



OPEN

DATA DESCRIPTOR

# A Singapore-centric Fungal Dataset of 518 Cultivated Strains with Visual Phenotypes and Taxonomic Identity

Darren Wei Xian Ten<sup>1,2</sup>, Fong Tian Wong<sup>1,2,3</sup>, Yee Hwee Lim<sup>1,2</sup> & Winston Koh<sup>1,2,4</sup> ✉

The fungal kingdom represents a greatly untapped resource to produce a wide range of bioactive secondary metabolites, including antibiotics, anticancer agents, industrially significant dyes and enzymes. To-date, it is estimated only less than 5% of all fungi have been characterised, a deficit that is especially pronounced in tropical regions like Singapore, where fungal diversity remains underexplored compared to northern hemisphere counterparts. This underlines the urgency and importance of our research which motivated the creation of our curated dataset, aiming to address this gap and contribute to understanding the broader ecosystem. We developed a generalisable cultivation workflow that enables systematic strain preparation, supports high-resolution imaging, and yields sufficient fungal biomass amenable for genomic analyses. This resulted in a diverse collection of 518 phylogenetically and ecologically varied fungal strains from both terrestrial and marine environments in biodiverse Singapore. The curated dataset from this project captures both taxonomic identity and colony-level morphological traits serving as a foundation for visual phenotype to taxonomy mapping through the integration of computer vision.

## Background & Summary

Fungi comprise one of the most diverse and ecological indispensable kingdoms on Earth. They are deeply integrated into the evolutionary history of life and continually serve as primary decomposers, facilitating nutrient cycling that sustains ecosystems<sup>1-3</sup>. As one of nature's most prolific chemists, fungi synthesise an extraordinary repertoire of molecules, with transformative applications across medicine, biotechnology and industry<sup>4-7</sup>. The vast potential and opportunities offered by fungal chemistry has pushed efforts to broaden the exploration of fungal diversity, as exemplified by several large-scale initiatives. Notably, the Joint Genome Institute (JGI)'s MycoCosm's platform currently hosts over 2,500 publicly available fungal genomes, compiled through a combination of community-contributed datasets and in-house sequencing efforts. This is primarily from the 1000 Fungal Genomes Project, which prioritises sequencing of reference genomes across more than 500 fungal families<sup>8,9</sup>. Another major initiative is MycoBank, a global database and inter-repository hub for the registration and standardisation of novel fungal species nomenclature, also providing reference access to associated fungal taxonomic and genomic resources<sup>10,11</sup>.

Despite this global momentum to elucidate the projected existence of 5 to 12 million fungal species, only ~150,000 species have been formally described, with an even smaller proportion explored for their biotechnological potential<sup>12-15</sup>. This disparity is particularly pronounced in tropical regions, where characterized fungi species are predominantly from the northern hemisphere<sup>16,17</sup>. Situated in the biodiverse equatorial region of

<sup>1</sup>Institute of Sustainability for Chemicals, Energy and Environment (ISCE2), Agency for Science, Technology and Research (A\*STAR), 8 Biomedical Grove, #07-01 Neuros Building, Singapore, 138665, Republic of Singapore.

<sup>2</sup>Singapore Integrative Biosystems and Engineering Research, Biosystems and Engineering (SIBER), Agency for Science, Technology and Research (A\*STAR), 2 Fusionopolis Way, #08-01 Innovis, Singapore, 138634, Republic of Singapore. <sup>3</sup>Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A\*STAR), 61 Biopolis Drive, #07-06, Proteos, Singapore, 138673, Republic of Singapore. <sup>4</sup>Bioinformatics Institute (BI), Agency for Science, Technology and Research (A\*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore, 138671, Republic of Singapore. ✉e-mail: [Winston\\_Koh@a-star.edu.sg](mailto:Winston_Koh@a-star.edu.sg)

Southeast Asia, Singapore harbours a remarkable diversity of flora and fauna, with an estimated 40,000 species, with a recent biodiversity survey formally documented over 3,000 species<sup>18,19</sup>, positioning it as a haven for concentrated biodiversity research. Comparatively, more studies can be emphasised on Singapore's fungal biodiversity, with much of its ecology and functional properties remained unexplored, hampering efforts on further biotechnological utilisation, conservation, and understanding of the broader ecosystem<sup>20–22</sup>.

To address this imbalance, we implemented a rapid, streamlined and scalable solid-state fermentation workflow that bins collected isolates by terrestrial and marine origins and employed two general purpose media, identified for its ability to support robust biomass production across phylogenetically diverse lineages. Applied to a curated collection of 1,136 cryopreserved fungal isolates, this workflow achieved an 80.2% revival rate ( $n = 911$ ; 99% CI: 77.1%–83.2%). High-resolution images were also collected across 3 key culture milestones [Day 7, Day 14, and pre-harvest (Day 11 to Day 71)] for some of the strains. Of the successfully revived isolates, 94.1% ( $n = 857$ ; 99% CI: 92.3%–95.9%) yielded sufficient biomass for DNA extraction, facilitating taxonomy curation with 18S. Quality assurances involved decontamination screening to omit non-fungal sequences and exclusion of fungal genomes lacking pre-harvest colony images.

The eventual curated collection comprises 518 diverse fungal strains, each paired with colony morphology images across various timepoints amounting to a total of 1,528 images. We processed the 518 fungal 18S rRNA sequences with multiple alignment using fast Fourier transform (MAFFT) to produce uniform base pair alignments<sup>23</sup>. With this alignment, we used FastTree to infer a maximum-likelihood phylogeny<sup>24</sup>, applied 100 bootstraps to obtain a majority-rule consensus tree, followed by visualisation with GGTREE<sup>25</sup>. Annotated with taxonomic assignments and ecological metadata, the phylogenetic tree offers a comprehensive overview of Singapore's fungal diversity. Spanning 127 genera and 3 unassigned strains across 16 classes and 4 phyla, the dataset represents fungal strains from a broad spectrum of habitats across the island, from soils and freshwater systems to marine environments (Fig. 1; Table 1).

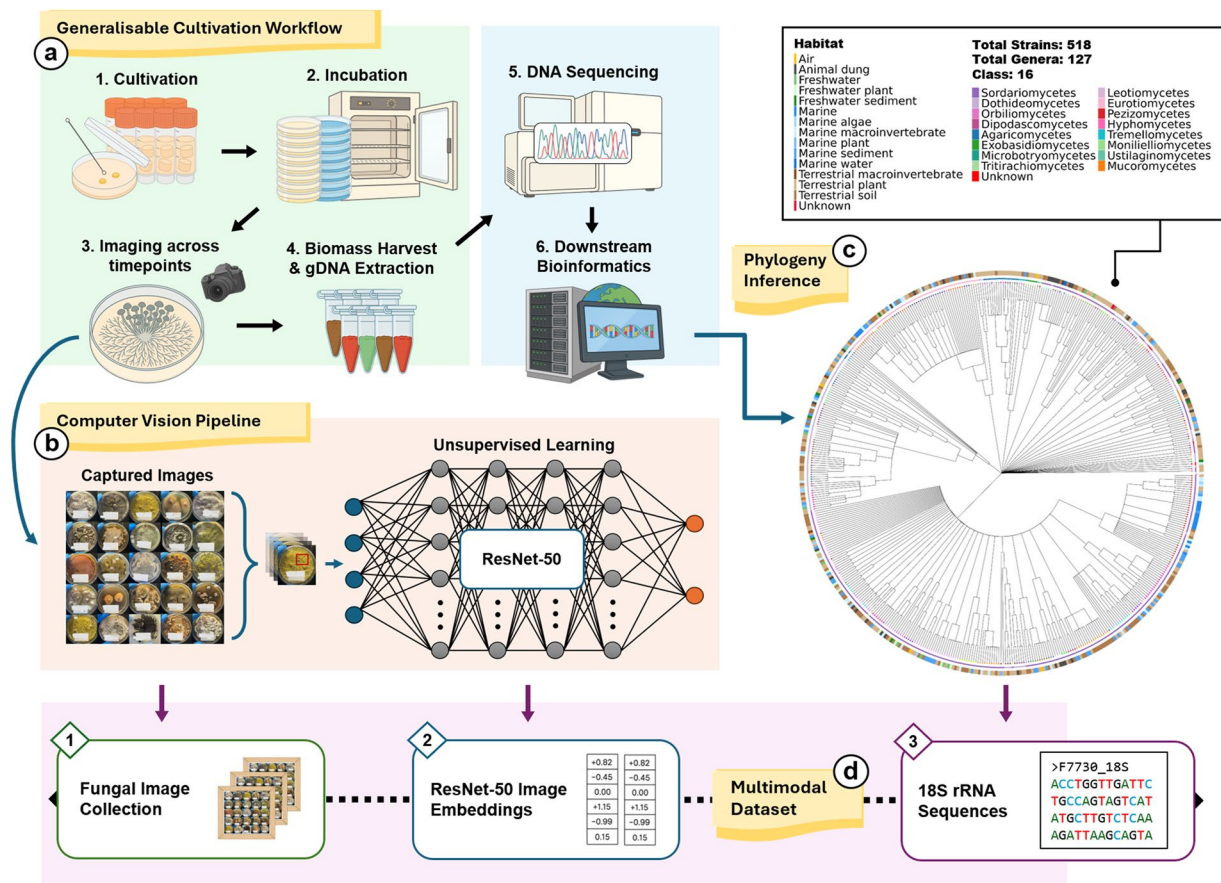
To demonstrate the utility of the multi-modal dataset, we focused on 518 pre-harvest colony images drawn from the complete image collection of 1,528. High-dimensional representations of image features of the strain collection were extracted using a Residual Network with 50 layers (ResNet-50) convolutional neural network (CNN)<sup>26</sup>. Prior to feature extraction, input images were pre-processed with the Efficient and Accurate Scene Text (EAST) detection model to identify and occlude text regions<sup>27</sup>, ensuring downstream analyses were driven solely by morphological traits. The resulting image embeddings effectively captured ecological niches, providing a direct linkage between phenotype and taxonomy. To evaluate the efficacies of the captured images and their corresponding embeddings in preserving taxonomic resolution and ecological niches, we projected the 2048-dimensional embeddings into two-dimensions using t-distributed stochastic neighbour embedding (t-SNE) (Fig. 2). Overlaid ecological metadata and captured pre-harvest images, this projection (Fig. 2a) revealed interesting phenotypic clusters, most notably by four coherent clusters defined by pigmentation: (2a-1) strains from diverse habitats exhibiting a deep black phenotype; (2a-2) clear white phenotypes from freshwater and marine sources; (2a-3) yellow strains predominantly from marine sources. Morphology-based clustering was also evident, highlighted by a cluster of strains displaying circular and lobate colony morphology (Fig. 2a-4). Alternative overlaying with taxonomical metadata enabled seamless toggling between phenotype and taxonomy context (Fig. 2b). For example, the black phenotype cluster (Fig. 2a-1) corresponds to Dothideomycetes genera such as *Hortaea* and *Cochliobolus* (Fig. 2b-1), both documented for their dark colony morphology<sup>28,29</sup>. This first-pass analysis validates our computer vision pipeline and further demonstrates the utility of the multimodal dataset to capture meaningful biological nuances.

Beyond these exploratory findings, this dataset provides not only a streamlined cultivation workflow adaptable to a wide spectrum of fungal taxa but also presents a unique opportunity to interrogate the distinct colony morphology and underlying genomic signatures of 518 fungal strains from Singapore. It includes time-stamped fungal images across key developmental milestones, high-dimensional ResNet-50 embeddings, predicted 18S rRNA sequences, and corresponding taxonomy and ecological metadata. Together, these components form a unified platform for bridging visual phenotype to taxonomy, serving both as a reference for fungal biodiversity in tropical Singapore and as a hypothesis-generation platform, spurring targeted future investigations and supporting downstream genomic exploration for deep phenotypic and taxonomic insights.

The current dataset provides valuable insights into fungal diversity through 18S sequencing and representative time-point imaging, enabling analysis of species-level patterns. While we recognise that distinguishing closely related taxa may benefit from additional data dimensions, this foundational approach significantly advances our understanding of fungal ecological dynamics. Incorporating high-confidence whole-genome assemblies and expanded temporal imaging in future studies will further improve taxonomic accuracy, uncover new morphotype patterns, and elucidate the molecular mechanisms driving phenotypic variation.

## Methods

**Cultivation, imaging and harvest.** A total of 1,136 uncharacterized fungal isolates from Singapore was cultivated. These isolates are part of the Natural Product Library (NPL) collection in Agency for Science, Technology and Research (A\*STAR)<sup>30</sup>. Cryovials containing preservation fluid (reverse osmosis water with 25% glycerol) and fungal agar plugs (agar with fungal mycelium) were retrieved from a  $-80^{\circ}\text{C}$  freezer and allowed to thaw completely at room temperature. Cryopreserved fungal plugs were cultured on either ME agar [malt extract (30 g/L), agar (15 g/L), mycological peptone (5 g/L)] or GYS agar [sea salt (40 g/L), agar (15 g/L), glucose (10 g/L), yeast extract (1 g/L)], selected based on terrestrial or marine origin. Cultures were incubated at  $25^{\circ}\text{C}$  ( $42^{\circ}\text{C}$  for thermophilic isolates), and colony morphology was photographed in a biosafety cabinet across key developmental stages (varied based on growth rate) using an iPhone 15 Pro (48-megapixel, macro mode).

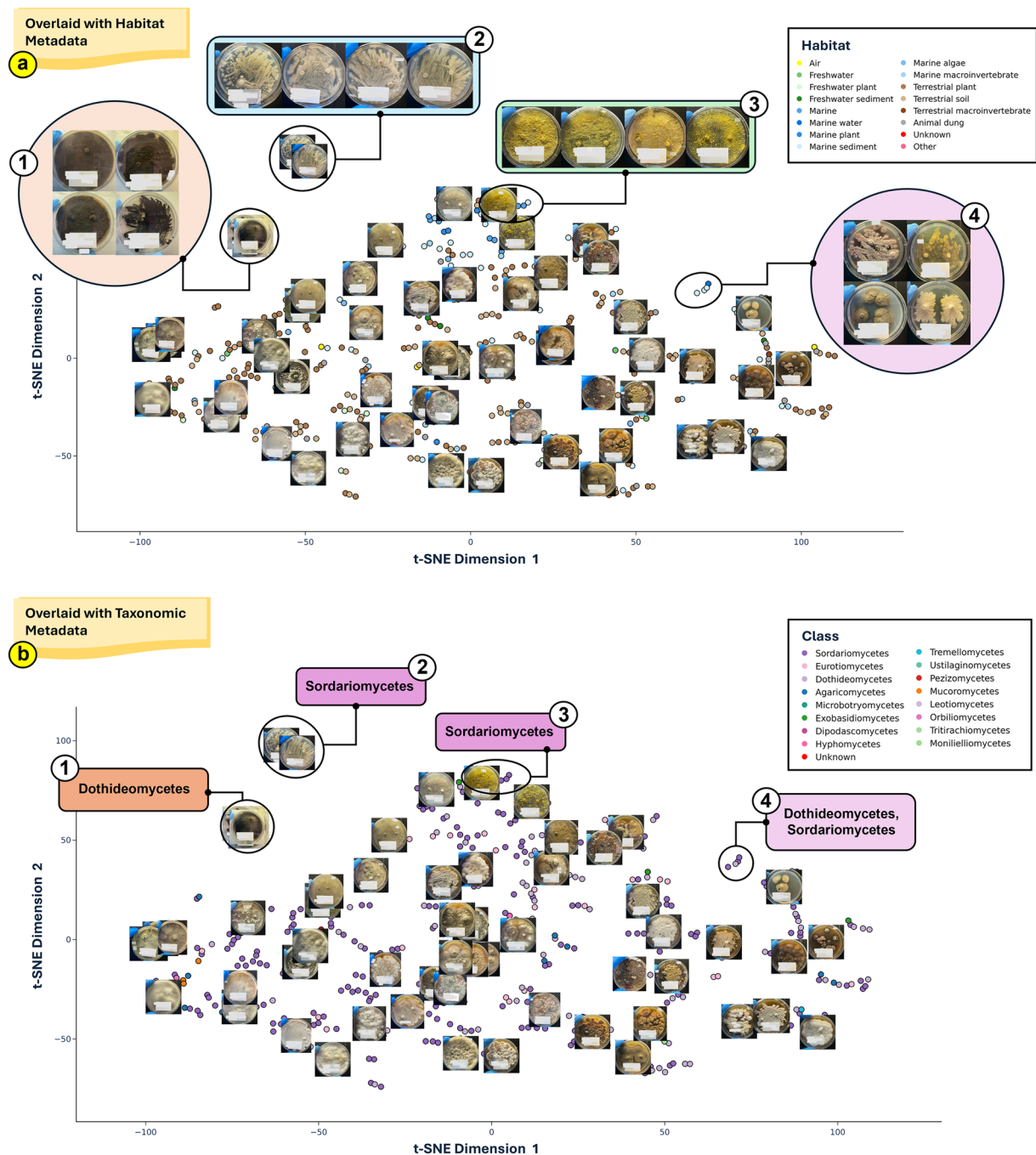


**Fig. 1** Overview of the multi-modal fungal dataset, providing an integrated perspective into the phenotypic and genomic diversity of Singapore's fungi. **(a)** The cultivation workflow involves the following steps: First, (1) cryopreserved isolates are revived through solid-state cultivation and (2) incubated under standardised conditions, with (3) high-resolution colony images captured at key developmental milestones. Once colonies reach target biomass, samples underwent (4) DNA extraction, library preparation, and sequencing. **(b)** Computer vision pipeline for the processing of fungal colony images to extract high-dimensional feature embeddings. **(c)** Phylogenetic tree of 518 fungal strains annotated with class-level taxonomy and habitat origin. **(d)** Generated multimodal dataset consisting of (1) fungal image collection across key development milestones, (2) ResNet-50 embeddings of all images, and (3) 18S rRNA sequences for taxonomic inference.

Taxonomic Distribution		Ecological Distribution	
Description	Count	Description	Count
Total Genera	127	Terrestrial Habitat	316 (61.0%)
Highest Represented Genus ( <i>Fusarium</i> )	56 (10.8%)	Marine Habitat	134 (25.9%)
Species with Unknown Lineages	3 (0.6%)	Animal Dung Habitat	35 (6.8%)
2–10% Genera ( <i>Hypocrea</i> , <i>Penicillium</i> , <i>Purpureocillium</i> , <i>Aspergillus</i> , <i>Verticillium</i> , <i>Phoma</i> , <i>Leptosphaeria</i> , <i>Hypoxylen</i> , <i>Proxiocillium</i> , <i>Myrothecium</i> , <i>Cochliobolus</i> )	216 (41.7%)	Freshwater Habitat	27 (5.2%)
Others	243 (46.9%)	Others	6 (1.2%)

**Table 1.** Taxonomic and Ecological Distribution of 518 Singapore Fungal Strains.

**DNA sequencing, taxonomic inference and phylogenetic analysis.** Fungal DNA was extracted from biomass and sequenced on the Illumina NovaSeq 6000 platform. 18S regions were predicted with (Barrnap, v0.9), applying eukaryote-specific hidden Markov model (HMM)<sup>31</sup>. To facilitate taxonomic assignments at a genus level, the predicted 18S regions of 518 strains were queried against NCBI core nucleotide database (accessed on 24 April 2025) with BLASTn (v2.14.1). Upon filtering by bit-score, each strain was provisionally assigned to the genus corresponding to its top BLAST hit, further supplemented with phylum, sub-phylum, and class assignments with records from JGI's MycoCosm (accessed on 4 September 2024) and MycoBank database (accessed on 1 June 2025).



**Fig. 2** t-SNE projections of 518 ResNet-50 derived feature embeddings from pre-harvest fungal colony images, facilitating phenotypic-genomic analysis. Embeddings (points) are coloured with different metadata, facilitating phenotypic-genomic analysis. **(a)** Embeddings coloured according to the isolate's habitat of origin, with emphasis on four key clusters with distinct morphotypes: (1) deep-black strains, (2) clear-white strains, (3) yellow strains, and (4) circular and lobate colony morphology. **(b)** The same embedding space as **(a)**, coloured with class-level taxonomy, highlighting the same four clusters.

Phylogenetic tree was constructed by performing multiple sequence alignment (MSA) on the 518 fungal 18S rRNA sequences using MAFFT (v7.525) using the-auto mode (FFT-NS-2 algorithm)<sup>23</sup>. The resulting alignment consisted of 518 sequences with equal length (2,287 bp). Phylogenetic robustness was assessed with 100 bootstrap replicates (resampled alignment columns with replacement) and a maximum-likelihood trees was reconstructed with FastTree 2 (v2.1.11)<sup>24</sup>, employing the Generalized Time Reversible (GTR) model with gamma-distributed site-rate variation. Clade support values were derived by generating a majority-rule consensus tree using the R-package Analyses of Phylogenetics and Evolution package (ape, v5.8.1)<sup>32</sup>, with clades retained if supported in  $\geq 50\%$  of replicates (518 tips, 262 internal nodes, mean bootstrap support of 84%). The tree was visualised with GGTree (Bioconductor v3.2)<sup>25</sup>, annotated with genus, class, and habitat metadata.

**Image Pre-processing, ResNet-50, and t-SNE Visualisation.** The 518 pre-harvest fungal colony images were first converted to .jpg format, resized and padded to 1024 pixel while preserving their aspect ratio, ensuring dimensional compatibility with the frozen EAST text detection model<sup>27</sup>. The pre-trained frozen EAST text detection model (inference-optimised) generated geometry maps and confidence scores for detected text presence, subsequently decoded into bounding boxes by translating model-predicted dimensions into input-image coordinates. Bounding boxes with confidence scores below 0.2 were discarded, retaining text regions within a generous confidence range. To mask these detected text regions and minimise bias in downstream phenotypic analysis, a Gaussian blur and semi-transparent overlay were applied to each bounding box.

The masked images were subsequently resized to a resolution of  $224 \times 224$  pixels to enable feature extraction by the ResNet-50 model. The images were then transformed into PyTorch tensors, converting pixel values to normalised floats in [Channel, Height, Width] format. These tensors were standardised using ImageNet dataset parameters (channel-wise mean [0.485, 0.456, 0.406], standard deviation [0.229, 0.224, 0.225]), aligning with ResNet-50 originally training conditions<sup>26</sup>. These vectors were iteratively fed into the ResNet-50 model for unsupervised feature extraction, extracting 2048-dimensional embeddings for each input image from the final average pooling layer of the model. Principal component analysis (PCA) and t-SNE, implemented via scikit-learn (v1.6.1), were applied to the 518 image embeddings to denoised the original high-dimensional feature space to 10 principal components, followed by further dimensionality reduction to two-dimensions ( $n\_components = 2$ ,  $perplexity = 5$ ,  $random\ state = 42$ ). Relevant metadata (e.g., taxonomy, ecological origin) were overlaid onto the t-SNE plot to guide interpretation of the ResNet-50-derived image clusters.

### Data Records

The 518 fungal dataset comprises of (1) fungal colony morphology images across key developmental stages (.jpg format), (2) ResNet-50 pre-harvest image embeddings (PyTorch format), (3) fungal strains taxonomy and ecological metadata, and (4) Barrnap-extracted 18S region (.fasta format). All components of the dataset are available at Figshare (<https://doi.org/10.6084/m9.figshare.29434469.v2>)<sup>33</sup>. The 18S sequences have also been deposited and is publicly accessible in the NCBI Sequence Read Archive (SRA) under the accession number SRP647581 (<https://identifiers.org/ncbi/insdc.sra:SRP647581>)<sup>34</sup>.

### Technical Validation

Captured fungal colony images were first subjected to automated text occlusion by the frozen EAST text detection model, followed by processing with the ResNet-50 CNN model. The extracted 2048-dimensional embeddings capture ecological and taxonomical features, highlighting the robustness of the dataset.

### Usage Notes

This dataset enables users to investigate underexplored fungal lineages and help address gaps in the fungal phylogeny, with emphasis on Singapore's rich mycobiota. Additionally, this dataset provides the unique opportunity to examine the dynamic colony morphology of 518 strains and explore its relationship to corresponding genomic signatures, presenting a valuable atlas for visual phenotype to taxonomic linkage. Accompanied by detailed strain metadata, the dataset, alongside Resnet-50 derived image embeddings and clustering, functions as a rapid hypothesis-generation platform. Users can identify fungal clusters of interest based on its phenotypic characteristics and rapidly relate its underlying genomics, creating a systematic and targeted exploration strategy for downstream functional or genomic studies. This phenotypic-taxonomic framework is readily amenable to additional visual and language model applications, to extract deeper insights from the rich mine of visual and genomic data.

### Data availability

All dataset components have been made publicly available on Figshare<sup>33</sup>. The 18S sequences have also been deposited to the NCBI SRA under the accession number SRP647581 (<https://identifiers.org/ncbi/insdc.sra:SRP647581>)<sup>34</sup>.

### Code availability

Scripts for (1) 18S rRNA prediction with Barrnap, (2) MAFFT alignment, (3) FastTree maximum-likelihood tree inference, (4) consensus tree construction and (5) ResNet-50 embedding generation are available on the GitHub repository (<https://github.com/twxdarren/fungal-trove.git>) and Figshare<sup>33</sup>.

Received: 4 July 2025; Accepted: 24 December 2025;

Published online: 22 January 2026

### References

1. Case, N. T. *et al.* The future of fungi: threats and opportunities. *G3 Genes|Genomes|Genetics* **12**, jkac224 (2022).
2. Peay, K. G., Kennedy, P. G. & Talbot, J. M. Dimensions of biodiversity in the Earth mycobiome. *Nat Rev Microbiol* **14**, 434–447 (2016).
3. Tedersoo, L. *et al.* Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
4. Bhattarai, K., Bhattarai, K., Kabir, M. E., Bastola, R. & Baral, B. Fungal natural products galaxy: Biochemistry and molecular genetics toward blockbuster drugs discovery. in *Advances in Genetics* vol. 107 193–284 (Elsevier, 2021).
5. Danner, C., Mach, R. L. & Mach-Aigner, A. R. The phenomenon of strain degeneration in biotechnologically relevant fungi. *Appl Microbiol Biotechnol* **107**, 4745–4758 (2023).
6. Hyde, K. D. *et al.* Current trends, limitations and future research in the fungi? *Fungal Diversity* **125**, 1–71 (2024).

7. Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H. & Lington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5601–5606 (2017).
8. Ahrendt, S. R., Mondo, S. J., Haridas, S. & Grigoriev, I. V. MycoCosm, the JGI's Fungal Genome Portal for Comparative Genomic and Multiomics Data Analyses. in *Microbial Environmental Genomics (MEG)* (eds. Martin, F. & Uroz, S.) vol. 2605 271–291 (Springer US, New York, NY, 2023).
9. Grigoriev, I. V. *et al.* MycoCosm portal: gearing up for 1000 fungal genomes. *Nucl. Acids Res.* **42**, D699–D704 (2014).
10. Crous, P. W., Grams, W., Stalpers, J. A., Robert, V. & Stegehuis, G. MycoBank: an online initiative to launch mycology into the 21st century. *Studies in Mycology* **50**, 19–22 (2004).
11. Robert, V. *et al.* MycoBank gearing up for new horizons. *IMA Fungus* **4**, 371–379 (2013).
12. Huberman, L. B. Developing Functional Genomics Platforms for Fungi. *mSystems* **6**, <https://doi.org/10.1128/msystems.00730-21> (2021).
13. Hyde, K. D. The numbers of fungi. *Fungal Diversity* **114**, 1–1 (2022).
14. Paterson, R. R. M., Solaiman, Z. & Santamaria, O. Guest edited collection: fungal evolution and diversity. *Sci Rep* **13**, 21438, s41598-023-48471-0 (2023).
15. Wu, B. *et al.* Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. *Mycology* **10**, 127–140 (2019).
16. Aime, M. C. & Brearley, F. Q. Tropical fungal diversity: closing the gap between species estimates and species discovery. *Biodivers Conserv* **21**, 2177–2180 (2012).
17. Stallman, J. K. *et al.* The contribution of tropical long-term studies to mycology. *IMA Fungus* **15**, 35 (2024).
18. Chisholm, R. A. *et al.* Two centuries of biodiversity discovery and loss in Singapore. *Proc. Natl. Acad. Sci. USA* **120**, e2309034120 (2023).
19. Davison, G. Special Ecology Feature: Biodiversity's Crucial Role in the Modern Singapore City. *CITYGREEN* **01**, 102 (2012).
20. Choong, M. F. A. Education on plants and fungi in Singapore: an urgent call. *Nature in Singapore Supplement* **1**, 279286 (2022).
21. Lee, S. & Choong, A. Checklist of Fungi Species with their Category of Threat Status for Singapore. in *The Singapore red data book: red lists of Singapore biodiversity* 515–518 (National Parks Board, Singapore, 2024).
22. Weerakoon, G., Ngo, K. M., Lum, S., Lumbsch, H. T. & Lücking, R. On time or fashionably late for lichen discoveries in Singapore? Seven new species and nineteen new records of *Graphidaceae* from the Bukit Timah Nature Reserve, a highly urbanized tropical environment in South-East Asia. *The Lichenologist* **47**, 157–166 (2015).
23. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
24. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).
25. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**, 28–36 (2017).
26. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778, <https://doi.org/10.1109/CVPR.2016.90> (IEEE, Las Vegas, NV, USA, 2016).
27. Zhou, X. *et al.* EAST: An Efficient and Accurate Scene Text Detector. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2642–2651, <https://doi.org/10.1109/CVPR.2017.283> (IEEE, Honolulu, HI, 2017).
28. Eliahu, N., Igbaria, A., Rose, M. S., Horwitz, B. A. & Lev, S. Melanin Biosynthesis in the Maize Pathogen *Cochliobolus heterostrophus* Depends on Two Mitogen-Activated Protein Kinases, Chk1 and Mps1, and the Transcription Factor Cmr1. *Eukaryot Cell* **6**, 421–429 (2007).
29. Plemenitaš, A., Vaupotič, T., Lenassi, M., Kogej, T. & Gunde-Cimerman, N. Adaptation of extremely halotolerant black yeast *Hortaea werneckii* to increased osmolarity: a molecular perspective at a glance. *Studies in Mycology* **61**, 67–75 (2008).
30. Ng, S. B. *et al.* The 160K Natural Organism Library, a unique resource for natural products research. *Nat Biotechnol* **36**, 570–573 (2018).
31. Seemann, T. barrnap 0.9: rapid ribosomal RNA prediction. <https://github.com/tseemann/barrnap> (2013).
32. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
33. Ten, D. W. X., Wong, F. T., Lim, Y. H. & Koh, L. C. W. A Singapore-centric Fungal Dataset of 518 Cultivated Strains with Visual Phenotypes and Taxonomic Identity. *figshare* <https://doi.org/10.6084/m9.figshare.29434469> (2025).
34. *NCBI Sequence Read Archive* <https://identifiers.org/ncbi/insdc.sra:SRP647581> (2025).

## Acknowledgements

This work is supported by the Agency for Science, Technology and Research (A\*STAR) under C233017006, and the Singapore Integrative Biosystems and Engineering Research Strategic Research & Translational Thrust (SIBER SRTT, A\*STAR). The authors extend their gratitude towards Dr. Siew Bee Ng (SIFBI, A\*STAR) for her support on strains retrieval.

## Author contributions

F.T.W., Y.H.L. and W.K. conceptualised, designed and coordinated the study. D.W.X.T. conducted the experiments and data acquisition. D.W.X.T. wrote the manuscript along with inputs from all the authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to W.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026