Briefings in Bioinformatics, 2025, **26(1)**, bbae679
https://doi.org/10.1093/bib/bbae679
Problem Solving Protocol

The iPhylo suite: an interactive platform for building and annotating biological and chemical taxonomic trees

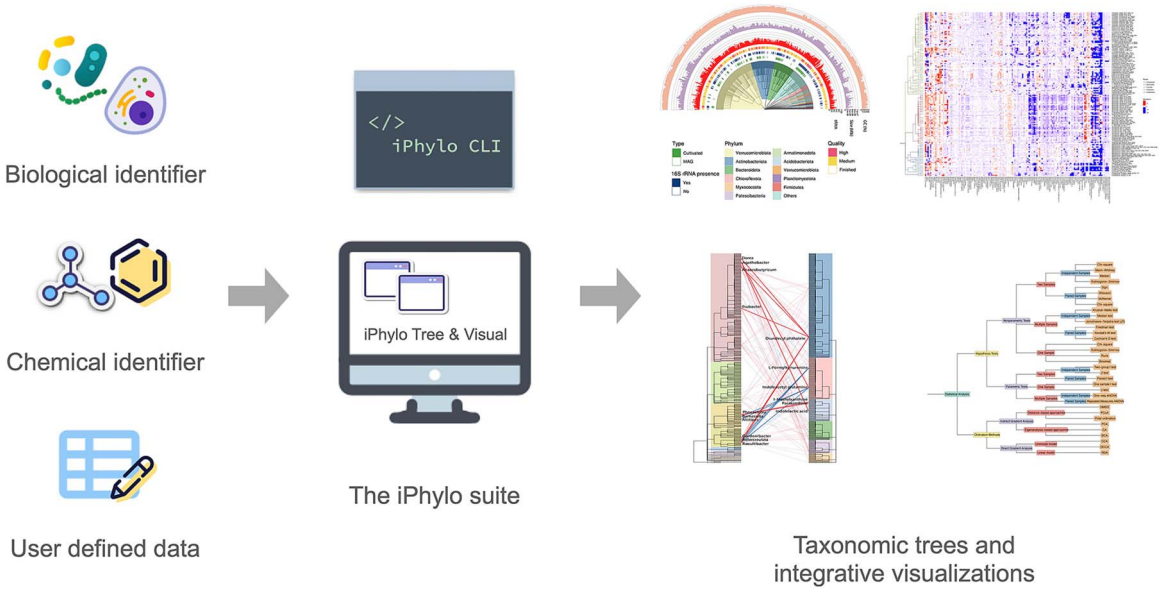
Yueer Li[1,‡], Chen Peng[1,‡], Fei Chi[2], Zinuo Huang[1], Mengyi Yuan[1], Xin Zhou[3], Chao Jiang[1,4,5,*]

[1]MOE Key Laboratory of Biosystems Homeostasis & Protection, and Zhejiang Provincial Key Laboratory of Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang 310030, China
[2]Innovation Center of Yangtze River Delta, Zhejiang University, 828 Zhongxing Road, Jiashan County, Jiaxing, Zhejiang 314103, China
[3]Department of Genetics, Stanford University, Stanford, 291 Campus Drive, Santa Clara County, CA 94305, United States
[4]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, First Affiliated Hospital, Zhejiang University School of Medicine, 79 Qingchun Road, Shangcheng District, Hangzhou, Zhejiang 310009, China
[5]Center for Life Sciences, Shaoxing Institute, Zhejiang University, 8 Nanbin East Road, Shangyu District, Shaoxing, Zhejiang 321000, China
*Corresponding author. MOE Key Laboratory of Biosystems Homeostasis & Protection, and Zhejiang Provincial Key Laboratory of Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310030, China. E-mail: jiang_chao@zju.edu.cn. X: @Chao_Jiang1
‡Yueer Li and Chen Peng are co-first authors (equal contribution).

Abstract

Accurate and rapid taxonomic classifications are essential for systematically exploring organisms and metabolites in diverse environments. Many tools have been developed for biological taxonomic trees, but limitations apply, and a streamlined method for constructing chemical taxonomic trees is notably absent. We present the iPhylo suite (https://www.iphylo.net/), a comprehensive, automated, and interactive platform for biological and chemical taxonomic analysis. The iPhylo suite features web-based modules for the interactive construction and annotation of taxonomic trees and a stand-alone command-line interface (CLI) for local operation or deployment on high-performance computing (HPC) clusters. iPhylo supports National Center for Biotechnology Information (NCBI) taxonomy for biologicals and ChemOnt and NPClassifier for chemical classifications. The iPhylo visualization module, fully implemented in R, allows users to save progress locally and customize the underlying R code. Finally, the CLI module facilitates analysis across all hierarchical relational databases. We showcase the iPhylo suite's capabilities for visualizing environmental microbiomes, analyzing gut microbial metabolite synthesis preferences, and discovering novel correlations between microbiome and metabolome in humans and environment. Overall, the iPhylo suite is distinguished by its unified and interactive framework for in-depth taxonomic and integrative analyses of biological and chemical features and beyond.

Graphical Abstract



Keywords: taxonomic analysis; chemical taxonomy; tree visualization; interactive annotation; integrative multi-omic analysis

Received: August 2, 2024. **Revised:** November 6, 2024. **Accepted:** December 13, 2024

Introduction

The systematic investigation of organisms and metabolites in human and diverse environments requires accurate and fast taxonomic classifications. Over the years, the comprehensive biological taxonomy, or Linnaean taxonomy, has been curated by NCBI [1] and European Molecular Biology Laboratory (EMBL) [2] to organize the ever-increasing number of species. The fast-growing need for integrating metabolomic perspective in microbiome and precision medicine research provokes the need to analyze chemical taxonomy conveniently. Similar to the biological classification system, several chemical classification systems have been developed, including the ChEBI ontology [3], LIPID MAPS [4], ChemOnt [5], and NPClassifier [6]. These systems use various hierarchical chemical criteria to organize chemical compounds. Specifically, ChemOnt classifies a large number of chemicals based on structural features. NPClassifier, conversely, employs machine learning models to classify natural products beyond the simple chemical substructures, incorporating factors such as the producing organism, biosynthetic pathways, and their biological properties. Utilizing a taxonomic tree structure to present a classification system is highly intuitive, allowing for a concise representation of the hierarchical relationships between features.

Several online and stand-alone tools offer taxonomic classification, visualization, and annotations in various combinations, including MEGA [7], ETE Toolkit [8], phangorn [9], taxtree [10], IQ-TREE [11], and PhyloT [12] for biological trees. The NCBI Taxonomy website offers an online tool (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi) for constructing taxonomic trees. ChemTreeMap [13], Qemistree [14], CluMSID [15], and BioDendro [16] are computational tools designed for clustering and identifying chemicals based on mass-spectrometry features, and they also provide the corresponding visualization. For visualization and annotation, ggtree [17] is a widely used R package for visualizing phylogenetic trees. Additionally, Interactive Tree Of Life (iTOL) [18] stands out as a well-established web-based option for visualization and annotation tools, while the newly developed web-based tool Tree Visualization By One Table (TVBOT) [19] offers similar functionalities.

However, several limitations apply. Software packages such as ETE Toolkit and phangorn require programming proficiency and have a steep learning curve for nonexperts, and taxtree suffers from undesirable processing time. The TaxBrowser of the NCBI website had limited functions, unattractive interfaces, and is incompatible with batch submissions. In summary, most software implementations only apply to different stages of the entire workflow of taxonomic analysis. Switching between software may lead to data format incompatibilities and other issues. PhyloT and iTOL, as mainstream and user-friendly tools, require payment to access their full features. The monetary requirements pose a severe limitation for researchers in less developed countries. More importantly, while constructing phylogenetic trees is popular and relatively easy for biological features, constructing and annotating chemical taxonomic trees is difficult due to a lack of a convenient method. Existing tools such as ChemTreeMap, Qemistree, CluMSID, and BioDendro rely on molecular similarity topologies, which do not provide a phylogenetic-like classification tree that could present the taxonomy, biological properties, biosynthetic pathways, and metabolic participation of chemicals. Additionally, most interactive tools, such as iTOL, do not have the free option for saving and uploading tree-building sessions for more complex projects. No current platform offers the underlying visualization codes in R, a popular language for data analysis. For advanced users, accessing the R source codes for plotting and annotation is highly desirable for further customizations. Finally, with the growing need for integrating metagenomic and metabolomic data in precision medicine and microbiome research, a unified platform for building, visualizing, annotating, and integrating biological and chemical taxonomic trees is imperative.

We present the iPhylo suite: a fully automated and interactive platform for biological and chemical taxonomic analysis. The iPhylo suite includes three modules. Two web-based modules, iPhylo Tree and iPhylo Visual, aim to streamline the workflow, encompassing tree generation, interactive and integrative tree visualization, and extensive graphic and textual annotations. A stand-alone module, iPhylo CLI, was designed for local use and high-performance computing applications and equipped with up-to-date biological and chemical taxonomic databases. To highlight the capabilities of the iPhylo suite, we employed iPhylo in diverse research contexts, including the analysis of glacier microbial genome catalogs, human gut metabolic profiles, integrative analysis of human metagenome and metabolome, and exploring the interactions between biological and chemical environmental exposomes. These case studies demonstrate that the iPhylo suite provides a comprehensive and integrated approach to exploring biological and chemical taxonomy across diverse scenarios. Compared to existing tools, the iPhylo suite offers comprehensive functionality and innovative features (Supplementary Table 1).

Materials and methods

The overall workflow of the iPhylo suite

The iPhylo suite features three modules (Fig. 1): (i) The iPhylo Tree rapidly generates biological or chemical taxonomic trees for up to tens of thousands of organisms and chemicals within minutes (https://www.iphylo.net/). (ii) The iPhylo Visual was developed based on the R framework [17, 20, 21] for visualizing and extensively annotating taxonomic trees (https://www.iphylo.net/visual/). The iPhylo Visual also offers the convenience of saving and uploading work sessions locally, as well as access to source codes for plotting and annotating. Importantly, the Tree and Visual modules are seamlessly integrated, enabling users to swiftly import trees and leaf data from construction to annotation with a single click. (iii) The iPhylo CLI is an offline command-line version of the Tree module with integrated databases that can be deployed locally or on high-performance computing clusters. Importantly, the iPhylo CLI can also construct customized taxonomic trees based on a user-defined hierarchical database, so the applications of the iPhylo suite can be extended beyond biological and chemical classifications.

The intuitive interface

The iPhylo suite server employs user-friendly interfaces, including but not limited to the navigation home, the tree construction page, the annotation dashboard, and the inspection pages (Fig. 2). The homepage displays an array of visually captivating dynamic effects that introduce the website's unique features and functionalities (Fig. 2A). The tree-constructing (Fig. 2B) and annotation dashboard (Fig. 2C) pages feature intuitive and functional form layouts. The inspection page adopts the dragging selector to convert the branches and nodes into structured data, facilitating the exploration of any part of the tree (Fig. 2D). Dedicated tutorials and a gallery page were prepared to showcase the examples.

iPhylo Tree

The iPhylo Tree module is an online web application for constructing taxonomic trees from biological species or chemical compounds. Historically, constructing phylogenetic trees may

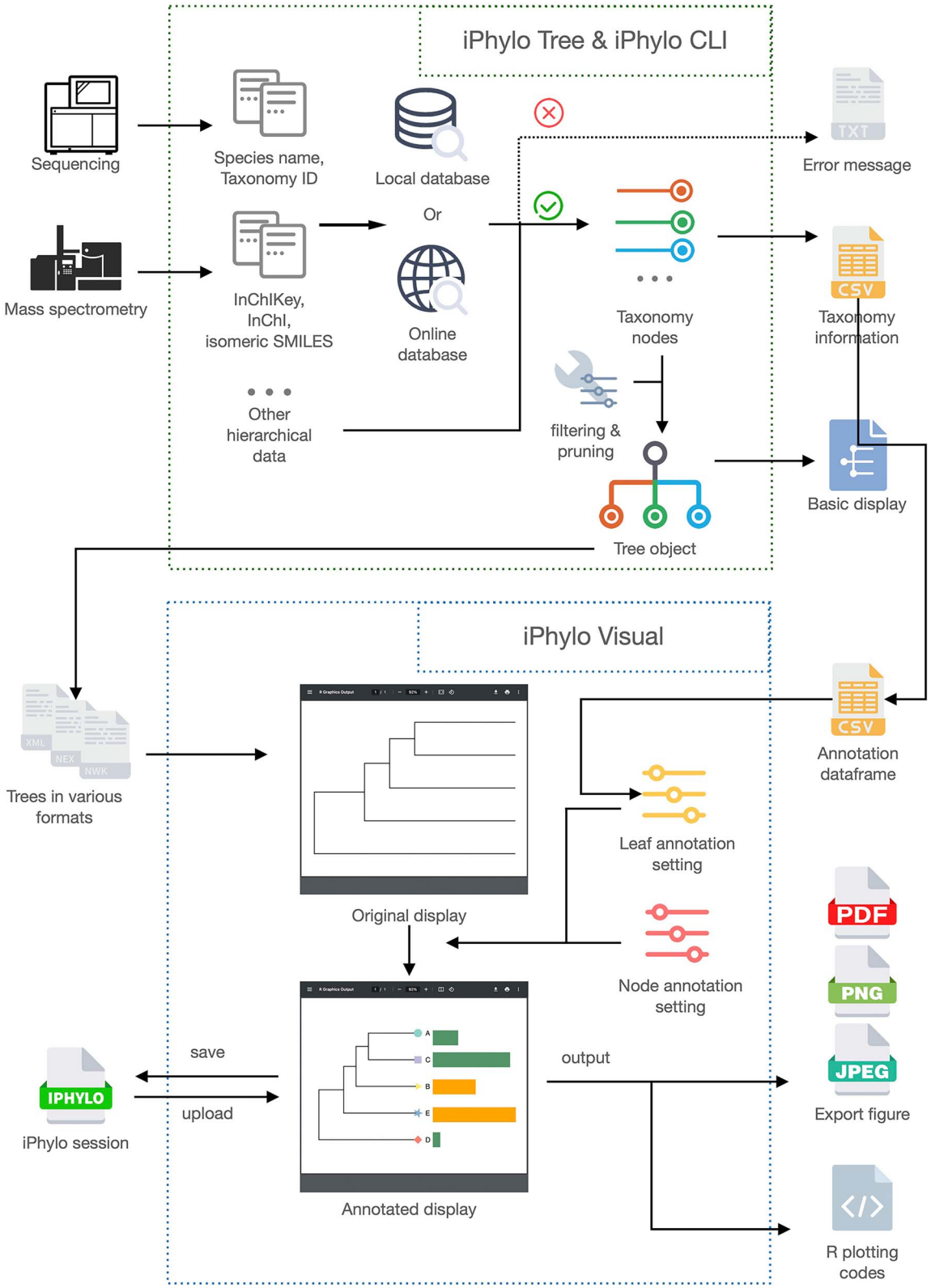


Figure 1. The overview of the iPhylo suite. The iPhylo Tree, CLI, visual modules, and the entire workflow, accompanied by the different formats of input and output files.

involve features such as morphology [22, 23], biochemical [24], or behavioral features [23]. Modern phylogenetic trees are inferred from DNA or protein sequence alignments using tools such as BEAST [25] and MEGA [7]. However, these alignment-based tools cannot easily construct trees spanning all life domains, require heavy data input, and can be time-consuming. Additionally, the results of chemical trees using methods based on molecular similarity, such as Qemistree [14], CluMSID [15], and BioDendro [16], start from spectrum data and depend on the clustering algorithm. Therefore, iPhylo Tree presents a swift and efficient

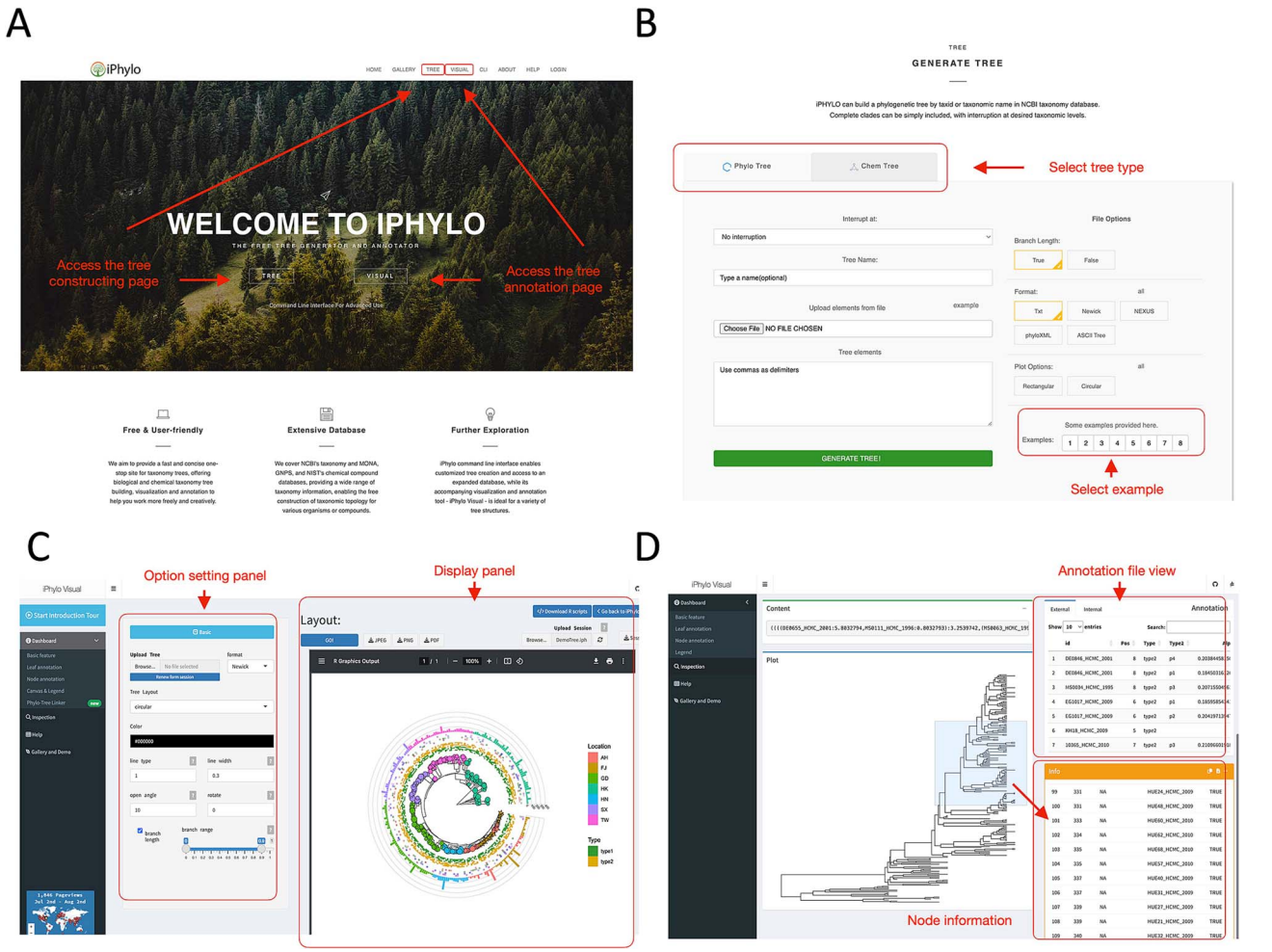


Figure 2. The user interface of the iPhylo suite. (A) The home page. (B) The tree construction page. (C) The annotation dashboard page. (D) The annotation inspection page.

alternative. It builds biological taxonomic trees based on the NCBI Taxonomy, enabling rapid retrieval of lineage information spanning all life domains. Similarly, it utilizes two chemical classification systems, ChemOnt and NPClassifier, to generate chemical taxonomic trees, thereby eliminating the need for extensive input and the impact of the algorithm. A complete pipeline of the iPhylo Tree includes the following steps:

Step 1: For constructing biological taxonomic trees, users provide a set of species names or corresponding taxonomy identifiers (TaxID). The species names are relatively flexible as long as they are indexed in the NCBI taxonomic database, and "scientific name," "synonym," or "common name" are all acceptable. For chemicals, the input can be InChIKey, InChI, and isomeric SMILES. Currently, common chemical names are not supported due to extensive heterogeneity and unnamed chemicals. The InChIKey is highly recommended because some chemicals lack InChI and isomeric SMILES information.

Step 2: Taxonomy querying involves querying input biological and chemical features against the iPhylo databases to extract taxonomic data. The iPhylo databases comprise biological and chemical divisions. The biological database includes the entire species classification information sourced from the NCBI taxonomy database, encompassing a total of 2 388 300 TaxIDs. The chemical database includes classification details for 801 308 functional compounds compiled from MassBank of North America (MoNA), Global Natural Product Social Molecular Networking (GNPS), and National Institute of Standards and Technology (NIST) databases, with a focus on human metabolome. The chemical taxonomic data are managed based on classification systems ChemOnt [5] and NPClassifier [6]. Moreover, the databases will be regularly updated to keep up with the expanding knowledge of biological and chemical features.

Step 3: To construct taxonomic trees, we used an object-oriented method to transform queried taxonomic data into an *n*-ary tree object, enabling easy representation in formats such as Newick. For output, we employed a depth-first post-order traversal algorithm to navigate the tree, preserving its topology as a Newick format string. Subsequently, iPhylo utilizes the phylo package from BioPython [26] to transform the Newick-formatted string to other formats, including Nexus and PhyloXML. Additionally, we provide an American Standard Code for Information Interchange (ASCII) representation and a Portable Document Format (PDF) visualization of the tree generated by the ggtree [17] package.

The resulting biological taxonomic tree is organized by a unified number of taxonomic levels: domain, kingdom, phylum, class, order, family, genus, and species, excluding informal ranks like subphylum and subclass. The chemical taxonomic tree, as structured by the ChemOnt classification system, organizes compounds hierarchically from the highest to the lowest levels: kingdom, superclass, class, subclass, parent level 1, and parent level 2. Similarly, the NPClassifier system categorizes

compounds into pathway, superclass, and class. Furthermore, iPhylo Tree is capable of filtering the tree by interrupting each branch at a specified level. The "|subtree" operator enables the extraction of a group of taxa that includes the common ancestor and all descendants within the specified clade, such as "Liliaceae|subtree" or "Cinnamic acid amides|subtree|subtree."

iPhylo CLI

The iPhylo CLI serves as an extension of the web-based iPhylo Tree. The iPhylo CLI includes several modules, namely, (i) the phylo tree module, (ii) the chemical tree module based on the local ChemOnt database, (iii) the chemical tree module based on the local NPClassifier database, (iv) the chemical tree module based on the online ClassyFire Application Programming Interface (API), (v) the chemical tree online module based on the online NPClassifier API, and (vi) the csv2tree module. The phylo tree module and chemical tree modules offer the same functionality as the online iPhylo Tree but run locally. Notably, the exclusive online chemical module enables access to chemical information via APIs from ClassyFire [5] and NPClassifier [6] API. ClassyFire's API currently supports queries for over 70 million chemicals. Similarly, NPClassifier's API allows for the retrieval of more than 1.3 million nature products, with both databases continually expanding. This feature significantly augments the capabilities of chemical taxonomy analysis. Additionally, the csv2tree module empowers users to create customized trees from Comma-separated Values (CSV) files directly.

The iPhylo CLI is tailored for efficiency and scalability. After its initial execution, which downloads the necessary database resources, the subsequent runs operate in an offline local mode. This design ensures rapid execution and is well suited for deployment on high-performance computing clusters (Supplementary Fig. S1).

iPhylo Visual

The iPhylo Visual is an interactive online tool designed to facilitate the display, annotation, and inspection of tree-based structures, including but not limited to phylogenetic and chemical taxonomic trees generated from iPhylo modules.

The iPhylo Visual simplifies the process of annotating taxonomic trees by adopting a data frame–compatible format, enabling users to encapsulate all required information within one data frame for leaf annotation and one data frame for node annotation, respectively. Within the data frame, rows correspond to tree nodes, and columns represent specific features. Users can efficiently navigate and manage these uploaded data frames through the provided online table viewer, with sorting and retrieval capabilities. This design avoids uploading multiple annotation files and is directly compatible with R.

The iPhylo Visual leverages the full graphical capabilities of ggtree [17] and ggtreeExtra [21] for visualizing, manipulating, and annotating tree-structured data. The iPhylo Visual provides the "Basic Feature," "Leaf Annotation," "Node Annotation," "Canvas and Legend," and "Phylo-tree Linker" tabs for visualization and annotation controls. Users can choose from multiple layouts (e.g. circular, rectangular; Supplementary Fig. S2), with adapting annotations. Customization options include branch thickness, color, and angle. The "Leaf Annotation" and "Node Annotation" tabs provide methods for adding specific markers to leaves and nodes, respectively (Supplementary Figs S3 and S4). Additionally, the "Legend" tab allows users to adjust the canvas size and legends for each annotation layer. Lastly, the "Phylo-tree Linker" integrates two taxonomic trees (e.g. biological and chemical trees) together, highlighting defined connections between them.

The iPhylo Visual emphasizes the ease of exporting and replicating tree displays with the ".iphylo" session files, packing all essential data for generating trees and maintaining compact sizes, e.g. 30 kB for a 1000-species tree. All sessions can be saved to and uploaded from the local computer, enabling easy storage, management, and collaboration.

iPhylo Visual supports exporting trees in PNG, PDF, and JPEG formats. Its "Export Code" feature allows a one-click download of visualization data, including the tree, annotations, parameters (JSON), and an R script. Running this script locally replicates the visualization, and users proficient with R programming can further modify the script.

Implementation

The iPhylo Tree web server stores taxonomy data with MariaDB 5.5.60 database. The back-end of the iPhylo Tree webserver was developed using Python and the Flask framework version 2.2.2. The iPhylo Visual uses the R/Shiny framework for advanced tree display and annotation. We employed the Twitter Bootstrap template version 4.6 for the front end to create a visually appealing and user-friendly graphical interface. The web-based applications are deployed on a server with 32-core CPUs (2 GHz), 96GB random-access memory (RAM), and 1TB of storage to ensure that the hosting is scalable with the large increase in use. The iPhylo CLI was developed alongside the server version, with Python 3.8.15 and SQLite 3.39.3.

Result management

The iPhylo suite offers free usage, including for commercial purposes, and does not require user registration. Besides, the iPhylo Tree module provides users with a login option to automatically save their generated tree history records. Notably, the registration does not require sensitive personal information, such as email addresses. The data for visualization and annotation in iPhylo Visual are managed with sessions. With the initiation of each new session within the user's web browser, we establish temporary storage on our server, each associated with a unique identifier, enabling user download activity. When the session terminates, these data are cleared to ensure privacy and security.

Database construction and management

The iPhylo suite utilizes NCBI Taxonomy for biological classification and chemical classification framework ClassyFire and NPClassifier for chemical classification. A local version of each database has been adapted for convenient accessibility through Structured Query Language (SQL) queries. For the NCBI Taxonomy database, we merge data originally distributed in a tab-separated format across separate dump files into complete lineages in a structured format. For chemical databases, we collect the compound classifications via the ClassyFire and NPClassifier APIs. These data are managed locally within SQL databases, enhancing data processing speed and capacity and reducing the risk of interruptions during server-side calls.

To keep the databases up to date, an automated process using custom Python scripts updates them every 6 months. For convenience, sqlite3 is employed in the iPhylo CLI module to avoid the complex configurations.

Statistical analysis

The majority of data processing and statistical analyses in case studies were done in Rstudio and R (at the time of writing,

2024.04.2 for Rstudio and 4.2.2 for R). The essential R packages used are tidyr (1.3.1), MetaNet (0.2.1), dplyr (1.1.4), pctax (0.1.1), ggtree (3.12.0), ape (5.8), reshape2 (1.4.4), devtools (2.4.5), and tidyverse (2.0.0).

In Case Study 3, Spearman correlation was used over the parametric counterparts due to the non-normality of our datasets. Significance was established using a correlation coefficient (R) >0.5, with P-values adjusted by the Benjamini & Hochberg method to control the false discovery rate at a threshold of <.05.

Results

To demonstrate the applications of the iPhylo suite, we constructed, visualized, and annotated a few phylogenetic and chemical taxonomic trees with data from published studies. The source data and iPhylo Visual session files for all case studies are available on the iPhylo website.

Case Study 1: Visualization of the genome catalog of diverse bacteria in the glacial microbiome

In this example, iPhylo was used to visualize the Tibetan glacier metagenome-assembled genomes (MAGs) and cultivated bacterial genomes [27]. The data comprised 2358 MAGs and 883 genomes from cultivated bacteria, resulting in a total of 3241 Tibetan glacier genomes. These genomes were clustered into 968 species-level operational taxonomic units. We utilized the iPhylo CLI to construct a customized tree that incorporated the Genome Taxonomy Database (GTDB) classification for each genome provided in the original study. To optimize the visualization process without compromising the display and functionality of the system, we sampled and formed a subtree from the full tree. This optimization resulted in a phylogenetic tree with 440 tips and 174 internal nodes. We next used iPhylo Visual to display and annotate the tree with various genome characteristics (Fig. 3), including the source of the genome, presence of 16S rRNA, number of tRNA genes, genome size, GC proportion, and genome quality standards based on the MAG [28]. We displayed the resulting figure in a style similar to the iTOL tree in the original study, but with a more streamlined and user-friendly approach. Instead of relying on specialized annotation templates for different dimensions, we prepared all the annotation data in spreadsheet format and configured the display interactively. This approach sets iPhylo apart from iTOL, making the annotation process more user-friendly and easier to replicate.

Case Study 2: Visualization and analysis of metabolic profiles for individual bacterial strains in the human gut microbiome

Gut microbes are associated with multiple metabolic pathways and play a significant role in modifying host phenotypes and overall health [29, 30]. Han *et al.* [31] reported a comprehensive metabolic profile of gut microbes encompassing 158 microbial strains and 833 metabolites. In this study, the culture of each microbial species was subjected to chemical detection using liquid chromatography–mass spectrometry (LC-MS). From this data, we selected 154 strains (the taxonomy of which was supported by NCBI Taxonomy) and 813 metabolites (chemical taxonomy supported by the ClassyFire database) for downstream analyses.

We first constructed the chemical taxonomic tree of the 813 metabolites produced by the microbes. This provided an ontology overview of 11 superclasses, 84 classes, and 157 subclasses of chemicals (Fig. 4A). Next, we constructed a phylogenetic tree for the strains producing selected metabolites, namely, putrescine, guanidine, ornithine, and citrulline. These four metabolites were highlighted in the original study as potential indicators for measuring metabolic phenotypes of strains. We used iPhylo Visual to plot the abundance of these selected metabolites in all strains using bubble plots (Fig. 4B). Additionally, we employed the heatmap to illustrate the taxonomy of microbes and their full metabolomic profiles (Supplementary Fig. S5).

To further investigate potential associations between microbes and metabolites, we employed the iPhylo Visual's phylo-tree linker to visualize the biosynthetic relationships between the gut microbes and the metabolites. Homovanillic acid (HVA) is a major catecholamine metabolite produced through the enzymatic actions of dopamine [32]. Research indicates that certain gut microbes can synthesize HVA, which has the capacity to inhibit autophagic cell death, thereby restoring synaptic function and potentially alleviating depression symptoms [33]. Our analysis identifies selected Firmicutes strains as the producers of HVA, underscoring their potential impact on neuro-metabolic pathways (Fig. 5A). Tartaric acid, a naturally occurring organic acid, is widely used as a food additive and serves as a crucial chiral-resolving agent in the pharmaceutical industry [34, 35]. Various studies and patents have documented that bacteria such as *Corynebacterium* sp., *Rhodococcus* sp., *Alcaligenes levotartaricus*, *Acinetobacter tartaricus*, and *Pseudomonas agrobacterium* exhibit high enzymatic activity through stereo-specific *cis*-epoxysuccinic acid hydrolase to produce L-(+)-tartaric acid [36]. Our results indicate that tartaric acid producers exclusively belong to the Bacteroidetes phylum (Fig. 5B). This discovery highlights the significant role of Bacteroidetes in the production of tartaric acid and suggests the need for further research to explore their biosynthetic pathways. Coproporphyrin III (CP III), a natural porphyrin derivative, plays a crucial role in the synthesis of heme and bilirubin [37]. In this case study, we found that CP III synthesis is confined to the Bacteroidetes and Firmicutes phyla, with no production observed in other phyla (Fig. 5C).

This case study highlights the versatility of the iPhylo suite in integrating and visualizing both chemical and biological taxonomic trees. When incorporated with abundance data on metabolite production, exclusive microbe–metabolite associations can be identified and visualized with ease.

Case Study 3: Exploring the associations between the human microbiome and metabolome using iPhylo

In this case study, we apply the iPhylo suite to analyze the interactions between the human microbiome and metabolome. We utilized longitudinal data from Zhou *et al.* [38] on the profiling of the human microbiome and corresponding metabolome. We constructed trees for the microbiome and metabolome quantitative data. The NP Classifier taxonomy was employed for metabolites. Additionally, we calculated correlations between microbial genera and host metabolites to identify potential microbe–metabolite links.

Our analysis revealed massive significant positive and negative correlations between specific microbes and metabolites (Fig. 6A), suggesting potential biological interactions. For example, we identified a positive correlation between alkaloids such as 1-methylxanthine and paraxanthine with the bacterial genus Phocaeicola. This finding is consistent with the study by Zhong *et al.* [39], suggesting that alkaloids may promote the growth and proliferation of Phocaeicola. Diundecyl phthalate is a type of phthalate ester. Research has shown that phthalates can be absorbed into the blood and fluids and impair physiological mechanisms, thus inducing significant endocrine disruption [40].

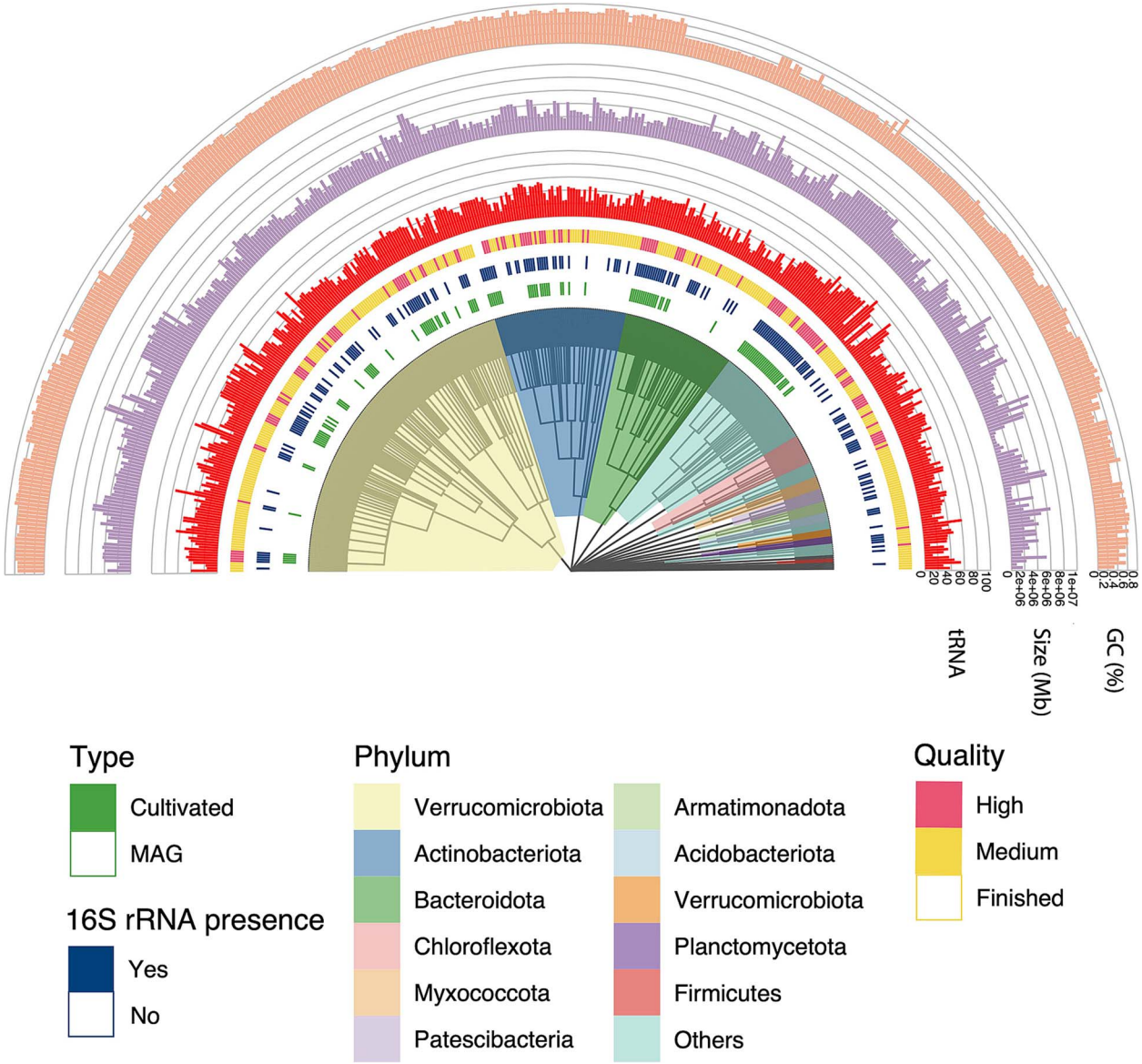


Figure 3. iPhylo visualizes phylogenetic trees incorporating extensive annotation tracks. The tree includes genome type (cultivated or MAG), 16S rRNA presence, phylum classification, and genome quality. Additional tracks show tRNA count, genome size, and GC content. Data source: Liu *et al.* [27].

Several studies have shown that the degradation of phthalates in the environment is predominantly carried out by strictly aerobic bacteria such as *Arthrobacter* sp. [41], *Acinetobacter* sp. [42], and *Pseudomonas* sp. [43], with a smaller proportion involving facultative anaerobes like *Bacillus* sp. [44], *Serratia* sp. [45, 46], and *Enterobacter* sp. [47]. In our study, we identified several bacterial genera that show positive correlations with diundecyl phthalate, namely, Dorea, Anaerobutyricum, Raoultibacter, Ihubacter, and Adlercreutzia. Our findings suggest that these gut microbes likely possess the capability to metabolize phthalates in humans.

Importantly, we discovered that several microbial genera negatively correlated with tryptophan metabolism. Raoultibacter, Adlercreutzia, and Gordonibacter showed negative correlations with L-formylkynurenine, which is an intermediate in the kynurenine pathway of tryptophan degradation and serves as a substrate for several subsequent reactions within this pathway. Similarly, Alistipes, Barnesiella, and Raoultibacter were negatively correlated with indoleacetyl glutamine (IAG), a derivative of the tryptophan metabolism breakdown product indole-3-acetic acid (IAA). These observations are supported by Jiang *et al.* [48], indicating that these microbes might participate in tryptophan metabolism and its subsequent breakdown [49]. Besides, our analysis shows that Agathobacter is positively correlated with indoleacetyl glutamine and indolelactic acid (ILA, a derivative of the intermediate indolepyruvate formed during tryptophan metabolism). Agathobacter is capable of producing tryptophan synthase, which catalyzes the synthesis of L-tryptophan from its precursor indole-3-glycerol phosphate [50]. Hence, the positive correlation suggests a causative relationship where Agathobacter actively contributes to the production of IAG and ILA through its enzymatic activity through the production of tryptophan synthase.

Case Study 4: Exploring the associations between environmental airborne biological and chemical exposomes using iPhylo

We next extended the application of the iPhylo suite to environmental exposome research beyond human subjects. Exposome depicts the total environmental exposures humans are exposed to in short and long timeframes [51, 52]. Huang *et al.* employed

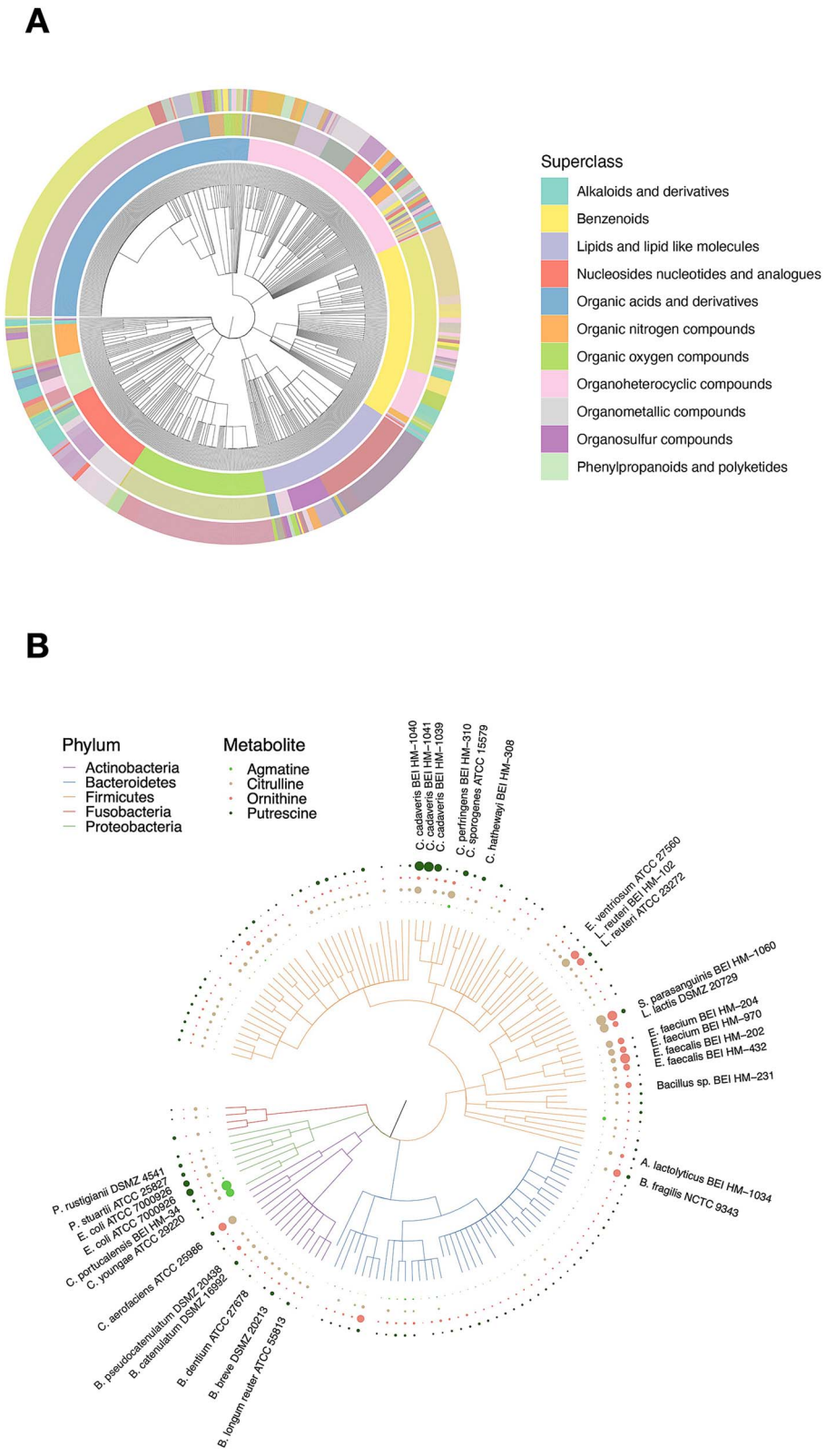


Figure 4. Chemical and annotated phylogenetic trees constructed by iPhylo for metabolite biosynthetic profiles of individual bacterial strains. (A) The representative chemical taxonomic tree generated based on chemical InChIKeys and annotated at the superclass, class, and subclass levels (from inside to outside, only legends for superclasses are shown). (B) The phylogenetic tree constructed using NCBI taxonomy IDs, with branches color-coded by phyla. The tree was further annotated with the abundance of four key metabolites: agmatine, citrulline, ornithine, and putrescine. The abundance was calculated from fold change data with internal standard correction.

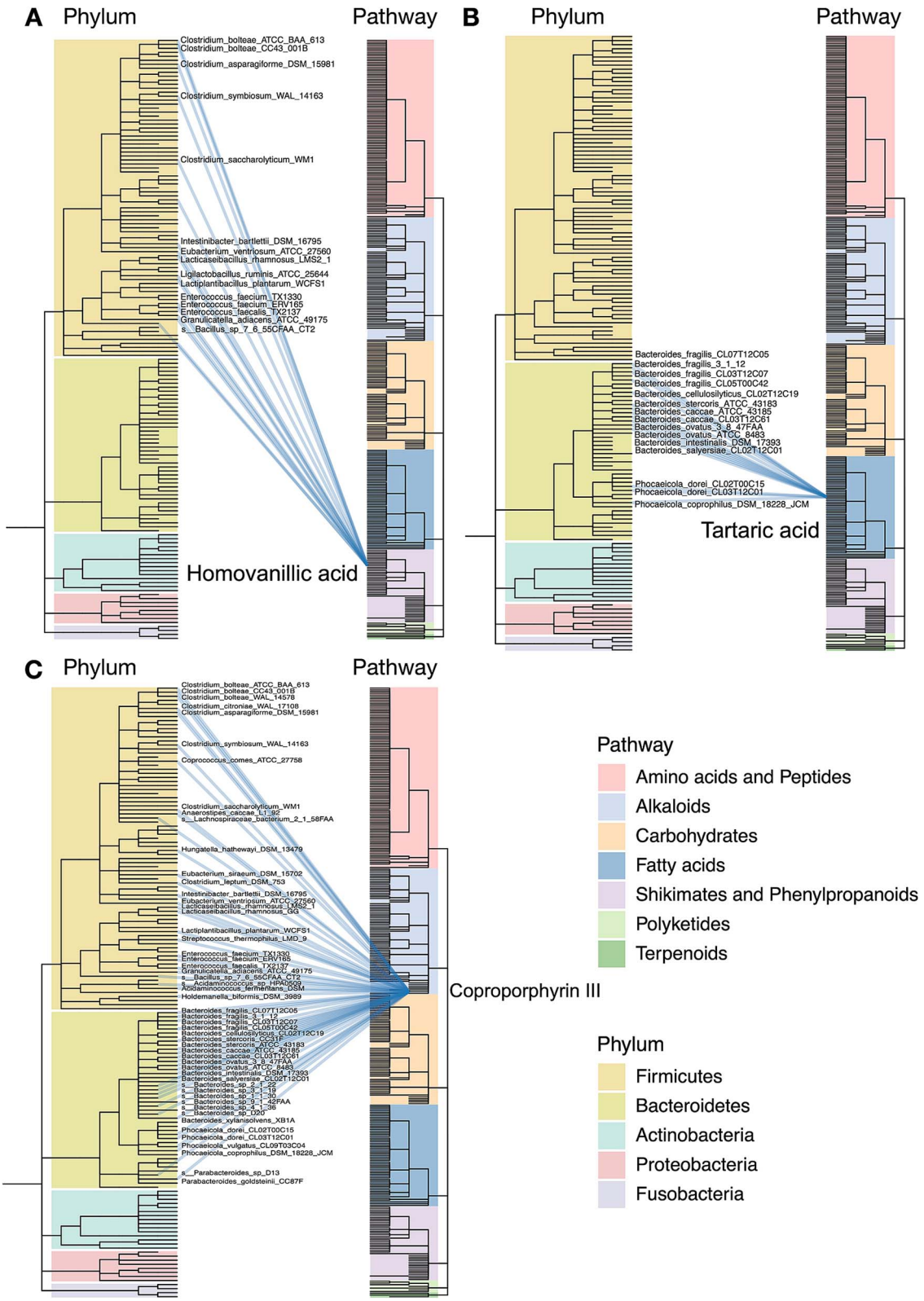


Figure 5. Examples of phylum-specific production of metabolites displayed using the iPhylo Visual's phylo-tree linker. (A) HVA is exclusively produced by the Firmicutes phylum. (B) Tartaric acid is exclusively produced by the Bacteroidetes phylum. (C) Coproporphyrin III is synthesized specifically by the Bacteroidetes and Firmicutes phyla.

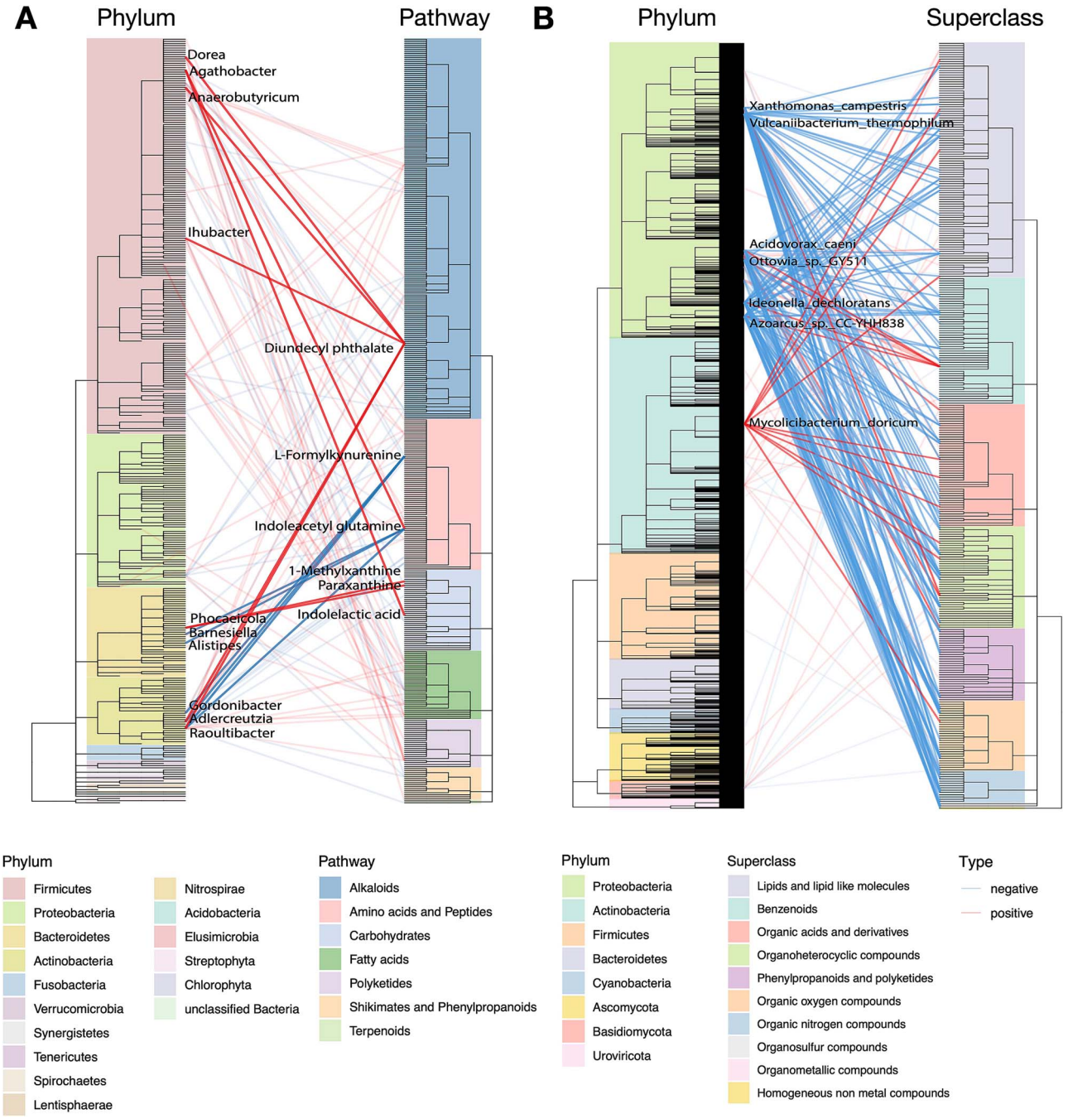


Figure 6. Visualizations of the correlations in human microbiome and metabolome and air exposome by iPhylo Visual. (A) Significant correlations between the human microbiome and metabolome visualized by iPhylo Visual's phylo-tree linker. Correlations with $R > 0.5$ and Adj. $P < .05$ are shown. (B) Significant correlations between the biological and chemical components of the air exposome visualized by iPhylo Visual's phylo-tree linker. Correlations with $R > 0.7$ and Adj. $P < .05$ are shown.

wearable silicone-based passive samplers to capture DNA and airborne chemicals in an underwater confined environment [53]. The chemical and biological exposomes were analyzed using liquid chromatography with high-resolution tandem mass spectrometry (LC-HRMS/MS), gas chromatography with high-resolution mass spectrometry (GC-HRMS), and metagenomic shotgun sequencing.

We calculated the correlations between biological and chemical exposomes to identify potential interactions between microbes and metabolites. A significant positive correlation ($R > 0.7$ and Adj. $P < .05$) was found between Mycolicibacterium doricum and superclasses of organic acids and derivatives, organoheterocyclic compounds, and lipids and lipid-like molecules (Fig. 6B). This finding is consistent with existing research indicating that the activation phase of Mycolicibacterium involves the biosynthesis of secondary metabolites, purine and pyrimidine metabolism, glycerophospholipid, and fatty acid metabolism [54]. The observed positive correlations suggest that *Mycolicibacterium doricum* thrives in environments rich in these compounds, likely due to their abundance in the air exposome. In addition, we identified selected microorganisms exhibiting negative correlations with various metabolites (Fig. 6B). For instance, the Azoarcus genus is reported to show specificity in metabolizing aromatic compounds [55], while *Ideonella dechloratans* is known to metabolize chlorate [56, 57]. *Ottowia* sp. can utilize phenylacetic

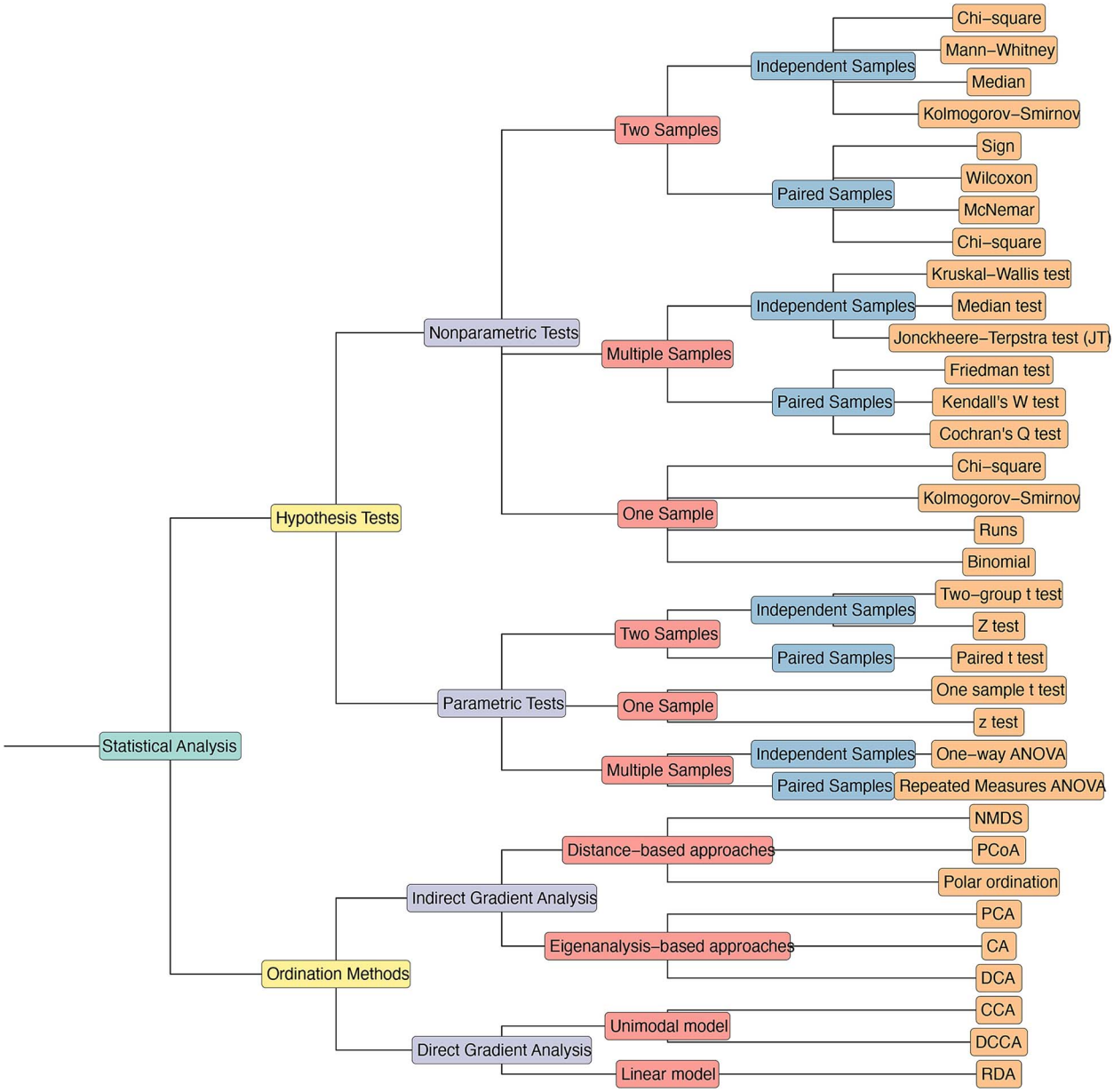


Figure 7. The iPhylo visualizes user-customized trees. A tree of common statistical methods was visualized as an example. We employed the iPhylo CLI module to organize the classifications of these statistical methods into a hierarchical tree structure and displayed it using iPhylo Visual. The tree is annotated with node labels that are color-coded based on taxonomic ranks.

acid as the sole nitrogen source to degrade phenylacetic acid [58]. These specific metabolic activities suggest that the degradation of organic pollutant compounds in the environment could be a reason for these negative correlations. Additionally, the presence of certain metabolites may inhibit the growth of microbes such as *Xanthomonas campestris*, *Vulcaniibacterium thermophilum*, and *Acidovorax caeni*, further contributing to the observed negative correlations.

These results highlight the capability of the iPhylo suite to illustrate complex interactions between microbiome and metabolome in humans and the environment.

Case Study 5: A user-customized tree of statistical analysis methods

The iPhylo suite is compatible with any customized hierarchical relational database, as well as user-defined trees. Statistical methods can be classified into different groups based on analytical purposes, such as hypothesis testing, ordination, factor analysis, clustering analysis, and correlation analysis. For demonstration purposes, we have selected several commonly used statistical methods in the field of bioinformatics. The hierarchical relationships for these methods were organized in CSV format (Supplementary Table 2) and processed by iPhylo CLI, resulting in a tree of common statistical methods and visualized in the iPhylo Visual (Fig. 7).

Discussion

As multi-omic studies rapidly develop, there is a growing demand for comprehensive analysis tools capable of handling both biological and chemical data [53, 59, 60]. Several tools for biological taxonomic analysis have been developed; none provide a fully integrated solution for both biological and chemical taxonomies. Additionally, some of these tools require monetary

subscriptions, limiting accessibility. The iPhylo suite aims to address these issues by providing a fast and convenient solution for building and annotating taxonomic trees. The iPhylo suite was developed with several principles: (i) cross-platform compatibility. The iPhylo suite offers web-based and command-line services accessible on Windows, Mac OS, and Linux. (ii) Integration. iPhylo seamlessly integrates the entire workflow of constructing, visualizing, and annotating taxonomic trees without switching platforms, thus minimizing potential errors. (iii) Customizability. iPhylo offers extensive customization options and extends its utility to all hierarchical relational databases.

The data frame format was adopted for annotation to minimize the input requirements from users. Additionally, users can download and upload tree-building sessions and access the underlying R plotting code for further modifications. Finally, it is worth noting that by leveraging the integrative functionality of the iPhylo suite, we were able to combine chemical classifications with biological taxonomies. This integrated approach afforded a comprehensive multi-omic perspective as demonstrated in Case Studies 2, 3, and 4, revealing critical insights into the biosynthetic capabilities of diverse microbes and the complex interactions between the human microbiome and metabolome, as well as between the biological and chemical exposomes. This unique capability, currently absent in other comparable tools, offers a convenient and innovative approach to investigating microbial–metabolite interactions and potential causal relationships. This approach also holds significant potential for a wide range of applications in environmental and exposome studies [61, 62].

Despite the strengths of iPhylo, some limitations apply. Currently, the iPhylo suite does not support phylogenetic trees computed from sequence alignments, as this approach requires an input of sequence alignment for all species considered and can be difficult to apply to multikingdom trees. At present, the embedded chemical database focuses on functional compounds, covering only a small subset of the vast array of chemical substances. This limitation was mitigated by the ability to query online data resources in iPhylo CLI, although the advantage of high querying speed in the locally embedded chemical database remains irreplaceable. Furthermore, other databases, such as the GTDB dataset, also present opportunities for future expansion. Looking forward, the iPhylo suite's development roadmap includes plans to incorporate a broader range of databases not limited to biologicals and chemicals. We believe that the iPhylo suite will greatly advance the general adoption of biological and chemical taxonomic analyses and their integration in diverse fields such as microbiome, metabolome, precision medicine, ecology, and environmental science.

Key Points

- A platform for both biological and chemical taxonomic analysis and visualizations.
- A stand-alone command line interface module for local or HPC deployment.
- R-based interactive visualization module with session-saving options.
- Supports multi-omic integrative analysis and customized hierarchical databases for analysis beyond taxonomy.

Acknowledgements

We are grateful to our colleagues at the core facility of the Life Sciences Institute, especially the NECHO high-performance computing cluster.

Author contributions

Yueer Li: Conceptualization, Methodology, Investigation, Software, Visualization, Formal Analysis, Validation, Writing—Original Draft and Writing—Review & Editing. **Chen Peng:** Software, Visualization, Validation, Writing—Review & Editing. **Fei Chi:** Resources, Writing—Review & Editing. **Zinuo Huang:** Data Curation, Writing—Review & Editing. **Mengyi Yuan:** Data Curation, Writing—Review & Editing. **Xin Zhou:** Data Curation, Writing—Review & Editing. **Chao Jiang:** Conceptualization, Methodology, Investigation, Writing—Original Draft, Writing—Review & Editing, Supervision, Project Administration and Funding Acquisition.

Conflicts of Interest: The authors declare no conflicts of interest.

Funding

This research was supported by grants from the National Natural Science Foundation of China (NSFC) (82173645 and 82341109).

Data availability

All the demonstration data for the case studies can be accessed on the iPhylo Visual website's Help page (https://iphylo.net/visual). The iPhylo database is available for download as a ".db" file from the following URLs:

Biological database: https://iphylo.net/resource/iphylo_db

Chemical Database for the ChemOnt system: https://iphylo.net/resource/ichem_db

Chemical Database for the NPClassifier system: https://iphylo.net/resource/ichem_np_db

Code availability

The iPhylo CLI is open-source and can be found on GitHub (https://github.com/ARise-fox/iPhylo-CLI). Additionally, the raw data and corresponding preprocessing R script can be downloaded from the GitHub repository (https://github.com/ARise-fox/iPhylo-CaseStudyData).

References

1. Federhen S. The NCBI taxonomy database. *Nucleic Acids Res* 2012;**40**:D136–43. https://doi.org/10.1093/nar/gkr1178.
2. Stoesser G, Baker W, van den Broek A. *et al.* The EMBL nucleotide sequence database. *Nucleic Acids Res* 2002;**30**:21–6. https://doi.org/10.1093/nar/30.1.21.
3. Degtyarenko K, de Matos P, Ennis M. *et al.* ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 2008;**36**:D344–50. https://doi.org/10.1093/nar/gkm791.
4. Fahy E, Subramaniam S, Murphy RC. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* 2009;**50 Suppl**:S9–14. https://doi.org/10.1194/jlr.R800095-JLR200.
5. Djoumbou Feunang Y, Eisner R, Knox C. *et al.* ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *J Chem* 2016;**8**:61.

6. Kim HW, Wang M, Leber CA. *et al.* NPClassifier: A deep neural network-based structural classification tool for natural products. *J Nat Prod* 2021;**84**:2795–807. https://doi.org/10.1021/acs.jnatprod.1c00399.
7. Kumar S, Stecher G, Li M. *et al.* MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;**35**:1547–9. https://doi.org/10.1093/molbev/msy096.
8. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, analysis, and visualization of Phylogenomic data. *Mol Biol Evol* 2016;**33**:1635–8. https://doi.org/10.1093/molbev/msw046.
9. Schliep KP. Phangorn: Phylogenetic analysis in R. *Bioinformatics* 2011;**27**:592–3. https://doi.org/10.1093/bioinformatics/btq706.
10. G C. *Taxtree: An R Package for Generating a Topology of Species Evolutionary Relationships Based on the Taxonomy Database*. R package version 0.1.0, 2024; Available from: https://github.com/nongxinshengxin/taxtree.
11. Minh BQ, Schmidt HA, Chernomor O. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;**37**:1530–4. https://doi.org/10.1093/molbev/msaa015.
12. Letunic, I. *phyloT: A Tree Generator*. Biobyte.de 2019; Available from: https://phylot.biobyte.de/.
13. Lu J, Carlson HA. ChemTreeMap: An interactive map of biochemical similarity in molecular datasets. *Bioinformatics* 2016;**32**:3584–92. https://doi.org/10.1093/bioinformatics/btw523.
14. Tripathi A, Vázquez-Baeza Y, Gauglitz JM. *et al.* Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. *Nat Chem Biol* 2021;**17**:146–51. https://doi.org/10.1038/s41589-020-00677-3.
15. Depke T, Franke R, Bronstrup M. CluMSID: An R package for similarity-based clustering of tandem mass spectra to aid feature annotation in metabolomics. *Bioinformatics* 2019;**35**:3196–8. https://doi.org/10.1093/bioinformatics/btz005.
16. Rawlinson C, Jones D, Rakshit S. *et al.* Hierarchical clustering of MS/MS spectra from the firefly metabolome identifies new lucibufagin compounds. *Sci Rep* 2020;**10**:6043. https://doi.org/10.1038/s41598-020-63036-1.
17. Yu G. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinformatics* 2020;**69**:e96. https://doi.org/10.1002/cpbi.96.
18. Letunic I, Bork P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;**49**:W293–6. https://doi.org/10.1093/nar/gkab301.
19. Xie J, Chen Y, Cai G. *et al.* Tree visualization by one table (tvBOT): A web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res* 2023;**51**:W587–92. https://doi.org/10.1093/nar/gkad359.
20. Chang W, Cheng J, Allaire J. *et al.* Shiny: Web Application Framework for R. R package version 1.10.0, 2024; https://shiny.posit.co/.
21. Xu S, Dai Z, Guo P. *et al.* ggtreeExtra: Compact visualization of richly annotated phylogenetic data. *Mol Biol Evol* 2021;**38**:4039–42. https://doi.org/10.1093/molbev/msab166.
22. Mishra S, Bhargava P, Adhikary SP. *et al.* Weighted morphology: A new approach towards phylogenetic assessment of Nostocales (cyanobacteria). *Protoplasma* 2015;**252**:145–63. https://doi.org/10.1007/s00709-014-0629-9.
23. de Queiroz AWPH, Wimberger PH. The usefulness of behavior for phylogeny estimation: Levels of HOMOPLASY IN BEHAVIORAL and morphological characters. *Evolution* 1993;**47**:46–60. https://doi.org/10.2307/2410117.
24. Sherbakov DY, Triboy TI. Effect of co-evolving amino acid residues on topology of phylogenetic trees. *Biochemistry (Mosc)* 2007;**72**:1363–7. https://doi.org/10.1134/S0006297907120103.
25. Drummond AJ, Suchard MA, Xie D. *et al.* Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;**29**:1969–73. https://doi.org/10.1093/molbev/mss075.
26. Cock PJ, Antao T, Chang JT. *et al.* Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;**25**:1422–3. https://doi.org/10.1093/bioinformatics/btp163.
27. Liu Y, Ji M, Yu T. *et al.* A genome and gene catalog of glacier microbiomes. *Nat Biotechnol* 2022;**40**:1341–8. https://doi.org/10.1038/s41587-022-01367-2.
28. The Genome Standards Consortium, Bowers RM, Kyrpides NC. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017;**35**:725–31. https://doi.org/10.1038/nbt.3893.
29. Mahmud MR, Akter S, Tamanna SK. *et al.* Impact of gut microbiome on skin health: Gut-skin axis observed through the lenses of therapeutics and skin diseases. *Gut Microbes* 2022;**14**:2096995. https://doi.org/10.1080/19490976.2022.2096995.
30. Stoeva MK, Garcia-So J, Justice N. *et al.* Butyrate-producing human gut symbiont, clostridium butyricum, and its role in health and disease. *Gut Microbes* 2021;**13**:1–28. https://doi.org/10.1080/19490976.2021.1907272.
31. Han S, van Treuren W, Fischer CR. *et al.* A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature* 2021;**595**:415–20. https://doi.org/10.1038/s41586-021-03707-9.
32. Lambert GW, Eisenhofer G, Jennings GL. *et al.* Regional homovanillic acid production in humans. *Life Sci* 1993;**53**:63–75. https://doi.org/10.1016/0024-3205(93)90612-7.
33. Zhao ML, Ren Z, Zhao A. *et al.* Gut bacteria-driven homovanillic acid alleviates depression by modulating synaptic integrity. *Cell Metab* 2024;**36**:1000-12.e6. https://doi.org/10.1016/j.cmet.2024.03.010.
34. Li MH, Su J, Yang H. *et al.* Grape tartaric acid: Chemistry, function, metabolism, and regulation. *Horticulturae* 2023;**9:1173**. https://doi.org/10.3390/horticulturae9111173.
35. Sui J, Wang N, Wang J. *et al.* Strategies for chiral separation: From racemate to enantiomer. *Chem Sci* 2023;**14**:11955–2003. https://doi.org/10.1039/D3SC01630G.
36. Ukaji Y, Soeta T. 3.6 Acetogenin (Polypriopionate) derived auxiliaries: Tartaric acid. *Comprehensive Chirality*. Elsevier, 2012;**3**:176–201. https://doi.org/10.1016/B978-0-08-095167-6.00306-2.
37. Guo Q, Xu J, Li J. *et al.* Synergistic increase in coproporphyrin III biosynthesis by mitochondrial compartmentalization in engineered Saccharomyces cerevisiae. *Synth Syst Biotechnol* 2024;**9**:834–41. https://doi.org/10.1016/j.synbio.2024.07.001.
38. Zhou X, Shen X, Johnson JS. *et al.* Longitudinal profiling of the microbiome at four body sites reveals core stability and individualized dynamics during health and disease. *Cell Host Microbe* 2024;**32**:506–26.e9. https://doi.org/10.1016/j.chom.2024.02.012.
39. Zhong S, Yang Y, Huo J. *et al.* Dissection of gut microbiota and metabolites reveals the hypolipidemic effect of green mulberry leaf tea / black mulberry leaf tea in mice. *J Funct Foods* 2023;**111**:105906. https://doi.org/10.1016/j.jff.2023.105906.
40. Miodovnik A, Engel SM, Zhu C. *et al.* Endocrine disruptors and childhood social impairment. *Neurotoxicology* 2011;**32**:261–7. https://doi.org/10.1016/j.neuro.2010.12.009.
41. Kong X, Bai Z, Jin T. *et al.* Arthrobacter is a universal responder to di-n-butyl phthalate (DBP) contamination in soils from various geographical locations. *J Hazard Mater* 2022;**422**:126914. https://doi.org/10.1016/j.jhazmat.2021.126914.

42. Vinay K, Maitra SS. Efficient degradation of Dibutyl phthalate and utilization of phthalic acid esters (Paes) by species isolated from Msw (municipal solid waste) leachate. *Global NEST J* 2016;**18**:817–30. https://doi.org/10.30955/gnj.002028.
43. Kumar V, Sharma N, Maitra SS. Comparative study on the degradation of dibutyl phthalate by two newly isolated pseudomonas sp. V21b and Comamonas sp. 51F. *Biotechnol Rep (Amst)* 2017;**15**:1–10. https://doi.org/10.1016/j.btre.2017.04.002.
44. Xu Y, Liu X, Zhao J. *et al.* An efficient phthalate ester-degrading Bacillus subtilis: Degradation kinetics, metabolic pathway, and catalytic mechanism of the key enzyme. *Environ Pollut* 2021;**273**:116461. https://doi.org/10.1016/j.envpol.2021.116461.
45. Gupta A, Kumar M, Thakur IS. Analysis and optimization of process parameters for production of polyhydroxyalkanoates along with wastewater treatment by Serratia sp. ISTVKR1. *Bioresour Technol* 2017;**242**:55–9. https://doi.org/10.1016/j.biortech.2017.03.110.
46. Ningthoujam R, Satiraphan M, Sompongchaiyakul P. *et al.* Bacterial community shifts in a di-(2-ethylhexyl) phthalate-degrading enriched consortium and the isolation and characterization of degraders predicted through network analyses. *Chemosphere* 2023;**310**:136730. https://doi.org/10.1016/j.chemosphere.2022.136730.
47. Lamraoui I, Eltoukhy A, Wang J. *et al.* Biodegradation of Di (2-Ethylhexyl) phthalate by a novel Enterobacter spp. strain YC-IL1 isolated from polluted soil, Mila, Algeria. *Int J Environ Res Public Health* 2020;**17**:7501. https://doi.org/10.3390/ijerph17207501.
48. Jiang Y, Yang J, Xia L. *et al.* Gut microbiota-tryptophan metabolism-GLP-1 Axis participates in beta-cell regeneration induced by Dapagliflozin. *Diabetes* 2024;**73**:926–40. https://doi.org/10.2337/db23-0553.
49. Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. *Nat Commun* 2018;**9**:3294. https://doi.org/10.1038/s41467-018-05470-4.
50. Sheridan PO, Walker AW, Duncan SH. *et al. Resolving the taxonomy of Roseburia spp., Eubacterium rectale and Agathobacter spp. through phylogenomic analysis. Submitted to EMBL/GenBank/DDBJ databases*. https://www.uniprot.org/citations/CIB6QN0RV324AF0.
51. Vermeulen R, Schymanski EL, Barabási AL. *et al.* The exposome and health: Where chemistry meets biology. *Science* 2020;**367**:392–6. https://doi.org/10.1126/science.aay3164.
52. Miller GW. The exposome: purpose, definitions, and scope. In: Miller GW (ed). *The Exposome*. San Diego: Academic Press, an Imprint of Elsevier, 2020, 1–26. https://doi.org/10.1016/B978-0-12-814079-6.00001-8.
53. Huang Z, Peng C, Rong Z. *et al.* Longitudinal mapping of personal biotic and abiotic Exposomes and transcriptome in underwater confined space using wearable passive samplers. *Environ Sci Technol* 2024;**58**:5229–43. https://doi.org/10.1021/acs.est.3c09379.
54. Wang ZK, Qiu H, Chen Y. *et al.* Microbial metabolism of diosgenin by a novel isolated sp. HK-90: A promising biosynthetic platform to produce 19-carbon and 21-carbon steroids. *Microb Biotechnol* 2024;**17**:e14415. https://doi.org/10.1111/1751-7915.14415.
55. Fernández H, Prandoni N, Fernández-Pascual M. *et al.* Azoarcus sp. CIB, an anaerobic biodegrader of aromatic compounds shows an endophytic lifestyle. *PLoS One* 2014;**9**:e110771. https://doi.org/10.1371/journal.pone.0110771.
56. Stenklo K, Danielsson Thorell H, Bergius H. *et al.* Chlorite dismutase from Ideonella dechloratans. *J Biol Inorg Chem* 2001;**6**:601–7. https://doi.org/10.1007/s007750100237.
57. Thorell HD, Stenklo K, Karlsson J. *et al.* A gene cluster for chlorate metabolism in Ideonella dechloratans. *Appl Environ Microbiol* 2003;**69**:5585–92. https://doi.org/10.1128/AEM.69.9.5585-5592.2003.
58. Pang FH, Yang HY, Sun J. *et al.* Ottowia caeni sp. nov., a novel phenylacetic acid degrading bacterium isolated from sludge. *Int J Syst Evol Microbiol* 2021;**71**:005144. https://doi.org/10.1099/ijsem.0.005144.
59. Gao P, Shen X, Zhang X. *et al.* Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. *Genome Res* 2022;**32**:1199–214. https://doi.org/10.1101/gr.276521.121.
60. Li F, Jiang L, Pan S. *et al.* Multi-omic profiling reveals that intra-abdominal-hypertension-induced intestinal damage can Be prevented by microbiome and metabolic modulations with 5-Hydroxyindoleacetic acid as a diagnostic marker. *mSystems* 2022;**7**:e0120421. https://doi.org/10.1128/msystems.01204-21.
61. Wei X, Huang Z, Jiang L. *et al.* Charting the landscape of the environmental exposome. *iMeta* 2022;**1**:e50. https://doi.org/10.1002/imt2.50.
62. Jiang C, Wang X, Li X. *et al.* Dynamic human environmental exposome revealed by longitudinal personal monitoring. *Cell* 2018;**1**:277–291.e31. https://doi.org/10.1016/j.cell.2018.08.060.

Briefings in Bioinformatics, 2025, **26(1)**, bbae679
https://doi.org/10.1093/bib/bbae679
Problem Solving Protocol