

Materials and methods

Classification of macrofungi

We compiled a hierarchical classification of all macrofungal genera up to the phylum level. For Ascomycota, we adopted the latest edition of the *Outline of Fungi and fungus-like taxa* (Hyde et al. 2024) as the primary reference. For Basidiomycota, we followed the comprehensive classification proposed by He et al. (2024). It is important to note that lichenized fungi are not included in the macrofungal outline presented here. For the purposes of this study, genera containing both macrofungal and microfungal species are nevertheless treated as macrofungal genera.

Construction of MTSEM database

Multidimensional data curation

The Macrofungal Taxonomic System and Economic Mushrooms (MTSEM) database was constructed to provide a comprehensive, continually updated classification system for macrofungi, integrating both Ascomycota and Basidiomycota. To link macrofungal classification, nomenclature and important information that critical for basic and applied research, we curated and integrated classification, species, taxon nomenclature status, and metadata of specimens, strains, genomes, herbaria, culture collections and scientific publications. Data are sourced from authoritative databases including the Fungal Names (FN, Wang et al. 2023, accessed on 4 August 2025), the Global Catalogue of Microorganisms (GCM, Fan et al. 2025, accessed on 2 July 2025), the Culture Collection Information Worldwide (CCINFO, <https://ccinfo.wdcm.org/>, accessed on 2 July 2025), the Analyzer of Bio-resource Citations (ABC, <https://abc.wdcm.org/>, accessed on 2 July 2025), the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>, accessed on 6 August 2025), and the Joint Genome Institute (JGI, Grigoriev et al. 2014, accessed on 11 August 2025).

Species curation

All genera from Ascomycota and Basidiomycota were manually checked with FN database to verify their nomenclature status and unique registration identifier with the help of Name Standardization Tools in FN (<https://nmdc.cn/fungalnames/standardization>). All species and registration identifier of current names within those genera were collected from FN to ensure traceability and interoperability across databases and reflect the most recent taxonomic revisions and nomenclatural changes. Species and genera with nomenclature status discrepancies were subjected to a rigorous manual reconciliation process against the latest literature and were retained only upon confirmation of their validity.

Specimen, strains and genomes curation

Guided by the curated species list, we integrated multidimensional data. Metadata for specimens and strains, encompassing accession numbers and preservation institutions, were aggregated from GCM and FN, as well as all available and not available genomic sequencing data from NCBI and JGI. For NCBI, our retrieval spanned all available genomic sequences, incorporating both reference and non-

reference genomes. The extracted data included assembly accession, organism, size, assembly level, GC content, and submitter information. For JGI, we collected all fungal genomes, including those marked as ‘not available’. A critical deduplication step was implemented to exclude redundant genomic data not present on our curated species list or had already been submitted to NCBI. Hyperlinks to NCBI, JGI, and the Global Catalogue of Fungal Genomes (gcFungi, Liang et al. 2023) are provided for each genome to enable users to access comprehensive details.

Herbaria and culture collections

A comprehensive inventory of herbaria and culture collections information housing the species in this study was compiled from the CCINFO and FN database. For each institution, we extracted its standard acronym, full name, country, inventory of relevant preserved species, associated strains and specimens in this study. Key external links to detail pages in FN, CCINFO and GCM were preserved to facilitate user-side deeper exploration.

Economic mushroom curation

We collected information on the edibility, toxicity and medicinal value for each taxon from publications (Wu et al. 2019, 2026; Li et al. 2021; He et al. 2022; Bau et al. 2024; Wu & Yang 2026) followed by manual verification. Each record is tagged with DOI or patent number and voucher strain (where available).

Taxonomy harmonization

A critical taxonomic harmonization pipeline was implemented to resolve discrepancies arising from heterogeneous data sources. This involved mapping all synonymous names to their current accepted names as per the FN database. Complex cases requiring expert judgment were resolved through manual curation against latest literature.

Construction of the Hierarchical Taxonomic Visualization

To create the interactive taxonomic tree, species data were organized within a standardized phylogenetic framework based on recent phylogenomic studies. For each taxonomic node (from kingdom to species), metadata on sequencing status, economic traits, nomenclatural status, and species richness were programmatically annotated. This structured data layer drives the dynamic rendering of the visualization interface.

Research output collection and cooperative network

To map the research productivity and collaborative landscape in mycology, data were programmatically retrieved and integrated from the ABC database. All analyses were performed via structured graph queries to extract and harmonize semantic triples linking publications, patents, genome records, authors, institutions, and taxonomic entities. The processed data underpin two analytical modules on the platform, as described below.

Temporal trends of research outputs. A quantitative time-series analysis was conducted to track the output of different research modalities for each species. All associated scholarly articles, patents, and publicly available genome records for a given species were extracted. The publication year, patent application year, or genome release year was harmonized for each record. Annual counts for each of these three output types

(article, patent, genome) were then calculated to form a continuous time series for trend assessment.

Overview knowledge graph. To generate the overview knowledge graph, the fifteen genera with the highest total publication counts and the one hundred most-published authors across these top genera were selected. The co-authorship relationships and author-genus attribution data among this filtered set constituted the input for the graph. Collaboration networks were constructed at two levels using a standardized filtering approach. For any genus, its network was derived from (a) the ten species within it with the highest publication counts, and (b) the 50 authors with the most publications on that genus. For any species, its dedicated network was generated from the fifty authors with the most publications specifically on that species.

Database architecture and dynamic updates

The database was developed using a core stack comprising Spring Boot (<https://spring.io/projects/spring-boot>) for the modular back-end, Vue2 (<https://v2.vuejs.org/index.html>) with Element-UI (Element-A Desktop UI Toolkit for Web) for the fluid interactive experience, MySQL (<https://www.mysql.com/>) for metadata storage, and Neo4j (<https://neo4j.com/>) for graph-based knowledge retrieval. This architecture supports a professional database that offers an intuitive interface, visualized data, and user-friendly operation.

Genomic and evolutionary analyses

Phylogenomic analyses

All publicly available fungal sequences used in this study were retrieved from GenBank. Genomes from representative genus of subkingdom Dikarya were selected for analysis (Supplementary Table S3). Genome completeness was assessed using BUSCO v5.2.0 (Simão et al. 2015) with default parameters, based on the presence or absence of 758 predefined single-copy orthologs from the fungi_odb10 database. For each BUSCO gene, consensus protein sequences were searched against genomes using tBLASTn, candidate regions were predicted with AUGUSTUS v3.2.2 (Stanke & Waack 2003), and resulting sequences were compared to the corresponding BUSCO HMM profiles. BUSCOs were classified as complete (single-copy or duplicated), fragmented (<95% of reference length), or missing. To construct the phylogenomic data matrix, we used the full-length, single-copy BUSCO genes from representatives of Ascomycota, Basidiomycota, and nine designated outgroups. Each of the 758 genes was individually aligned using MAFFT v7.490 (Katoh & Standley 2013) with default settings. Ambiguously aligned regions were trimmed using trimAl v1.4 (Capella-Gutiérrez et al. 2009) with the gappyout option. Amino acid alignments of genes with >70% taxon occupancy were concatenated into a final supermatrix. Phylogenetic analyses were conducted using IQ-TREE v2.0.3 (Minh et al. 2020), a tool demonstrated to perform reliably in maximum likelihood (ML) frameworks for phylogenomic datasets. The phylogenomic tree was visualized using iTOL v6 (Letunic & Bork 2024).

Comparative genomics

Genome size, GC content, and gene numbers were obtained from standardized annotations, with gene models predicted using GeneMark-ES version 4.71_lic

(Besemer & Borodovsky 2005). Horizontally transferred genes were identified using a combined similarity- and phylogeny-based approach, incorporating DIAMOND searches against non-fungal databases, Alien Index filtering, and phylogenetic validation with IQ-TREE following established procedures (Liu et al. 2025). Biosynthetic gene clusters (BGCs) were predicted using the fungal module of antiSMASH (Blin et al. 2021). Differences between macrofungi and microfungi in genome size, GC content, gene number, HGT abundance, and BGC content were evaluated using Wilcoxon rank-sum tests (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

KEGG functional annotation and enrichment analysis

KEGG functional annotations were obtained from eggNOG-mapper v2 (Cantalapiedra et al. 2021). KEGG Orthology (KO) assignments from all annotated protein-coding genes were extracted and aggregated for each genome. KO identifiers were subsequently mapped to the KEGG BRITE hierarchy, and functional abundances were summarized at BRITE level 2. Disease-related pathway categories were excluded from downstream analyses. Differences in normalized KO abundances between macrofungi and microfungi were assessed using Wilcoxon rank-sum tests, and P-values were corrected using the Benjamini–Hochberg method ($FDR < 0.05$).

References

- Bau T, Li HJ, Bao HY, Li Y (2024) A revised checklist of poisonous mushrooms in China. *Journal of Fungal Research* 22(4): 301–321. <https://doi.org/10.13341/j.jfr.2024.0010>
- Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in Prokaryotes, Eukaryotes and Viruses. *Nucleic Acids Research* 33: W451–W454. <https://doi.org/10.1093/nar/gki487>
- Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T (2021) antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* 49: W29–W35. <https://doi.org/10.1093/nar/gkab335>
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15): 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* 38(12): 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Fan GM, Sun QL, Sun Y, Liu D, Li S, Li M, Chen Q, Wang F, Moriya O, Itoh T, Kawasaki H, Yu Y, Cai M, Kim SG, Lee JS, Ma JC, Wu LH (2025) GCM and gcType in 2024: comprehensive resources for microbial strains and genomic data. *Nucleic Acids Research* 53(D1): D763–D771. <https://doi.org/10.1093/nar/gkae1057>
- Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* 42(D1): D699–D704. <https://doi.org/10.1093/nar/gkt1183>

He MQ, Wang MQ, Chen ZH, Deng WQ, Li TH, Vizzini A, Jeewon R, Hyde KD, Zhao RL (2022) Potential benefits and harms: a review of poisonous mushrooms in the world. *Fungal Biology Reviews* 42: 56–68. <https://doi.org/10.1016/j.fbr.2022.06.002>

He MQ, Cao B, Liu F, Boekhout T, Denchev TT, Power N, Denchev CM, Kemler M, Gorjón SP, Begerow D, Valenzuela R, Davoodian N, Niskanen T, Vizzini A, Redhead SA et al. (2024) Phylogenomics, divergence times and notes of orders in Basidiomycota. *Fungal Diversity* 126: 127–406. <https://doi.org/10.1007/s13225-024-00535-w>

Hyde KD, Noorabadi MT, Thiyagaraja V, He MQ, Johnston PR, Wijesinghe SN, Armand A, Biketova AY, Chethana KWT, Erdoğdu M, Ge ZW, Groenewald JZ, Hongsanan S, Kušan I, Leontyev DV et al. (2024) The 2024 Outline of Fungi and fungus-like taxa. *Mycosphere* 15(1): 5146–6239. <https://doi.org/10.5943/mycosphere/15/1/25>

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780. <https://doi.org/10.1093/molbev/mst010>

Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* 52(W1): W78–W82. <https://doi.org/10.1093/nar/gkae268>

Li H, Tian Y, Menolli N Jr, Ye L, Karunarathna SC, Perez-Moreno J, Rahman MM, Rashid MH, Phengsintham P, Rizal L, Kasuya T, Lim YW, Dutta AK, Khalid AN, Huyen LT et al. (2021) Reviewing the world's edible mushroom species: A new evidence-based classification system. *Comprehensive Reviews in Food Science and Food Safety* 20: 1982–2014. <https://doi.org/10.1111/1541-4337.12708>

Liang JM, Huang JE, Wu LH, Cai L (2023) The Chinese microbial type-strain genome project. *The Innovation Life* 1(3): 100045. <https://doi.org/10.59717/j.xinn-life.2023.100045>

Liu F, Wang SH, Cheewangkoon R, Zhao RL (2025) Uneven distribution of Prokaryote-derived horizontal gene transfer in Fungi: a lifestyle-dependent phenomenon. *mBio* 16: e02855-24.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37(5): 1530–1534. <https://doi.org/10.1093/molbev/msaa015>

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19): 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19: 215–225. <https://doi.org/10.1093/bioinformatics/btg1080>

Wang F, Wang K, Cai L, Zhao M, Kirk PM, Fan G, Sun Q, Li B, Wang S, Yu Z, Han D, Ma J, Wu L, Yao Y (2023) Fungal names: a comprehensive nomenclatural

repository and knowledge base for fungal taxonomy. *Nucleic Acids Research* 51(D1): D708–D716. <https://doi.org/10.1093/nar/gkac926>

Wu F, Zhou LW, Yang ZL, Bau T, Li TH, Dai YC (2019) Resource diversity of Chinese macrofungi: edible, medicinal and poisonous species. *Fungal Diversity* 98: 1–76. <https://doi.org/10.1007/s13225-019-00432-7>

Wu F, Liu SB, Wang XH, Zhang M, Yang ZL, Dai YC (2026) Species cataloging of edible and medicinal fungi in China. *Mycosystema* 45(1): 250237.

Wu G, Yang ZL (2026) Chapter 1: Diversity and taxonomy of edible and medicinal mushrooms. In: Xu JP, James TY, Bao DP (eds) *Mushroomology*. Brill/Wageningen Academic, Leiden.