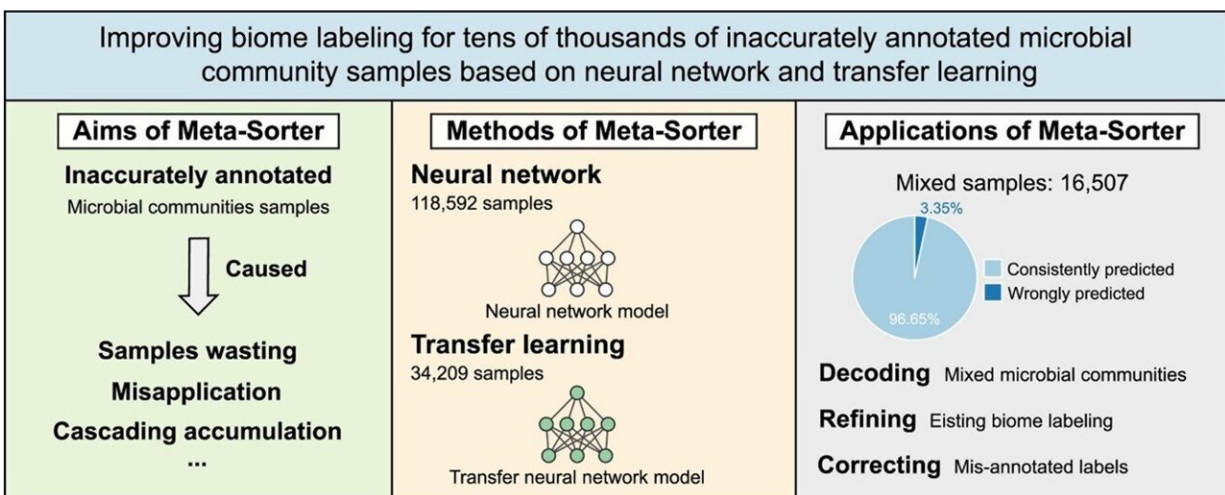


# Refining biome labeling for microbial community samples

September 8 2023



Graphical abstract. Credit: *Environmental Science and Ecotechnology*

In a [study](#) published in the journal *Environmental Science and Ecotechnology*, researchers from Huazhong University of Science and Technology have introduced "Meta-Sorter," an AI-based method that leverages neural networks and transfer learning to significantly improve biome labeling for thousands of microbiome samples in the MGnify database, especially those with incomplete information.

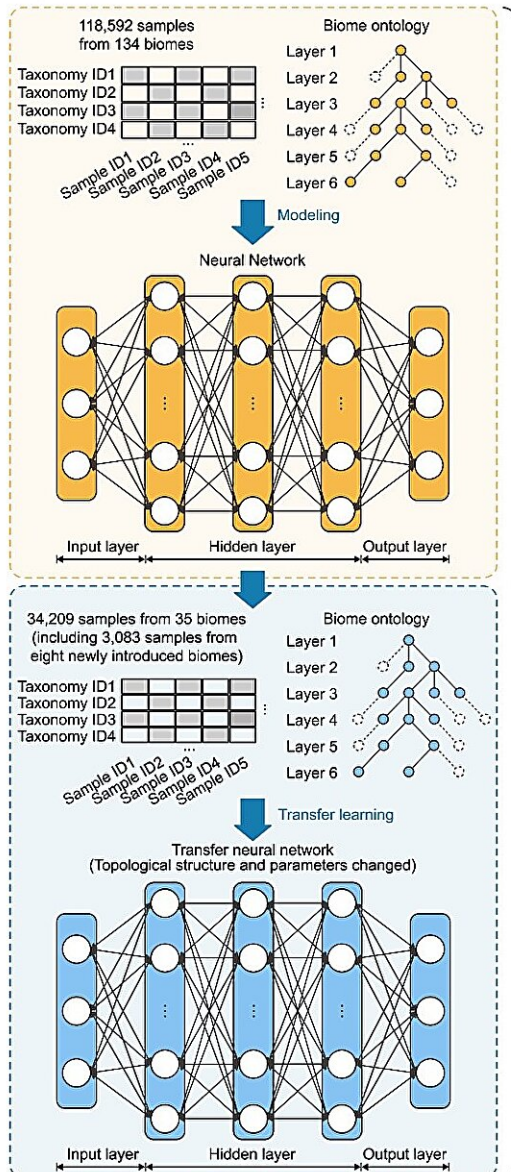
The Meta-Sorter approach comprises two crucial steps. Firstly, a [neural network model](#) is meticulously constructed using 118,592 [microbial](#)

[samples](#) from 134 biomes and their respective biome ontology, boasting an impressive average AUROC of 0.896. This model accurately classifies samples with detailed biome information, serving as a strong foundation for further analyses.

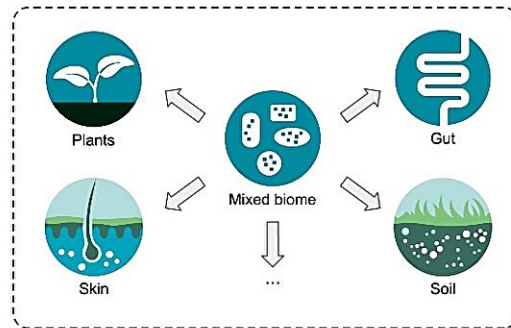
Secondly, to address the challenge of newly introduced samples with different characteristics, researchers incorporated transfer learning with 34,209 newly added samples from 35 biomes, including eight novel ones. The transfer neural network model achieved an outstanding average AUROC of 0.989, successfully predicting biome information for newly introduced samples annotated as "Mixed biome."

The results of Meta-Sorter are indeed impressive, achieving an overall accuracy rate of 96.7% in classifying samples among the 16,507 lacking detailed biome annotations. This breakthrough effectively resolves the issue of cascading errors and opens up exciting new possibilities for knowledge discovery across various scientific disciplines, particularly in [environmental research](#).

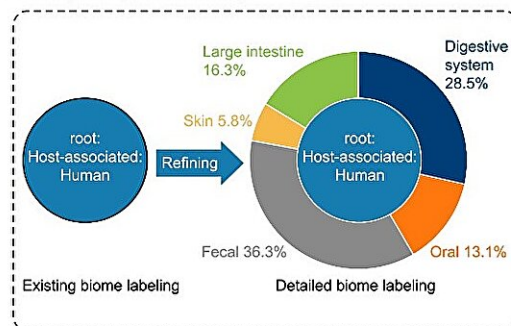
**a** The process of model construction and transfer learning of Meta-Sorter



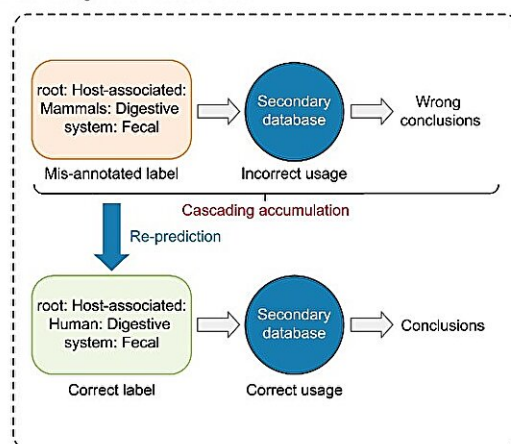
**b** Decoding mixed microbial communities



**c** Refining existing biome labeling



**d** Correcting mis-annotated labels



The rationale and applications of Meta-Sorter. a, The process of model construction and transfer learning of Meta-Sorter. Two input files, biome ontology and samples' taxonomic structures with detailed biome information, are required in model construction and transfer learning. The yellow box shows that the neural network model was constructed based on 118,592 existing samples with detailed information on 134 biomes and their biome ontology. The blue box shows that the transfer neural network model was constructed using 34,209 newly introduced samples from 35 biomes (including 3083 samples from eight newly introduced biomes) and transfer learning to the existing neural network

model. b–d, The applications of Meta-Sorter. Meta-Sorter decoded the samples' biome labels annotated as “Mixed biome” into detailed biome labels (b). Meta-sorter refined the biome labels in more detail to obtain more valuable information for reference (c). Meta-Sorter corrected the mis-annotated samples' labels to avoid cascading accumulation (d). Credit: *Environmental Science and Ecotechnology*

Moreover, Meta-Sorter's success extends to refining the biome annotation for under-annotated and mis-annotated samples. Its intelligent and automatic assignment of precise classifications to ambiguous samples provides valuable insights beyond the original literature, while the differentiation of samples into specific environmental categories enhances the reliability and validity of research conclusions.

With the ongoing development of standardized protocols for data submission and incorporation of additional meta-data information, Meta-Sorter is set to revolutionize the way researchers analyze and interpret microbial community samples. Ultimately, it will lead to more accurate and insightful discoveries in the realm of microbiome research and beyond.

**More information:** Nan Wang et al, Refining biome labeling for large-scale microbial community samples: Leveraging neural networks and transfer learning, *Environmental Science and Ecotechnology* (2023). [DOI: 10.1016/j.esec.2023.100304](https://doi.org/10.1016/j.esec.2023.100304)

Provided by Chinese Society for Environmental Sciences

Citation: Refining biome labeling for microbial community samples (2023, September 8) retrieved 15 May 2024 from <https://phys.org/news/2023-09-refining-biome-microbial->

[community-samples.html](https://community-samples.html)

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.