annotation in *C. albicans* (**f**) and *S. pombe* (**g**). Source data are provided as a Source Data file.

reconstructed models in essential gene prediction. By contrast, both AlphaGEM and CarveFungi exhibited lower accuracy than the manually curated model in predicting carbon source utilization for *C. albicans*, highlighting a persistent bottleneck for automated tools in the absence of iterative manual refinement. Such discrepancies are particularly pronounced in opportunistic pathogens like *C. albicans*, where host-driven genomic plasticity leads to two primary modeling artifacts: false-positive over-annotations arising from expanded paralogous gene families with ambiguous functions[40], and false-negative omissions of highly divergent enzymes that elude traditional

https://doi.org/10.1038/s41467-026-75549-w

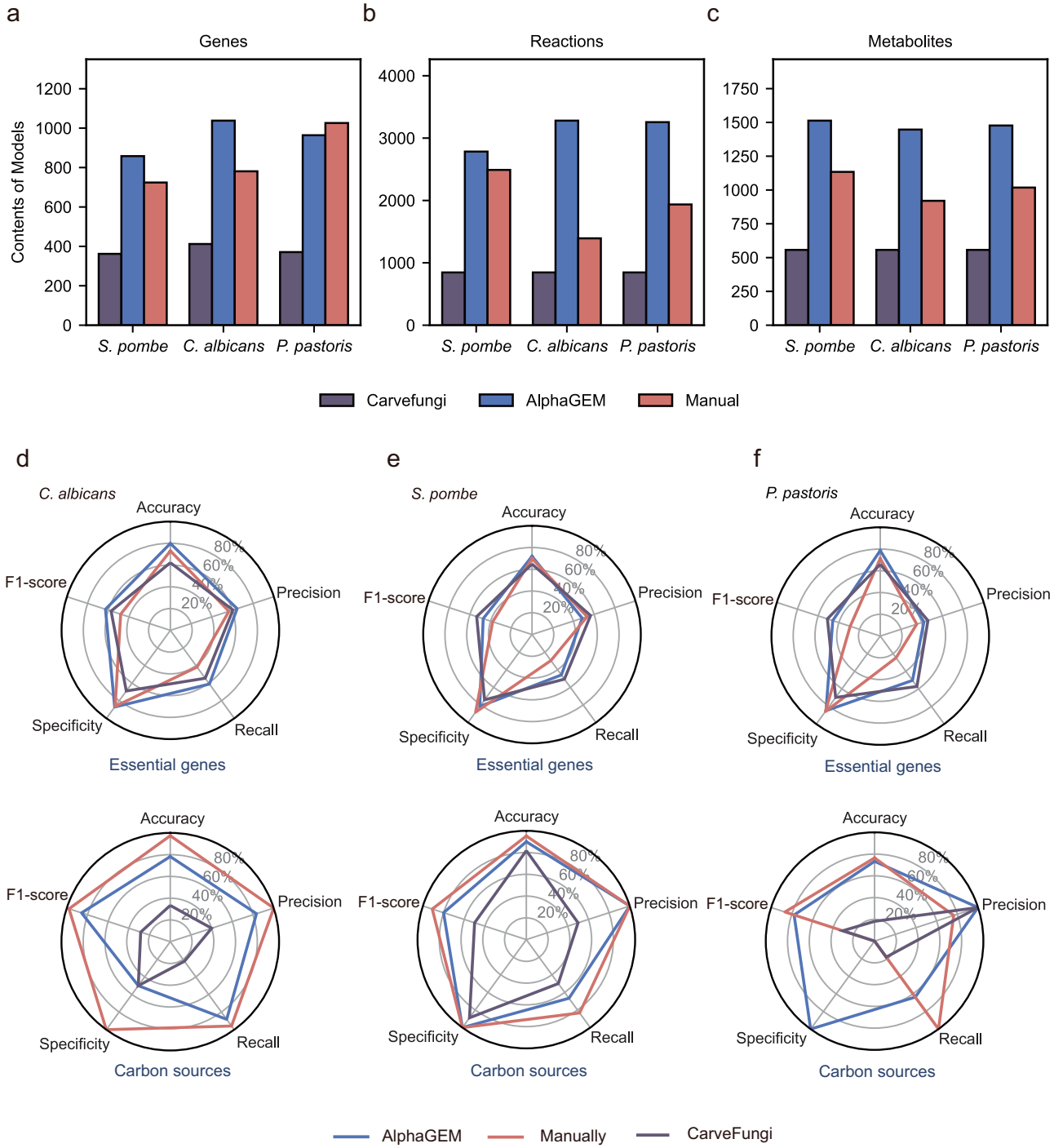

Fig. 4 | Quality assessment of GEMs for three unicellular eukaryotes reconstructed by AlphaGEM. Comparison of the number of genes (**a**), reactions (**b**), metabolites (**c**) among GEMs of *C. albicans*, *S. pombe*, and *P. pastoris* from manual curation, AlphaGEM, and CarveFungi. Comparison of accuracy in predicting carbon source utilization and essential genes by different GEMs from manual curation, AlphaGEM, and CarveFungi for *C. albicans* (**d**), *S. pombe* (**e**), and *P. pastoris* (**f**). Source data are provided as a Source Data file.

sequence-based detection[41]. Specifically, we discovered that the GEMs built by AlphaGEM could predict the phenotypes of *C. albicans* using xylitol and D-glucosamine without introducing additional false-positive predictions. The enzymes implicated in the major pathways for these two carbon sources were structurally recognized as proteins homologous to *Saccharomyces cerevisiae* enzymes (D-glucosamine 6-phosphotransferase and xylitol transporter), but no homologous proteins were discovered using sequence-based clustering (Fig. 2d). This demonstrates that our strategy could, to some extent, enhance the reliability of GEMs by refining the protein function annotation inferred solely by sequence-based analysis. Similarly, the GEMs built by AlphaGEM for *S. pombe* also achieved comparable accuracy in predicting carbon source utilization and essential genes (90.00%, 71.93%) compared to manually curated counterparts (95.24%, 69.89%) (Fig. 4e) and to the CarveFungi model (81.82%, 64.82%). Although CarveFungi showed a slightly higher F1-score for essential gene prediction in *S. pombe* (53.5% vs. 47.2%), this metric is biased by its severely limited gene representation (362 genes versus 858 genes in AlphaGEM;

Fig. 4a). While manually curated models benefit from iterative manual curation in carbon source utilization predictions by comparing these predictions with experimental datasets, AlphaGEM demonstrates comparable performance in generating high-quality GEMs for downstream applications.

Furthermore, to demonstrate the generalizability of AlphaGEM beyond the initial optimization species, we constructed GEMs for *R. toruloides* and *P. pastoris* using the same established parameters. Similarly, we evaluated the prediction performance of these two models. The results showed that these models achieved high accuracy in the prediction of carbon source utilization and gene essentiality (Fig. 4f and Supplementary Fig. 6a, b), comparable to the performance observed in our primary test species. Notably, due to the lack of reliable experimental data for carbon source utilization in *R. toruloides*, only gene essentiality prediction was performed for this species. The successful cross-species prediction underscores that AlphaGEM provides a robust and unbiased framework for the automated reconstruction of high-quality GEMs for diverse yeast species.

AlphaGEM enhances the reconstruction of high-quality GEMs for both bacteria and mammals

To further demonstrate AlphaGEM's performance across diverse organisms, we constructed GEMs for bacteria and mammals (Fig. 1a), respectively, and compared them with models generated by one of the state-of-the-art toolboxes and with manually curated counterparts.

As for bacterial modeling, we comprehensively assessed the quality of models generated by CarveMe[18], ModelSEED[10], gapseq[19], and AlphaGEM. *K. pneumoniae* is a ubiquitous human pathogen that can cause pneumonia, bloodstream infections, urinary tract infections, and other life-threatening infections. We generated GEMs for *K. pneumoniae* with these tools, respectively, and the models were compared to an experimentally validated model, iYL1228[42]. The efficacy of these generated models for predicting carbon-source utilization was compared to experimental datasets. AlphaGEM achieved a prediction accuracy of 75.32%, comparable to the manually curated model iYL1228 (81.8%) (Fig. 5a). In contrast, the models constructed by CarveMe, ModelSEED, and gapseq exhibited poor performance (Fig. 5a). Specifically, the CarveMe model yielded a low overall prediction accuracy of 45.45%. The model generated by ModelSEED inherently lacked exchange reactions for unutilized carbon sources, identifying only 8 utilizable ones; even after supplementing all corresponding exchange reactions, its accuracy remained poor at 41.56%. Furthermore, gapseq exhibited poor specificity (16.67%), indicating that reactions and pathways added by gapseq were associated with excessive FP, leading to poor substrate utilization specificity and functional deficiencies. In terms of model content, the AlphaGEM-constructed model annotated more genes (1588) compared to gapseq (1260) and iYL1228 (1228) (Fig. 5c). This expanded coverage is biologically consistent with genome size variations: the *K. pneumoniae* genome has 5728 genes, whereas the *Escherichia coli* genome consists of 4401 genes, and the *E. coli* model *i*ML1515 has 1515 annotated metabolic genes. The lower number of genes annotated by gapseq may be due to its less accurate gene function annotation method, as evidenced by the fact that gapseq identified 63 reactions during gap filling, while AlphaGEM only required 2 reactions to fill gaps, which suggests that the model constructed by gapseq suffers from insufficient gene annotation, resulting in more reactions required for gap filling. However, the AlphaGEM model contained fewer reactions (3068) and metabolites (2475) than the gapseq model (3207 reactions and 2577 metabolites), likely due to gapseq's tendency to incorporate entire pathways from databases such as MetaCyc, and the model constructed by gapseq exhibited an excessive number of genes corresponding to each reaction and an overabundance of reactions encoded by each gene, which results in the phenomenon of having fewer genes but more reactions (Supplementary Fig. 7a, b). This often introduces redundant reactions and metabolites, resulting in poor phenotype prediction, consistent with the phenomenon that gapseq-based GEM has a low specificity in predicting carbon source utilization (Fig. 5a). We further constructed GEMs for *B. subtilis* as the second case study and benchmarked it against models generated by gapseq and manual curation[43]. As GEMs for this species have not been extensively benchmarked for single carbon source utilization, we validated model performance in essential gene predictions. AlphaGEM demonstrated prediction accuracy higher than gapseq, CarveMe, and ModelSEED (AlphaGEM = 85.52%, gapseq = 79.35%, CarveMe = 83.09%, ModelSEED = 77.24%), though it still slightly lagged behind the manually curated GEM (accuracy = 89.26%) (Fig. 5b). Beyond predictive capabilities, the quality of models generated by AlphaGEM was evaluated using the MEMOTE[44] suite, demonstrating that these models are of high quality, consistent with their respective reference models (Supplementary Table 2). Furthermore, regarding computational efficiency, although AlphaGEM requires more running time (excluding the protein 3D structure prediction) than ModelSEED and CarveMe, it operates faster than gapseq (Supplementary Table 3). This moderate time trade-off is well-justified by its reliability, thus achieving a balance between computational efficiency and predictive accuracy. Together, these results indicate that AlphaGEM could generate GEMs comparable to manually curated ones and, to some extent, outperforms the existing toolboxes in bacterial modelling.

For mammals, after increasing the threshold for homologous relationship identification, we used the latest Human1 as the reference model to build GEMs for *M. musculus* and *C. griseus*, respectively, then compared the models' performance with previously published ones. AlphaGEM achieved comparable essential gene prediction accuracy compared to the existing *M. musculus* GEM—Mouse1[45] (model by AlphaGEM: F1-score = 4.32%, Matthews Correlation Coefficient (MCC) = 0.0 184, accuracy = 91.10%; Mouse1: F1-score = 4.71%, MCC = −0.0016, accuracy = 88.62%) (Fig. 5d). AlphaGEM's multi-scale protein annotation approach, leveraging structural data for homology-based annotation, resulted in 488 more annotated genes than Mouse1 (Fig. 5f and Supplementary Fig. 2e) while the numbers of reactions and metabolites from the two types of models were comparable. To evaluate the accuracy of our annotations, we used multiple deep-learning-based annotation tools to mine the function of proteins corresponding to the unique genes from each model. We applied a higher threshold to filter these proteins, and those annotated with reactions after filtering were considered reliable metabolic proteins, while others were considered unreliable metabolic proteins. We found that, among the differential genes in Mouse1, 9.4% encoded reliable metabolic proteins, while 64.5% of the genes in the GEM constructed by AlphaGEM encoded proteins considered reliable metabolic proteins (Supplementary Fig. 8a, b). Here, if we select only the genes encoding reliable proteins as the target GEM's genes, the number of genes in the constructed model (3103) will be similar to that of Human1 (2887), which further supports the validity of our approach. For models constructed for *C. griseus*, AlphaGEM demonstrated accuracy and specificity in essential gene prediction comparable to those of a manually curated model (accuracy 85.32% compared to 88.56%, specificity 98.52% compared to 99.29%). However, due to the limited number of essential genes identified through our approach, the F1-score performance was lower (10.43% compared to 21.09%) (Fig. 5e). For higher mammals, the complexity and abundance of intracellular pathways, coupled with relaxed flux constraints, can lead to essential reactions appearing as non-essential for growth, whereas the fact that some metabolic reactions are not being uncovered in the model would result in the opposite outcome. Adding additional constraints, such as transcriptomic data and enzymatic constraints, could improve prediction accuracy, further enhancing its usability for complex organisms[46]. Overall, these findings initially demonstrate that AlphaGEM could be leveraged to generate functional, comprehensive GEMs for mammals.

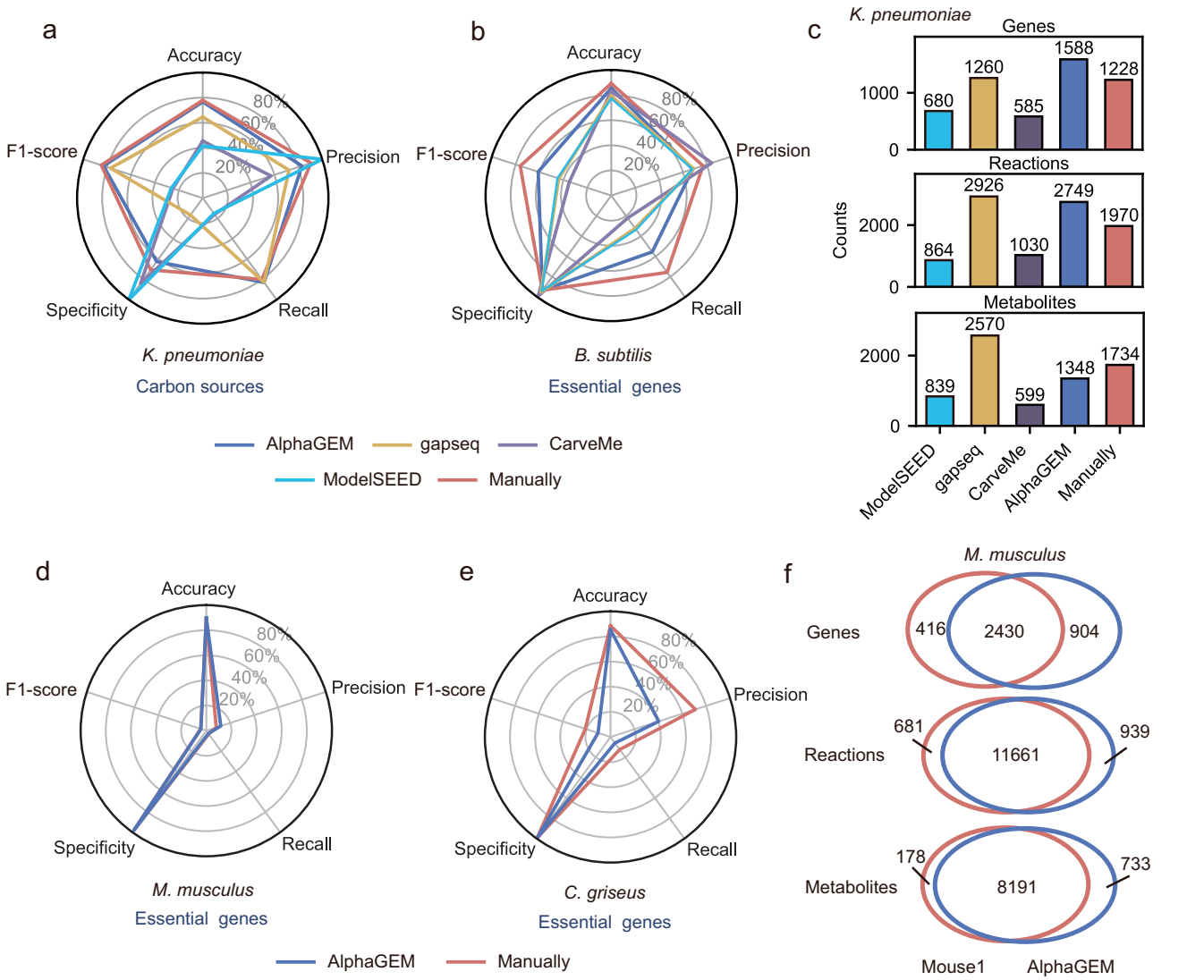

Fig. 5 | Quality assessment of GEMs for both bacteria and mammals reconstructed by AlphaGEM. Comparison of accuracy in predicting carbon source utilization by different GEMs from manual curation, CarveMe, ModelSEED, gapseq, and AlphaGEM for *K. pneumoniae* (**a**). Comparison of accuracy in essential gene prediction by different GEMs from manual curation, CarveMe, ModelSEED, gapseq, and AlphaGEM for *B. subtilis* (**b**). Comparisons of the number of genes, reactions, metabolites from different versions of GEMs for *K. pneumoniae* (**c**). Comparison of accuracy in essential gene prediction by different GEMs from manual curation and AlphaGEM for *M. musculus* (**d**) and *C. griseus* (**e**). Analysis of common and unique genes, reactions, and metabolites between two different GEMs from manual curation and AlphaGEM for *M. musculus* (**f**). Source data are provided as a Source Data file.

AlphaGEM accelerates the generation of GEMs at a large scale

The advancement of high-throughput sequencing technology has led to a large amount of genomic information from non-model organisms. Traditional procedures for generating GEMs for these non-model organisms often undergo multiple rounds of iterative upgrades, and require considerable time and resources[47]. In this aspect, AlphaGEM could enable high-throughput reconstruction of GEMs to bridge the gaps between cellular phenotypes and genotypes at a large scale. To confirm this, we leveraged AlphaGEM to reconstruct GEMs in an automated manner for 332 sequenced yeast species[48,49], which were previously modeled using manually curated GEMs[50]. By clustering the models using t-SNE based on their reaction content (Fig. 6a), we found that models from AlphaGEM exhibited distinct clustering patterns, in which the corresponding yeast species from the same order displayed comparable aggregation with manually curated ones, thus, to some extent, reflecting species classification based on phylogenetic analysis[48,49,51]. Statistical validation using the Kruskal–Wallis test strongly supports this observation, confirming significant clustering structure in AlphaGEM-generated models (x: P value = 1.00×10^{-23}, y: P value = 3.29×10^{-56}, z: P value = 4.46×10^{-37}) and those in manually created models (x: P value = 8.53×10^{-45}, y: P value = 7.94×10^{-53}, z: P value = 1.30×10^{-17}). This phylogenetic clustering was further quantitatively supported by a higher Silhouette score for AlphaGEM (0.138) compared to the manually curated models (0.103). Furthermore, models constructed by AlphaGEM exhibited greater hamming distances between reaction profiles compared to manually curated ones, indicating GEMs built by AlphaGEM could reflect species-specific metabolic differentiation (Fig. 6b).

We compared structural characteristics between AlphaGEM-generated and manually curated models. Linear correlations existed ($R^2 > 0.45$) between two types of models in the numbers of genes and metabolites (Supplementary Fig. 9a, b). While the number of reactions showed a comparatively weak correlation ($R^2 = 0.35$), AlphaGEM could identify more reactions for most yeast species (Supplementary Fig. 9c and f). These findings indicate that the models reconstructed by AlphaGEM could be comparable to manually curated counterparts. Further analysis of GEMs across 332 yeast species demonstrated remarkably similar core reaction sets (AlphaGEM: 1944 vs. manual:

https://doi.org/10.1038/s41467-026-75549-w

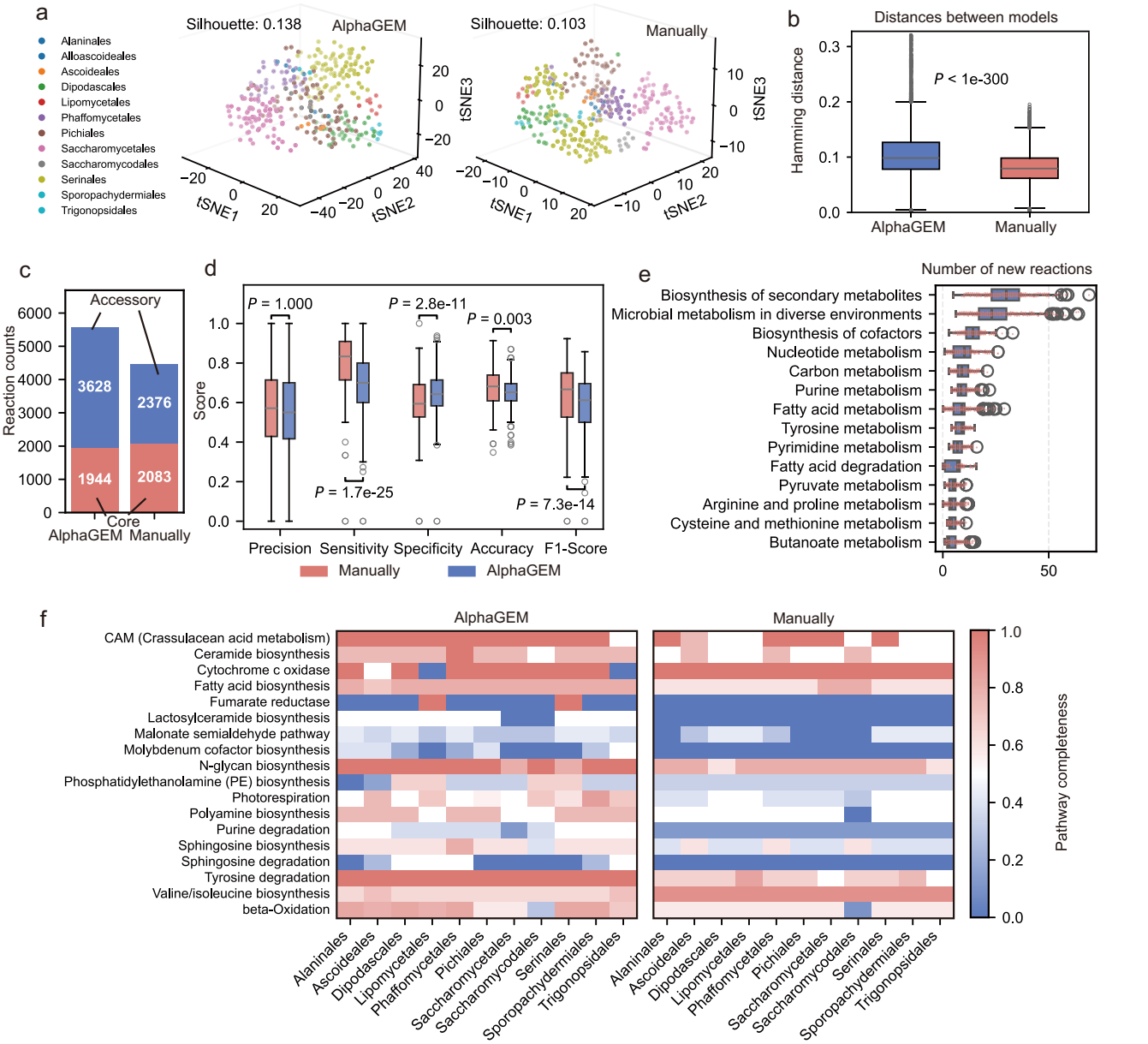

Fig. 6 | AlphaGEM accelerates high-throughput modelling for 332 distinct yeast species. The t-SNE clustering analysis of yeast species based on reaction presence in the manually curated GEMs and GEMs built by AlphaGEM, with different colors representing different orders (**a**). Comparison of Hamming distances between 332 yeast-species-specific GEMs constructed by AlphaGEM and manually curated GEMs (two-sided paired *t*-test; unadjusted) (**b**). Distribution of pan-, core-, and accessory-reactions across 332 yeast species (**c**). Validation in prediction of substrate utilization using GEMs constructed by AlphaGEM and the manually curated GEMs (two-sided paired *t*-test; Bonferroni-adjusted) (**d**). Main subpathways with newly added reactions that occurred only in GEMs generated by AlphaGEM compared to manually curated GEMs (**e**). Heatmap of pathway completeness between GEMs constructed by AlphaGEM and the manually curated GEMs, focusing on pathways with the largest mean inter-order differences. Color intensity indicates completeness values (red: 1.0; blue: 0) (**f**). Box plots show the median (center line), upper and lower quartiles (box limits), and 1.5× interquartile range (whiskers). Outliers are plotted as individual points. For all correlations and comparisons, $n = 332$ distinct yeast species were analyzed. Statistical details are provided in Supplementary Note 4. Source data are provided as a Source Data file.

2083) (Fig. 6c), confirming AlphaGEM's accuracy in identifying conserved yeast metabolic functions. Notably, AlphaGEM annotated more accessory reactions (3628 vs. 2376) (Fig. 6c), consistent with its enhanced species-specific metabolic differentiation observed earlier. This expanded accessory repertoire showcases AlphaGEM's potential to uncover broader metabolic pathway diversity across distinct species.

To validate the functional quality of the models, we performed substrate utilization predictions using the newly reconstructed GEMs and compared them with experimentally verified phenotypes (Fig. 6d). It shows that the GEMs built by AlphaGEM achieved comparable accuracy compared with manually constructed ones (Cohen's $d_z = -0.231254$). Notably, AlphaGEM models exhibited higher specificity (P value < 0.001), albeit with lower sensitivity and F1-score (P value < 0.001). However, because the manually constructed GEMs were explicitly refined using experimental data, they still exhibited higher accuracy for certain specific predictions.

Lastly, AlphaGEM's capability in uncovering dark metabolism was explored. Here, reactions uniquely annotated by AlphaGEM in each species were mapped onto metabolic pathways. We found that the novel reactions identified by AlphaGEM were primarily distributed in pathways related to secondary metabolism, microbial metabolism in

diverse environments, and cofactor biosynthesis (Fig. 6e). This distribution pattern demonstrates AlphaGEM's potential to elucidate untapped metabolic capabilities within secondary metabolic pathways. Furthermore, we evaluated the potential of AlphaGEM for uncovering gaps in metabolic pathways by analyzing pathway completeness. We then evaluated pathway completeness to pinpoint gaps in metabolic networks, concentrating on those pathways showing the largest inter-order differences between GEMs built by AlphaGEM and manually curated ones (Fig. 6f). In these pathways, AlphaGEM exhibited greater variability across orders, demonstrating its capacity to uncover dark metabolism in a order-specific manner by tailoring reaction sets to each lineage rather than applying a one-size-fits-all approach.

Discussion

In this work, we present AlphaGEM, an integrated toolbox for automated, high-throughput construction of genome-scale computational models for target organisms, uniquely addressing the critical gap in high-quality modelling of numerous newly sequenced species[52,53]. Our tool fundamentally differs from existing methods (e.g., gapseq[19]) by achieving two breakthrough capabilities: (1) high-throughput generation of eukaryotic GEMs with quality approaching manually curated benchmarks and (2) extension to prokaryotic modelling with performance comparable to or superior to state-of-the-art tools. This dual advantage stems from our integration of protein structural alignment and deep learning-based protein function inference, enabling precise detection of distant homologies widely present in understudied organisms. Importantly, we developed an ensemble algorithm that could expand metabolic network coverage for target organisms through the in-depth annotation of non-homologous proteins using multiple recent deep learning models. To demonstrate these capabilities, we systematically assess the performance of AlphaGEM in reconstruction of GEMs across diverse domains, including fungi (332 different yeast species), prokaryotes (e.g., *K. pneumoniae*), and mammalian cells (e.g., *M. musculus*), demonstrating AlphaGEM's excellent capability for large-scale, multi-species metabolic characterization and revealing its potential for uncovering dark metabolism in less characterized organisms through enhanced annotation of non-canonical enzymatic functions.

Unlike gapseq and CarveMe, which rely on a universal model as a template to generate GEMs for target species, our toolbox integrates the existing high-quality GEMs from reference organisms (i.e., *S. cerevisiae*, *E. coli*, *Homo sapiens*) alongside the universal reaction pool to mine the genome-level metabolic activities as comprehensively as possible. A key distinction of our approach lies in leveraging the evolutionary principle that protein structure is more conserved than protein sequence[54]. By employing structural alignment, AlphaGEM achieves high specificity in identifying distant homologies often overlooked by traditional sequence-based methods. Consequently, our pipeline could efficiently build a high-quality backbone of the reaction network for non-model organisms. Meanwhile, a comprehensive universal reaction pool was leveraged to annotate the remaining non-homologous proteins based on our ensemble pipeline, which effectively enlarges the scope of GEMs to cover more species-specific metabolic reactions. As a result, AlphaGEM exhibits highly versatile capabilities, capable of reconstructing both prokaryotic and eukaryotic GEMs and overcoming the domain-specific limitations of tools like gapseq[19] (prokaryote-oriented) and CarveFungi[39] (fungus-specific).

Although our strategy offers a robust framework for developing high-quality GEMs for non-model organisms, several challenges still remain. Firstly, generating all protein structures for one species requires substantial computational resources, which may lower the efficiency of our pipeline as a whole. Additionally, using protein-language-model-based approaches, i.e., PLMSearch, could identify homologous relationships on a large scale, but, to some extent, compromise accuracy at the same time. Moreover, the predictive accuracy of deep learning models employed in our study could directly influence AlphaGEM's effectiveness, as training datasets used in different deep learning models often contain inconsistencies, resulting in potential errors in assigning key reactions to specific proteins. Furthermore, while AlphaGEM successfully inherits subcellular compartment annotations from reference models, we excluded automated compartment prediction for novel reactions identified via our ensemble method due to the inadequate performance of current protein localization tools[55,56]. Similarly, given the limitations in annotating transporter locations, extracellular transporters are comprehensively sourced from reference models and the Rhea database, whereas intracellular transporters remain strictly dependent on reference models. The relatively limited performance on mammalian data reflects a systemic bottleneck in eukaryotic modeling, where functional redundancy and the discrepancy between generic models and tissue-specific biological contexts inherently constrain predictive sensitivity. Future advances in the development of advanced computational tools, the integration of multi-omics data[11], and high-quality datasets will be critical to address these limitations, thus further helping to enhance the overall performance of AlphaGEM. Additionally, future integration of other high-quality reaction databases, such as MetaCyc[57], could further improve the annotation capability for target species-specific reactions, thus further helping to enhance the overall performance of AlphaGEM.

In conclusion, AlphaGEM enables automated, high-quality metabolic modelling across diverse species. It paves the way for efficient, scalable, and holistic analysis of newly sequenced organisms, deepening our understanding of metabolic diversity in microbial and higher-order organisms on Earth and driving transformative progress in biological research.

Methods

Inference of homologous relationships

In this study, two methods were employed to infer protein homologous relationships between target and reference species as the initial step in reconstructing draft GEMs for target organisms.

Homologs based on structural alignment

Initially, OrthoFinder (v3.0.1b1)[31] was utilized to cluster proteins from annotated genome assemblies, generating pairwise protein similarity matrices. Then, protein structure alignment was conducted for each protein pair using US-align (v20241108)[58], yielding a structural similarity score (TM-score). The results were filtered based on the TM-score and the predicted protein structure quality score (pLDDT)[59]. Protein pairs with a TM-score exceeding the threshold (default: 0.5), coupled with an alignment coverage ≥ 0.8, were classified as exhibiting high structural similarity, while those with a pLDDT below the threshold (default: 0.7) were excluded from structure alignment due to low-quality structure predictions (Supplementary Fig. 2c). These filtered protein pairs constituted the first protein homology set (Protein Homology Set One).

Foldseek (v9.427df8a)[54] was employed to conduct a structure-based homology search, expanding the collection of homologous protein pairs. Homologous pairs identified through Foldseek-based alignment were screened using the following criteria: a TM-score ≥ 0.7, coverage ≥ 0.8, and probability = 1 (Supplementary Fig. 2a, b). Pairs with a pLDDT below the default threshold of 0.7 were excluded due to low-quality structure predictions. Metabolic proteins were categorized into transporters and non-transporters, with transporters further filtered using an identity threshold of 30%. To enhance the accuracy of identification of structure-based homologs, homologous pairs were clustered separately for each enzyme type. For clustering homologous proteins from the reference species, the DBSCAN algorithm (default

eps = 1) was applied[60]. Clusters were selected based on the highest combined TM-score and coverage values, ensuring all pairs met the established TM-score and coverage thresholds. This process generated the second protein homology set, referred to as Protein Homology Set Two. Finally, the two homology sets were combined to determine the final protein homologous relationships.

Homologs based on PLMSearch

The embeddings of the target strain and the reference strain genomes generated by a protein language model were integrated as input to predict homologous relationships using PLMSearch (2024.05.08)[24]. Based on a predefined SS-predictor threshold (default = 0.8), the results were filtered by blastp with coverage ≥ 0.6 and identity ≥ 30%. Subsequently, BBHs identified through global BLAST alignment performed by diamond (v2.1.10)[61] were added to the final set of homologs (with identity ≥ 60%, coverage ≥ 0.5).

Strategy to solve the conflict between the results from sequence and structure annotation

To robustly integrate sequence and structural data and resolve potential metric conflicts, AlphaGEM employed a structure-first prioritization logic. When sequence and structural similarity diverged (e.g., high sequence identity but sub-threshold TM-score), structural conservation was strictly prioritized, and such pairs were excluded to prevent the inclusion of functionally divergent paralogs. Notably, for the metabolic enzymes targeted in this study, our structure-first prioritization rarely, if ever, results in the false exclusion of genuine homologs that exhibit high sequence identity but pronounced structural divergence—such as those with high conformational flexibility (Supplementary Fig. 2d). Empirical validation shows that no correct homologs in our dataset occupied the high-identity/low-TM-score quadrant. Ultimately, this filtering process represents a strategic trade-off aimed at maximizing annotation precision while maintaining a robust functional baseline for automated GEM reconstruction. Conversely, in the remote homology zone where sequence identity was negligible, stringent structural criteria (e.g., TM-score ≥ 0.7) were imposed to confidently infer functional homology solely from structural evidence.

Furthermore, to account for the taxon-dependent differences in evolutionary conservation across diverse taxa[62–64], the structural alignment thresholds applied during this screening process were dynamically adjusted according to the target organism's taxonomic classification (Supplementary Table 1). This taxon-specific parameterization ensures a robust balance between sensitivity and specificity across varying evolutionary distances.

Homolog filtering using ProtTrans

Our algorithm identified several proteins with an unusually high number of corresponding homologous proteins (i.e., > 6) from the input organism (Supplementary Fig. 10b, c), which is biologically implausible. To address these issues, ProtTrans (ProtT5-XL-UniRef50), a protein language model[65], was utilized to encode the pre-identified homologous proteins. Principal Component Analysis (PCA) was then applied for dimensionality reduction to refine these homologous relationships ("Methods", Supplementary Fig. 10a). This approach effectively reduced the number of homologs for proteins with multiple relationships to fewer than three, minimizing FP (Supplementary Fig. 10b, c). This strategy to prune the number of homologous proteins could be used as an optional step for building GEMs with our toolbox. This approach proved critical for constructing the mouse GEM, successfully differentiating functions between Cytochrome P450 26B1 and Cytochrome P450 26C1 (Supplementary Fig. 10e), thereby enhancing essential protein prediction accuracy. The detailed method is as follows:

For each protein (RP) executing a specific reaction in the reference model and its homologous proteins (P1, P2,...), they were grouped together. If the group contains only one protein apart from RP, this step was skipped. For the prediction groups, ProtTrans was used to generate embeddings for the sequences of all proteins in the group. PCA was then applied to reduce the dimensionality of these embeddings to analyze the distances between different sequences. The distance between the reduced PCA coordinates was defined as:

$$r_{i,j} = \sqrt{\left(\mathrm{pca1}_i - \mathrm{pca1}_j\right)^2 + \left(\mathrm{pca2}_i - \mathrm{pca2}_j\right)^2} \quad (1)$$

$$r_0 = \min_{P \in [P1, P2, \ldots]}\left(r_{RP,P}\right) \quad (2)$$

$$r_t = \mu r_0 \quad (3)$$

Where i and j represent different proteins within the group. The minimum distance r_0 to the reference protein RP was selected, and a threshold r_t was applied (default $\mu = 1.1$). Homologous proteins with a distance greater than this threshold were removed, and the remaining group was considered the target homologous relationship.

Validation of homologous protein relationships

KO numbers were assigned to proteins based on homologous protein relationships using data from UniProt and KEGG[66]. Proteins sharing the same KO number were designated true homologs; those with different KO numbers were designated false homologs. Accuracy was calculated as the proportion of true homologs among all evaluated protein pairs (true homologs/(true homologs + false homologs)).

Selection of high-quality reference models

The latest published manually curated functional GEMs were selected as reference GEMs. For modelling eukaryotic single-cell organisms, the metabolic model of *S. cerevisiae*, Yeast9[14], was chosen as the reference model. Species with a closer phylogenetic relationship to *S. cerevisiae* can use this model as a reference. For prokaryotic cells, the *E. coli* model, *i*ML1515[15], was selected as the reference model, and species with a closer phylogenetic relationship to *E. coli* can also use it as a reference. The latest published *H. sapiens* GEM, Human1[16], was selected as the reference model for mammalian modelling. These reference models have been updated with metabolite and reaction IDs from the Rhea and ChEBI databases, using mapping relationships from MetaNetX[67] to ensure compatibility when adding new reactions. AlphaGEM further provides an interface for manual submission of custom reference models, which must be processed using the supplied curation script model_curation.py to align with framework requirements.

Draft model reconstruction for non-model organisms

The GEMs for target species were reconstructed based on the homologous relationships derived from the reference models. First, reactions lacking homologous proteins in the reference model were removed according to the gene-protein-reaction (GPR) relationships. Subsequently, the retained reactions were re-annotated through homologous mapping to generate draft models. The proteins were mapped to their encoding genes and added to the GEMs.

For isoenzyme-catalyzed reactions, the reaction was considered to exist if at least one corresponding isoenzyme was present. The reaction GPR was defined based on the isoenzymes present. For reactions catalyzed by protein complexes, the reaction was considered to be present if over 80% of the complex's proteins were identified, and the GPR for the reaction was reconstructed based on the complex. For

transporters, an optional stricter filtering threshold was offered for their identification.

Protein annotation through multiple computational tools

Proteins excluded from the homologous protein pairs were annotated using an ensemble method. EggNOG-mapper (v2.1.12)[28] was employed to annotate proteins with KO numbers, CLEAN (v1.0.1)[26], and DeepECtransformer (v1.0)[27] were used for EC number annotations, and UniProt protein data were used to obtain corresponding Rhea IDs. CLEAN annotations were filtered at a confidence threshold of 0.7 before downstream analysis. PLMSearch[24] identified homologous protein pairs and annotated target proteins with Rhea IDs by performing homology searches against UniProt's Rhea-annotated proteins. The results were filtered using a threshold determined by SSpredictor (default = 0.9) to obtain the final output. All annotation results were standardized to Rhea IDs by rhea.rdf for consistent protein function annotation.

To evaluate the accuracy of these computational tools, the standardized protein-to-reaction pairs were derived from the related Rhea annotations recorded in UniProt database, serving as the benchmark for evaluating protein–reaction pairs annotated by computational tools. Accuracy was quantified using: (1) true positives (TP): protein-reaction pairs annotated by the tool that were also present in the database annotations; (2) FP: protein-reaction pairs annotated by the tool that were not present in the database annotations; (3) false negatives (FN): protein-reaction pairs present in the database annotations that were not annotated by the tool. These metrics (TP, FP, FN) were then used to calculate precision, sensitivity, and F1-score. Notably, FP assessments exhibit inherent limitations due to incomplete reaction annotations in UniProt[68].

Guided by the aforementioned evaluations, and taking into account the varying annotation performance and inherent limitations of each individual tool, an ensemble weighting scheme was developed to systematically integrate the preliminary results (Fig. 3a and Supplementary Note 3). Within this scheme, distinct weight scores were assigned to the annotations from each tool, and a cumulative weight threshold was utilized to determine the final predictions. The specific weight assignments and target threshold were initially optimized based on annotation performance in *S. pombe* and *C. albicans*. To confirm robustness and prevent overfitting, this weighting strategy was rigorously validated by analyzing F1-score distributions across independent fungal species, including *R. toruloides* and *P. pastoris* (Supplementary Fig. 11a). Furthermore, the reliability of the selected parameters was verified by Receiver Operating Characteristic curve analyses, with the resulting Area Under the Curve values demonstrating consistent predictive capability across diverse taxa (Supplementary Fig. 11b). The selected parameters were further validated using *B. subtilis* and *M. musculus*, leveraging their curated UniProt annotations (Supplementary Fig. 11b). Notably, in the mammalian context, the observed trade-off favoring high precision over recall in *M. musculus* is justified, as it effectively minimizes false-positive annotations. Justified by this empirical validation, the final weight scores were established (EggNOG-mapper: 0.3, CLEAN: 0.2, DeepECtransformer: 0.2, PLMSearch: 0.1) (Fig. 4c), and candidate annotations achieving a cumulative weight score of ≥ 0.4 were ultimately selected as the final integrated predictions.

Extension of GEMs using the Rhea database

To accelerate reaction integration, biochemical reactions from Rhea, a standardized reaction-metabolite database, were retrieved. These reactions were converted into a pool of ready-to-use cobra.Reaction objects. All reaction and metabolite information was extracted from the rhea.rdf file, including the mapping relationships with other databases, reaction reversibility, metabolite composition, charge, and ChEBI IDs[69], to generate reaction classes that can be used with COBRApy. During this integration phase, strict quality control measures were implemented: the mass and charge balance of the supplementary reactions were explicitly verified to prevent the introduction of new stoichiometrically unbalanced reactions (e.g., artificial generation of metabolites or ATP). Reactions were subsequently added to the model based on prior Rhea ID annotations for genes. It should be noted that while this integration pipeline successfully prevents the integration-induced errors, the final reconstructed models may still inherit pre-existing stoichiometric artifacts from their chosen reference models.

Compartmentalization and subcellular annotation

Reactions and associated metabolites were assigned to specific subcellular compartments based on the reference models. As explicit subcellular localization is generally absent in the Rhea database, newly integrated reactions from Rhea were assigned to the cytosol by default. For transport reactions, extracellular transporters (mediating exchange between the cytosol and the extracellular environment) were identified and incorporated using both the reference models and the Rhea database. Conversely, inter-compartmental transporters (mediating exchange between intracellular organelles) were reconstructed exclusively based on the reference models.

Subsystem annotation

In the AlphaGEM reconstructions, the assignment of subsystem information was executed via a dual strategy. First, subsystem annotations for a subset of reactions were directly inherited from the selected reference models. Furthermore, for metabolic reactions possessing KEGG[66] annotations, the relevant subsystem categorizations were retrieved and integrated directly from the database.

Improving the quality of GEMs using gap-filling

As the final refinement step, AlphaGEM employed a sequence-based gap-filling strategy to ensure both metabolic functionality and biological relevance for the newly added reactions. Initial gap-filling was performed using the gap-filling function cobra.flux_analysis.gapfill() in COBRApy (v0.29.1)[17]. During the gap-filling process, the reference model or the combined one, which merged the reference model and the universal model together, could be selected depending on the users' demands. Gap-filling was optionally performed using either minimal or complete growth media. At the same time, the objective reactions (e.g., biomass production) and the biomass composition were consistent with reference GEMs. To further refine and filter potential gap reactions identified through gap-filling, AlphaGEM performed tblastn (v2.16.0) searches against the genome of the target organism using the proteins catalyzing these gap reactions. Homology-based replacement was then applied: for hits with a bit score > 200, the protein corresponding to the hit with the highest bit score was used to replace the original homologous protein in the reaction. If no qualifying hit (bit score > 200) was found, AlphaGEM attempted to remove the reaction from the model, while ensuring the integrity of the objective reaction. However, if removing a reaction without sequence support would seriously weaken the model's prediction of cellular growth, it was retained as a purely algorithmic necessity. In our tested cases, such reactions added purely as algorithmic necessities ranged from 0 to 14 across species (e.g., only 0 reactions for *C. albicans*). As a result, out of the 3–19 total reactions added, gap-filling further complemented GEMs by filling essential pathway gaps (e.g., three reactions for *C. albicans*).

Hybrid strategy for cross-species model reconstruction

A hybrid reconstruction strategy was implemented by integrating the above processes. To accommodate the metabolic diversity across fungi, bacteria, and mammals, a hybrid reconstruction strategy was implemented. Instead of relying solely on a single reference organism,

a high-quality reference model was integrated with a comprehensive pool of reactions identified through AI-based functional annotation. For genes lacking direct homology to the reference, functional annotations were supplemented by this reaction pool to ensure the capture of species-specific metabolic pathways. To mitigate the limitations of sequence-based ortholog inference across evolutionarily distant taxa, a structure alignment-based strategy was employed. This approach enabled the identification of distant homologous relationships, allowing for the effective use of manually curated reference backbones even when phylogenetic proximity was limited. By combining a refined reference backbone with a supplemental reaction pool, the pipeline ensures robust and accurate metabolic reconstructions across diverse taxonomic groups.

Essential gene prediction

To conduct gene essentiality analysis, the essential gene list from published studies[34,37,38,43,45,70,71] was used. To accurately reflect the experimental conditions, the in silico medium was formulated based on the standard M9 minimal medium. The base medium was supplemented with additional non-carbon compounds to mirror the specific experimental media described in the respective literature. The lower bounds (maximum uptake rates) for the exchange reactions of these supplementary compounds were set to −1000 mmol/gDW·h to ensure they were not growth-limiting. In silico growth after gene knockout was calculated as follows: (1) Each gene in the model was iteratively knocked out using the model.genes.knock_out() function. (2) Growth rates were simulated using FBA under identical minimal glucose medium conditions. (3) A gene was classified as computationally essential if the simulated growth rate was less than $1 \times 10^{-4}\ h^{-1}$.

Predictions were compared against the experimental essentiality datasets. To ensure the reliability of the comparison, only genes present in both the experimental datasets and the metabolic model were included in the evaluation. The results were categorized using a confusion matrix as follows: (1) true positive (TP): experimental non-essential/in silico non-essential; (2) true negative (TN): experimental essential/in silico essential; (3) false negative (FN): experimental non-essential/in silico essential; (4) FP: experimental essential/in silico non-essential (essential genes with low in silico growth were calculated as negative). The specific statistics and details of the evaluated datasets are summarized (Supplementary Table 4 and Supplementary Fig. 12).

Carbon source utilization prediction

To conduct carbon source utilization analysis, the phenotype data from published studies[37,38,42,45,72] were used. In silico growth with carbon source utilization was calculated as follows: (1) The model medium was set to the minimal growth medium with glucose replaced by alternative carbon sources; that is, the lower bound for the exchange reaction of the alternative carbon source was constrained to −10 mmol/gDW·h, while other environmental constraints were set to simulate aerobic M9 minimal medium. (2) Growth rates were simulated using FBA. (3) A carbon source was classified as non-utilization if the simulated growth rate was less than $1 \times 10^{-3}\ h^{-1}$.

Predictions of carbon source utilization were evaluated using a confusion matrix framework, with carbon source utilization defined as positive and non-utilization defined as negative. True positive (TP) was defined as experimental utilization/in silico utilization; TN as experimental non-utilization/in silico non-utilization; false negative (FN) as experimental utilization/in silico non-utilization; and FP as experimental non-utilization/in silico utilization.

Metrics to assess the prediction performance of GEMs

To evaluate model performance, the following metrics were computed based on the confusion matrix, consisting of TP, TN, FP, and FN.

MCC[73]: The MCC, a balanced measure of classification quality, was computed as:

$$\mathrm{MCC} = \frac{(\mathrm{TP} \cdot \mathrm{TN}) - (\mathrm{FP} \cdot \mathrm{FN})}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}} \tag{4}$$

Sensitivity (Recall)[74]: Sensitivity, the ability to correctly identify positive samples, was computed as:

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{5}$$

Accuracy[74]: Accuracy, the proportion of correctly classified samples, was calculated as:

$$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{6}$$

Specificity[74]: Specificity, the ability to correctly identify negative samples, was computed as:

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{7}$$

Precision[74]: Precision, the proportion of TP among predicted positives, was calculated as:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{8}$$

F1-Score[74]: The F1-score, the harmonic mean of precision and sensitivity, was calculated as:

$$\mathrm{F1-score} = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Sensitivity}}{\mathrm{Precision}+\mathrm{Sensitivity}} \tag{9}$$

Ablating structural features for sequence-only alignment

For this ablation study, the results from OrthoFinder were directly used as the final protein homologous relationships. All other steps in the pipeline were kept unchanged. Then, the generated GEMs were evaluated for carbon source utilization. Finally, these results were compared with GEMs constructed using the standard AlphaGEM.

Construction of GEMs using other methods

GEMs were reconstructed starting from protein sequences retrieved from UniProt. For prokaryotic species, including *K. pneumoniae* and *B. subtilis*, models were constructed using CarveMe (v1.6.6)[18], ModelSEEDpy (v0.4.2)[10] and gapseq (v1.4.0)[19] under their respective default configurations with minimal medium. For eukaryotic species, including *S. pombe*, *C. albicans*, *P. pastoris*, and *R. toruloides*, GEMs were constructed using CarveFungi (2024.03.21)[39]. To ensure reproducibility and scalability, an automated reconstruction script was developed based on the official Tutorial-Model_reconstruction.ipynb provided by the CarveFungi framework. All reconstructions were performed using minimal medium settings to maintain consistency across the benchmark. The benchmark was performed on a local standard server (CPU: AMD Ryzen 9950x, GPU: NVIDIA GeForce RTX 5080).

Construction of GEMs for 332 yeast species

We constructed GEMs for the annotated yeast genomes using AlphaGEM's PLMSearch method under default parameters. Gap-filling was performed using a minimal medium formulation. To evaluate the metabolic capacity of each strain, substrate utilization was predicted

across a panel of 23 distinct carbon and nitrogen sources for 232 species with available phenotype data, following the procedure described in the preceding section. The resulting models were analyzed directly without modification. For dimensionality reduction, t-SNE (t-Distributed Stochastic Neighbor Embedding) was applied using the SciPy library (v1.14.1)[75]. This cluster analysis was conducted based on reaction presence/absence profiles.

Comparison of inter-order pathway completeness differences

Reactions from AlphaGEM-constructed GEMs and manually constructed GEMs were extracted and mapped to KEGG-defined pathways. The pathway completeness for a single species is calculated as:

$$\text{PathwayCompleteness}_{\text{species}} = \frac{N(\text{ModelReactions} \cap \text{Pathway})}{N(\text{Pathway})} \quad (10)$$

The order-level assessment is defined as:

$$\text{PathwayCompleteness}_{\text{Order}} = \text{median}_{\text{species} \in \text{Order}} \left(\text{PathwayCompleteness}_{\text{species}}\right) \quad (11)$$

Where N denotes the cardinality of the reaction set, and median is computed over all valid species annotations in the target order.

Statistical analysis

All statistical tests were two-sided unless otherwise stated. Exact sample sizes, test statistics, degrees of freedom, confidence intervals, effect sizes and *P* values for the main statistical comparisons are reported in Supplementary Note 4 and in the Source Data file. *P* values smaller than the numerical precision of the statistical software were reported as *P* value < 10^{-300}. The statistical analyses were performed using Python (v3.10.15) with the SciPy and Scikit-learn (v1.5.2) libraries.

Welch's *t*-test for independent groups

For comparisons between two independent groups, Welch's *t*-test was used unless equal variance was explicitly assumed. The Welch's *t*-test statistic was calculated as:

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \quad (12)$$

where $\bar{X}_1$ and $\bar{X}_2$ are the group means, s_1^2 and s_2^2 are the sample variances, and n_1 and n_2 are the sample sizes. The degrees of freedom were estimated using the Welch-Satterthwaite approximation. Hedges' *g* was used as the effect size for independent-group comparisons.

Paired *t*-test for matched comparisons

For paired comparisons, paired two-sided t-tests were used when measurements were matched by species, model pair or pairwise species comparison. The paired *t* statistic was calculated from the vector of paired differences:

$$t = \frac{\bar{d}}{\left(\frac{\text{sd}}{\sqrt{n}}\right)} \quad (13)$$

where $\bar{d}$ is the mean paired difference, sd is the standard deviation of paired differences, and n is the number of paired observations. Cohen's d_z was calculated as the mean paired difference divided by the standard deviation of paired differences. Mean paired differences were reported with 95% confidence intervals.

Pearson and Spearman correlation analyses

For correlation analyses, Pearson correlation tests were used to evaluate linear associations between paired model statistics. Pearson's correlation coefficient r was used as the test statistic and effect-size estimate. The 95% confidence intervals for Pearson's *r* were calculated using Fisher's *z* transformation. Spearman's rank correlation was additionally calculated as a non-parametric sensitivity analysis.

Kruskal–Wallis test

For comparisons across more than two orders in the t-SNE analyses, Kruskal–Wallis tests were performed on each target t-SNE dimension (x, y or z coordinates) across orders. The Kruskal–Wallis statistic *H* was calculated as:

$$H = \left[\frac{12}{(N(N+1))}\right] \sum_{i}^{k} \left(\frac{R_i^2}{n_i}\right) - 3(N+1) \quad (14)$$

where *N* is the total number of samples, *k* is the number of orders, n_i is the number of species in the *i*-th order, and R_i is the sum of ranks for the *i*-th order.

Multiple-comparison correction and statistical reporting

Where multiple primary tests were performed within the same figure panel group, *P* values were adjusted using the Bonferroni method by multiplying the raw *P* value by the number of primary comparisons and capping the adjusted value at 1. Bonferroni correction was applied across five carbon-source performance metrics in Fig. 6d, across two panel-wise comparisons in Supplementary Fig. 7, and across six primary tests in Supplementary Fig. 9. No multiple-comparison adjustment was applied to Fig. 6b because it involved a single planned comparison.

Box plots show the median, upper and lower quartiles, and 1.5 times the interquartile range; outliers are plotted as individual points where shown. Bar charts and box plots are accompanied by the underlying numerical values in the Source Data file.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data are provided with this paper. The source data underlying the main and Supplementary Figures are provided as a Source Data file. Additional data for model validation can be found in the Supplementary Information and Supplementary Tables. The processed analysis data generated in this study, including figure-generation code, processed data tables, statistical analysis outputs, model-comparison results, species mapping files, reaction/module completeness tables, and intermediate analysis outputs, have been deposited in the GitHub repository https://github.com/HanWeishang/AlphaGEM-figures. The reconstructed GEMs generated in this study have been deposited in the GitHub repository https://github.com/HanWeishang/AlphaGEM-GEMs-reconstructed. The software dependency data required to run AlphaGEM have been deposited in Figshare under https://doi.org/10.6084/m9.figshare.30630524. The data are available upon request from the corresponding author. The yeast genome data used in this study were sourced from the Y1000+ project[48,49], and manually curated GEMs for 332 yeast species were obtained from https://github.com/sysbiochalmers/yeast-species-gems.[50] Genomic data for other species were sourced from standard datasets in UniProt[68], and the protein function and homology data used for validation were obtained from UniProt and KEGG[66] as well. Source data are provided with this paper.

Code availability

The source code for AlphaGEM is available on GitHub at https://github.com/hongzhonglu/AlphaGEM. The analysis code is available on GitHub at https://github.com/HanWeishang/AlphaGEM-figures.

References

1. Oberhardt, M. A., Palsson, B. Ø & Papin, J. A. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* **5**, 320 (2009).
2. Steuer, R. Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry* **68**, 2139–2151 (2007).
3. Fang, X., Lloyd, C. J. & Palsson, B. O. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* **18**, 731–743 (2020).
4. Nielsen, J. & Petranovic, D. Modeling for understanding and engineering metabolism. *QRB Discov.* **6**, e11 (2025).
5. Stelling, J. et al. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Comput. Biol.* **9**, e1003091 (2013).
6. Lularevic, M., Racher, A. J., Jaques, C. & Kiparissides, A. Improving the accuracy of flux balance analysis through the implementation of carbon availability constraints for intracellular reactions. *Biotechnol. Bioeng.* **116**, 2339–2352 (2019).
7. Duarte, N. C. et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* **104**, 1777–1782 (2007).
8. Zomorrodi, A. R., Islam, M. M. & Maranas, C. D. d-OptCom: dynamic multi-level and multi-objective metabolic modeling of microbial communities. *ACS Synth. Biol.* **3**, 247–257 (2014).
9. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 121 (2019).
10. Henry, C. S. et al. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* **28**, 977–982 (2010).
11. Agren, R. et al. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* **10**, 721 (2014).
12. Mardinoglu, A. et al. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083 (2014).
13. Mukherjee, S. et al. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res.* **47**, D649–D659 (2019).
14. Zhang, C. et al. Yeast9: a consensus genome-scale metabolic model for *S. cerevisiae* curated by the community. *Mol. Syst. Biol.* **20**, 1134–1150 (2024).
15. Monk, J. M. et al. *i*ML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **35**, 904–908 (2017).
16. Robinson, J. L. et al. An atlas of human metabolism. *Sci. Signal* **13**, eaaz1482 (2020).
17. Heirendt, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).
18. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
19. Zimmermann, J., Kaleta, C. & Waschina, S. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol.* **22**, 81 (2021).
20. Capela, J. et al. *merlin*, an improved framework for the reconstruction of high-quality genome-scale metabolic models. *Nucleic Acids Res.* **50**, 6052–6066 (2022).
21. Hanemaaijer, M., Olivier, B. G., Röling, W. F. M., Bruggeman, F. J. & Teusink, B. Model-based quantification of metabolic interactions from dynamic microbial-community data. *PLoS ONE* **12**, e0173183 (2017).
22. Aite, M. et al. Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models. *PLoS Comput. Biol.* **14**, e1006146 (2018).
23. Mendoza, S. N., Olivier, B. G., Molenaar, D. & Teusink, B. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.* **20**, 158 (2019).
24. Liu, W. et al. PLMSearch: protein language model powers accurate and fast sequence search for remote homology. *Nat. Commun.* **15**, 2775 (2024).
25. Radivojac, P. et al. A large-scale evaluation of computational protein function prediction. *Nat. Methods* **10**, 221–227 (2013).
26. Yu, T. et al. Enzyme function prediction using contrastive learning. *Science* **379**, 1358–1363 (2023).
27. Kim, G. B. et al. Functional annotation of enzyme-encoding genes using deep learning with transformer layers. *Nat. Commun.* **14**, 7370 (2023).
28. Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
29. Potter, S. C. et al. HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
30. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
31. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
32. Koonin, E. V. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **39**, 309–338 (2005).
33. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823–826 (1986).
34. Tiukova, I. A., Prigent, S., Nielsen, J., Sandgren, M. & Kerkhoven, E. J. Genome-scale model of *Rhodotorula toruloides* metabolism. *Biotechnol. Bioeng.* **116**, 3396–3408 (2019).
35. Kim, J. et al. Multi-omics driven metabolic network reconstruction and analysis of lignocellulosic carbon utilization in *Rhodosporidium toruloides*. *Front. Bioeng. Biotechnol.* **8**, 612832 (2021).
36. Yaegashi, J. et al. *Rhodosporidium toruloides*: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotechnol. Biofuels* **10**, 241 (2017).
37. Viana, R. et al. Genome-scale metabolic model of the human pathogen *Candida albicans*: a promising platform for drug target prediction. *J. Fungi* **6**, 171 (2020).
38. Grigaitis, P. et al. A computational toolbox to investigate the metabolic potential and resource allocation in fission yeast. *mSystems* **7**, e0042322 (2022).
39. Castillo, S., Peddinti, G., Blomberg, P. & Jouhten, P. Reconstruction of compartmentalized genome-scale metabolic models using deep learning for over 800 fungi. Preprint at https://www.biorxiv.org/content/10.1101/2023.08.23.554328v1 (2023).
40. Mayer, F. L., Wilson, D. & Hube, B. *Candida albicans* pathogenicity mechanisms. *Virulence* **4**, 119–128 (2013).
41. Butler, G. et al. Evolution of pathogenicity and sexual reproduction in eight Candida genomes. *Nature* **459**, 657–662 (2009).
42. Liao, Y.-C. et al. An experimentally validated genome-scale metabolic reconstruction of *Klebsiella pneumoniae* MGH 78578, *i*YL1228. *J. Bacteriol.* **193**, 1710–1717 (2011).
43. Bi, X. et al. etiBsu1209: a comprehensive multiscale metabolic model for Bacillus subtilis. *Biotechnol. Bioeng.* **120**, 1623–1639 (2023).
44. Lieven, C. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).

45. Wang, H. et al. Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proc. Natl. Acad. Sci. USA* **118**, e2102344118 (2021).
46. Chen, Y. et al. Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO Toolbox 3.0. *Nat. Protoc.* **19**, 629–667 (2024).
47. Rawat, M., Chauhan, M. & Pandey, A. Extremophiles and their expanding biotechnological applications. *Arch. Microbiol.* **206**, 247 (2024).
48. Shen, X.-X. et al. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**, 1533–1545.e20 (2018).
49. Opulente, D. A. et al. Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts. *Science* **384**, eadj4503 (2024).
50. Lu, H. et al. Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection. *Mol. Syst. Biol.* **17**, e10427 (2021).
51. Hillis, D. M. Phylogenetic analysis. *Curr. Biol.* **7**, R129–R131 (1997).
52. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
53. McAlister, J. S. et al. An interdisciplinary perspective of the built-environment microbiome. *FEMS Microbiol. Ecol.* **101**, fiae166 (2025).
54. van Kempen, M. et al. Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).
55. Xiao, H., Zou, Y., Wang, J. & Wan, S. A review for artificial intelligence based protein subcellular localization. *Biomolecules* **14**, 409 (2024).
56. Jiang, Y., Wang, D., Wang, W. & Xu, D. Computational methods for protein localization prediction. *Comput. Struct. Biotechnol. J.* **19**, 5834–5844 (2021).
57. Karp, P. D. et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **20**, 1085–1093 (2019).
58. Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **19**, 1109–1115 (2022).
59. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
60. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (eds Simoudis, E., Han, J. & Fayyad, U.) 226-231 (AAAI Press, 1996).
61. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
62. Pál, C., Papp, B. & Lercher, M. J. An integrated view of protein evolution. *Nat. Rev. Genet* **7**, 337–348 (2006).
63. Waterhouse, R. M., Zdobnov, E. M. & Kriventseva, E. V. Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. *Genome Biol. Evol.* **3**, 75–86 (2010).
64. Slodkowicz, G. & Goldman, N. Integrated structural and evolutionary analysis reveals common mechanisms underlying adaptive evolution in mammals. *Proc. Natl. Acad. Sci. USA* **117**, 5977–5986 (2020).
65. Elnaggar, A. et al. ProtTrans: toward understanding the language of life through self-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 7112–7127 (2022).
66. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
67. Moretti, S., Tran, V. D. T., Mehl, F., Ibberson, M. & Pagni, M. MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Res.* **49**, D570–D574 (2021).
68. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617 (2025).
69. Hastings, J. et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219 (2016).
70. Strain, B., Morrissey, J., Antonakoudis, A. & Kontoravdi, C. How reliable are Chinese hamster ovary (CHO) cell genome-scale metabolic models? *Biotechnol. Bioeng.* **120**, 2460–2478 (2023).
71. Zhu, J. et al. Genome-wide determination of gene essentiality by transposon insertion sequencing in yeast *Pichia pastoris*. *Sci. Rep.* **8**, 10223 (2018).
72. Raja, W. A. & Çalık, P. Biochemical production from sustainable carbon sources by *Komagataella phaffii*. *Biochem. Eng. J.* **219**, 109702 (2025).
73. Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **405**, 442–451 (1975).
74. Fisher, R. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936).
75. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

Acknowledgements

We are grateful to the Center for High Performance Computing (HPC) of Shanghai Jiao Tong University for its support.

Author contributions

H.L. conceived the study. W.H., L.X., and H.S. contributed to the development of AlphaGEM. H.L., W.H., L.X., H.S., G.X., Q.J., H.W., B.J., C.Z., E.J.K., and J.N. read, edited, and approved the final manuscript.

Funding

This work is supported by grant 25HC2810600 and 24HC2810800 from Shanghai Municipal Science and Technology Program, grant 2022YFA0913000 from the National Key R&D Program of China, Shanghai Municipal Science and Technology Major Project, and grant 22208211 and 22378263 from the National Natural Science Foundation of China (NSFC). The funding body has no role in the design of the study, analysis and interpretation of the data, preparation of the manuscript, and decision to submit the manuscript for publication.

Competing interests

H.L., W.H., L.X., and H.S. filed two Chinese patent applications (202410340910.8 and 202411938706.2) based on this work. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-75549-w.

Correspondence and requests for materials should be addressed to Hongzhong Lu.

Peer review information *Nature Communications* thanks Christoph Kaleta and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.