MetaTree: an interactive web platform for hierarchical data visualization and multi-group comparison

Qing Wu[1, 2 †], Ailing Zhang[1, 2 †], Zhibin Ning[1], Daniel Figeys[1, 3, 4]

[1] School of Pharmaceutical Sciences, Faculty of Medicine, University of Ottawa, Ottawa, Canada

[2] Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Canada

[3] Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk, UK.

[4] University of East Anglia, Norwich, Norfolk, UK.

[†] These authors contributed equally to this work.

[*] Correspondence: Daniel Figeys, daniel.figeys@quadram.ac.uk

Abstract

Motivation

Hierarchically structured data are common in biology but are difficult to visualize in a way that both preserves structure and supports systematic comparison of many samples and experimental groups. Conventional approaches either discard the hierarchy or focus on single annotated trees, making it difficult to consistently compare clade-level patterns across multiple samples or treatments. There is a need for a browser-based framework for interactive visualization and multi-group comparison through hierarchical trees.

Results

We present MetaTree, an entirely browser-based platform for interactive visualization of hierarchical quantitative data, mapping quantitative values onto tree color, node size and branch width. MetaTree requires no installation or programming and accepts both hierarchical abundance tables and statistical result tables. It provides dedicated analysis modes that together support single-sample exploration, group-level summaries, pairwise group comparisons and automated all-versus-all comparison matrices on a shared reference tree. MetaTree also supports flexible figure arrangement and export in raster and vector formats, making tree-based comparative analysis accessible to any users.

Availability and Implementation

MetaTree is freely accessible at https://byemaxx.github.io/MetaTree and its source code and documentation are available on GitHub.

1 Introduction

Many biological datasets, particularly taxonomic classifications and hierarchical functional category systems (Ashburner, et al., 2000; Kanehisa, et al., 2025; Parks, et al., 2018), are organized across multiple levels. In microbiome workflows such as 16S rRNA sequencing (Ward, et al., 1990), shotgun metagenomics (Quince, et al., 2017) and metaproteomics (Wilmes and Bond, 2006), thousands of features are mapped onto such hierarchies. Understanding how clades respond to perturbations requires visualization methods that preserve this structure (Moon, et al., 2019). However, common approaches have significsnt limitations. Stacked bar charts require aggregation to coarse taxonomic ranks and are constrained by the small number of distinguishable colors, which obscures important patterns at deeper, biologically relevant levels (Foster, et al., 2017; McMurdie and Holmes, 2013). Clustered heatmaps, even when accompanied by dendrograms, visually separate the tree from the quantitative matrix, forcing users to mentally reconstruct their relationship and quickly becoming cluttered as feature counts increase (Wilke, 2019).

Several tools exist for visualizing phylogenetic trees, including PhyD3 (Kreft, et al., 2017), GraPhlAn (Asnicar, et al., 2015), ETE Toolkit (Huerta-Cepas, et al., 2016) and iTOL (Letunic and Bork, 2024), but they are primarily designed for static tree rendering rather than systematic comparison of multiple samples or groups. Most require programming or configuration files, and even user-friendly platforms such as iTOL provide only limited mechanisms for encoding quantitative information (Ondov, et al., 2011). Critically, these tools do not offer a native approach to display many samples or experimental groups on a shared tree topology and to compare them directly. Heat-tree visualizations address part of this gap by embedding quantitative information within the tree. The R package Metacoder (Foster, et al., 2017) introduced this concept for microbial community analysis, but its reliance on the R environment, static ggplot2 (Wickham, 2011) output and limited label control make it difficult to use for large or complex datasets. Label placement is particularly challenging, with labels frequently overlapping or becoming unreadable at deeper levels, and assembling multi-panel, multi-group comparisons typically requires substantial manual editing of static figures (Wilke, 2019).

To address these limitations, we developed MetaTree, a fully browser-based platform for interactive heat-tree visualization of hierarchical quantitative data. MetaTree requires no

installation or coding and provides an integrated workflow for uploading data, exploring hierarchical patterns and exporting publication-quality figures. Implemented in JavaScript with D3.js (Bostock, et al., 2011), it constructs a shared reference tree from all hierarchical items so that all samples and groups are visualized on the same topology. Users can adjust color mappings, value domains, label density, node sizes and branch widths, and can switch between several tree layouts to suit different datasets. Beyond visualization, MetaTree supports differential-abundance summaries, operational taxon-function (OTF) (Wu, et al., 2025) structures and dedicated multi-group comparison modes, enabling clade-level patterns to be examined across samples or experimental conditions within a single, interactive framework.

2 Implementation

2.1 Visualization framework

MetaTree first constructs a global reference tree by combining all hierarchical items present in the input data. This ensures that all tree panels share the same topological structure, so differences between samples or groups appear only as changes in color, size and edge appearance. In this way, panels can be compared directly across samples and groups, which is difficult with conventional tree-based visualizations. MetaTree also supports synchronized zooming across panels, making it easier to explore differences between trees in comparative settings.

Quantitative values are mapped onto three main visual channels:

- **Tree color (nodes and branches)** encodes a user-selected variable such as abundance, intensity or log2 fold-change. MetaTree provides multiple built-in color palettes and allows users to adjust the value domain and the direction of the color scale. Transformations such as logarithmic or square-root scaling can be applied before mapping, making it possible to handle both highly skewed abundance distributions and centered contrast statistics.
- **Node size** reflects the magnitude of a quantity of interest, typically an aggregated abundance or intensity. Global size scaling controls allow users to emphasize or down-weight differences in node area for dense or sparse trees.
- **Branch width** can be used to enhance structural readability or to emphasize clades with higher overall signal, depending on the analysis goal.

MetaTree offers extensive controls for adjusting the visual encoding. Users can select color palettes, define value ranges or quantile-based domains and apply optional

transformations to the mapped variable. Label presentation is highly configurable: label density and hierarchical depth can be adjusted to control which levels of the tree are shown, individual clades can be highlighted by assigning custom label colors and font size can be tuned to balance readability with visual compactness. Node and branch scaling parameters allow users to emphasize or de-emphasize structural features, while the tree layout can be switched among radial, tidy and circular styles. To improve readability in dense trees, MetaTree includes an optional auto-hide mechanism for overlapping labels, which retains only a non-overlapping subset while preserving the underlying layout; when this option is disabled, all labels are drawn. All visualization settings are applied interactively, and the resulting panels can be arranged and exported as high-resolution PNG or SVG files, supporting both exploratory analysis and publication-quality figures.

2.2 Data input and hierarchical model

The primary input to MetaTree is a sample-level hierarchical table. Two general formats are supported:

- **Wide format**, where each row corresponds to a hierarchical item (e.g. a taxon or an OTF unit) and columns store per-sample quantitative values such as abundance. This includes conventional taxon abundance tables produced by 16S, metagenomics and metaproteomics pipelines.
- **Long format**, where each row encodes an item–sample or item–contrast pair with associated statistics, for example log2 fold-change, effect size and P-value obtained from statistical tools.

Users may optionally provide a metadata table describing sample-level attributes (e.g. treatment, age group), or define groups directly within the interface. MetaTree links measurements to the user-provided hierarchy and, where needed, aggregates sample values within groups before plotting.

Although MetaTree was motivated by microbiome and metaproteomics applications, the framework is not restricted to these domains and can visualize any dataset that can be expressed as features attached to a rooted hierarchy.

2.3 Analysis modes

On top of the shared visualization framework, MetaTree implements four analysis modes:

- **Individual samples:** multiple samples are shown simultaneously as separate heat-tree panels on the shared reference tree, with sample-specific values mapped directly to

tree color and node size. This supports side-by-side inspection of sample-level heterogeneity and outlier detection.

- **Group samples:** samples are aggregated into user-defined groups based on metadata or interactive grouping, and group-level summaries (e.g. mean or median abundance) are visualized as one heat tree per group. This mode facilitates comparison of overall hierarchical profiles across conditions.
- **Group comparison:** two groups are selected for pairwise comparison. For each hierarchical item, MetaTree applies one of two built-in tests (Wilcoxon rank-sum (Mann and Whitney, 1947) or Welch's t-test (Welch, 1947)) and uses Benjamini–Hochberg (Benjamini and Hochberg, 1995) adjustment to control the false discovery rate. The resulting statistics, such as log2 fold-change and adjusted P-values, are then used primarily as filtering criteria: users can restrict the display to items exceeding a chosen effect-size threshold and passing an FDR cutoff, thereby focusing the tree on clades with consistent between-group differences.
- **Comparison matrix:** when more than two groups are present, MetaTree can automatically compute pairwise comparisons between all group pairs and arrange the results in a matrix of heat-tree panels. Any cell can be selected for detailed inspection using the same visualization and filtering controls as in group comparison mode, providing an overview of relationships between conditions while preserving access to individual contrasts.

Across all modes, additional filters allow users to subset the tree by metadata (e.g. selected groups), by item identity (e.g. specific clades or keyword matches) and by simple thresholds on abundance or other numeric variables. P-values and FDR can be used as filtering criteria to control which items are shown, thereby reducing visual clutter and highlighting the most relevant features.

3 Results

To demonstrate MetaTree's capabilities, we applied it to a previously published human gut microbiome metaproteomics dataset generated from stool-derived communities of three donors cultured in vitro with several structurally distinct oligosaccharides (Zhang, et al., 2025). Taxonomic and operational taxon–function (OTF) tables were generated using MetaX, a peptide-centric metaproteomics workflow (Wu, et al., 2025), and used as input to MetaTree. Representative visualizations from this dataset are shown in Figure 1.

Using the individual samples and group samples modes, MetaTree visualizes how clade-level taxonomic and taxa–function patterns vary across donors and treatments on a shared hierarchical topology. In group samples mode (Figure 1A), group-level heat trees for the control (PBS) and five oligosaccharide treatments are displayed side by side on the same reference tree, facilitating visual comparison of overall hierarchical profiles across conditions. In individual samples mode (Figure 1B), tidy-tree layouts for selected samples highlight clade-specific differences while preserving the underlying hierarchy.

In the group comparison mode, MetaTree combines built-in hypothesis testing with FDR-based filtering to highlight taxa and OTFs that differ significantly between a selected oligosaccharide and the control. The comparison matrix mode (Figure 1C) extends this analysis to all pairs of treatments, producing a grid of heat-tree panels in which node and branch color encode log2 fold change for each pairwise contrast. Each small panel in the matrix is fully interactive and can be opened in a larger, dedicated view, where users can further adjust visualization parameters or export publication-quality figures. The matrix thus provides both a compact overview of relationships between treatments and an entry point for detailed exploration of individual contrasts.

Together, these examples illustrate how MetaTree integrates interactive visualization, quantitative comparison and hierarchical interpretation within a single platform, and how it can be applied to real microbiome metaproteomics data to explore treatment-specific and donor-specific responses.

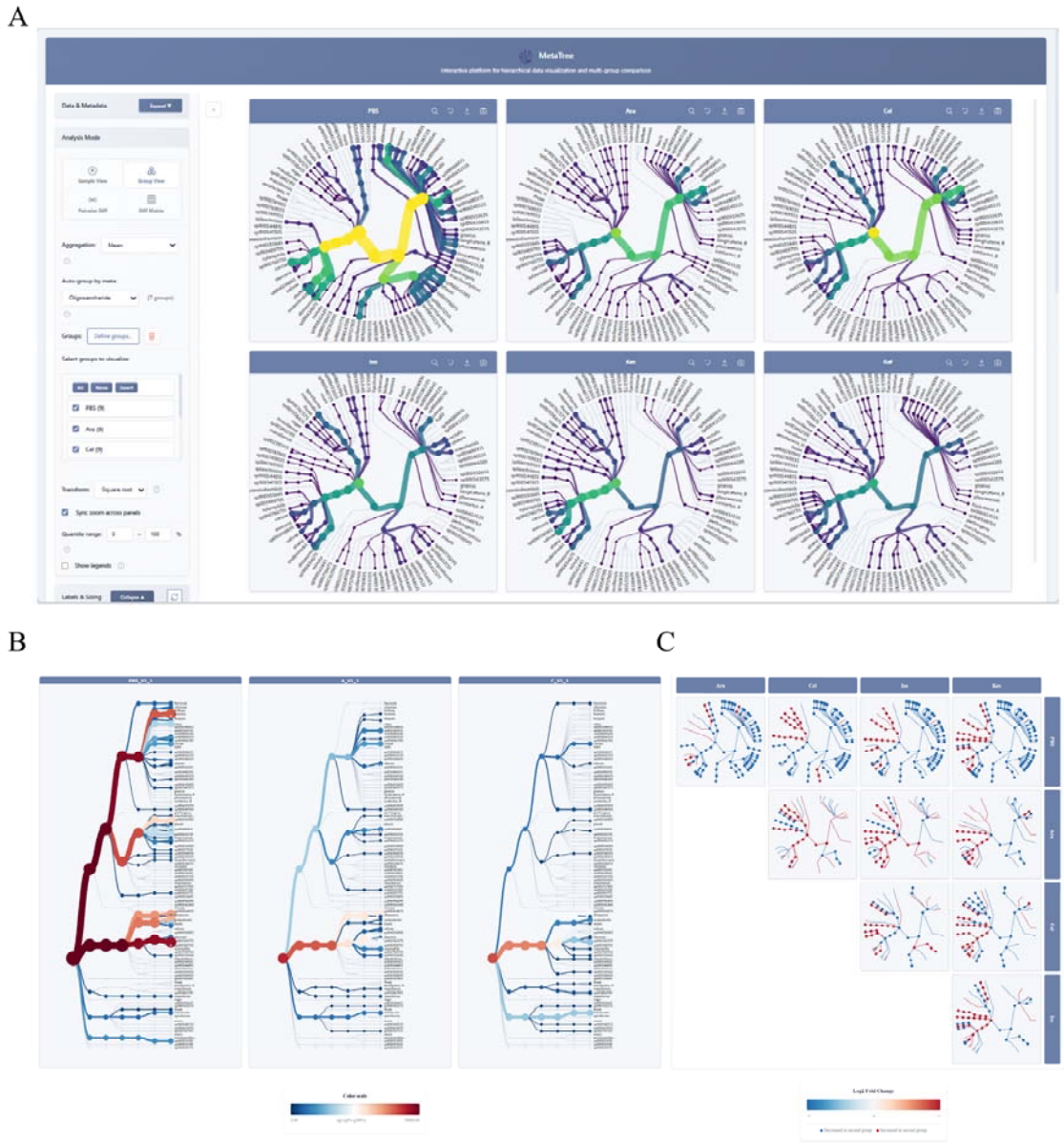


Figure 1. MetaTree interface and main analysis modes for hierarchical microbiome data.

(A) MetaTree web interface in group samples mode, showing group-level taxonomic heat trees for the control (PBS) and five oligosaccharide treatments. All panels share a common reference tree, and tree color and node size encode aggregated taxon abundance, enabling direct comparison of clade-level profiles across groups.

(B) Example of individual samples mode using a tidy-tree layout. Three samples from different treatments are displayed on the same taxonomy, with tree color representing transformed taxon abundance. This layout highlights clade-specific differences along the vertical hierarchy while preserving the shared topology.

(C) Comparison matrix mode for taxonomic data, summarizing all pairwise contrasts between treatments as a grid of heat trees. Each cell corresponds to one group comparison on the shared reference tree, with node and branch color encoding log2 fold change between the two groups (blue, decreased; red, increased). This provides a compact overview of treatment effects and allows targeted inspection of individual contrasts.

Acknowledgements

The authors acknowledge the use of ChatGPT (OpenAI) for assistance with English language editing and polishing during manuscript preparation.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through grant RGPIN-03905-2018 to D.F.

Availability

MetaTree is freely accessible at https://byemaxx.github.io/MetaTree

The source code, documentation and example datasets are available on GitHub at https://github.com/byemaxx/MetaTree. MetaTree supports current versions of major web browsers and can be used either as an online service or deployed locally for offline analysis.

References

Ashburner, M., *et al.* Gene ontology: tool for the unification of biology. *Nature genetics* 2000;25(1):25-29.

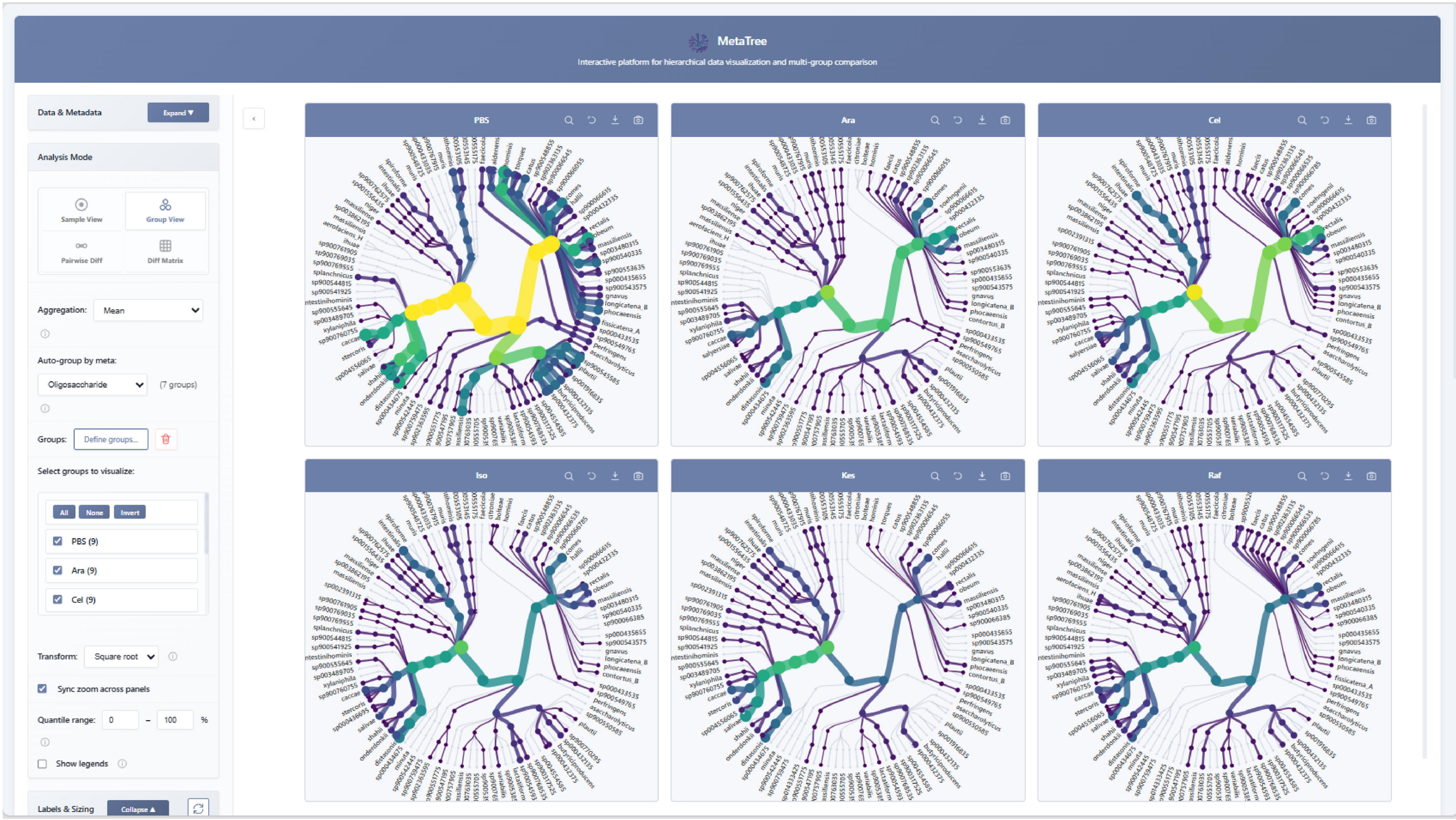
Asnicar, F., *et al.* Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* 2015;3:e1029.

Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 1995;57(1):289-300.

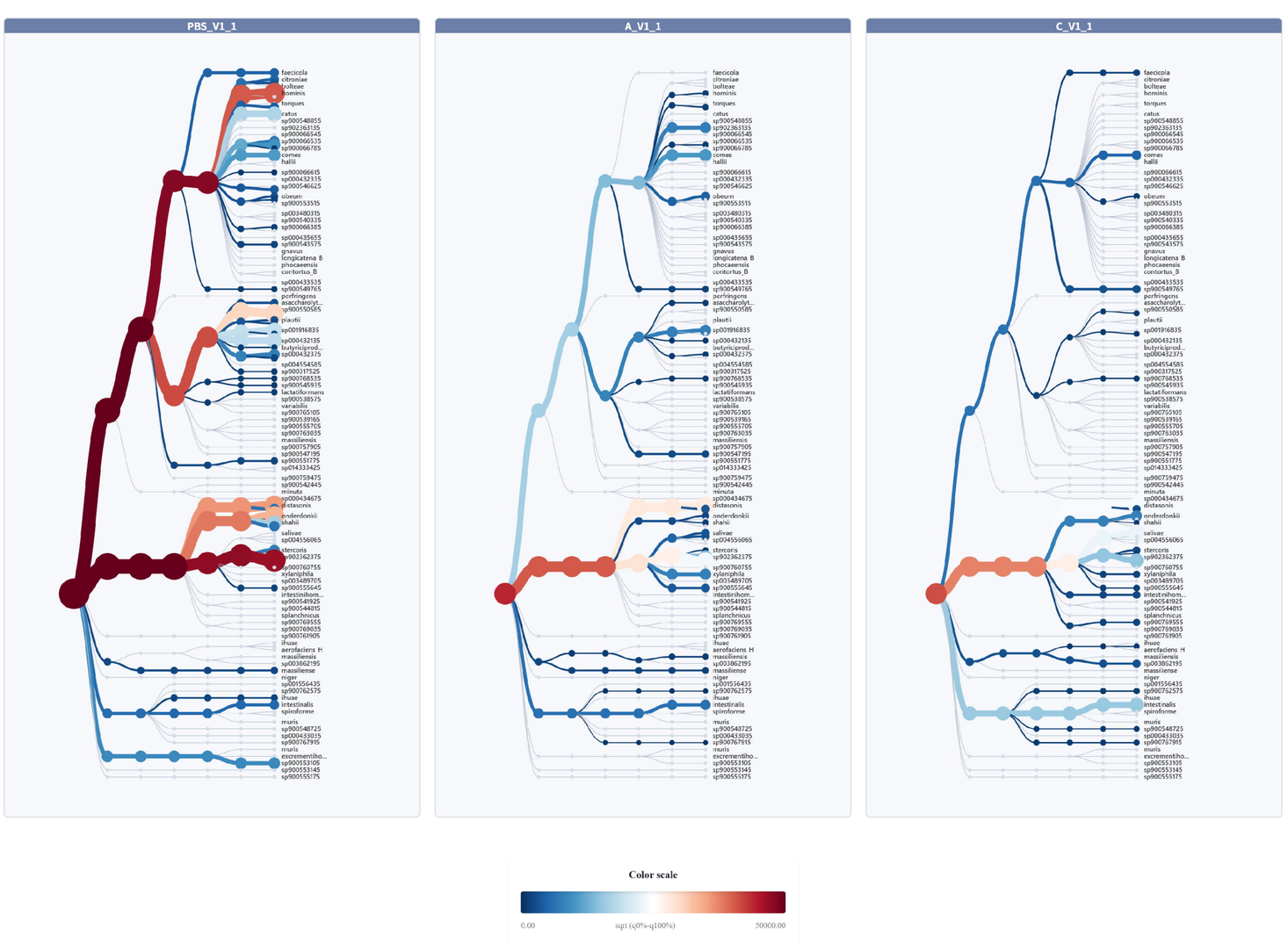
Bostock, M., Ogievetsky, V. and Heer, J. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics* 2011;17(12):2301-2309.

Foster, Z.S., Sharpton, T.J. and Grünwald, N.J. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS computational biology* 2017;13(2):e1005404.
Huerta-Cepas, J., Serra, F. and Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular biology and evolution* 2016;33(6):1635-1638.
Kanehisa, M., *et al.* KEGG: biological systems database as a model of the real world. *Nucleic acids research* 2025;53(D1):D672-D677.
Kreft, Ł., *et al.* PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics* 2017;33(18):2946-2947.
Letunic, I. and Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic acids research* 2024;52(W1):W78-W82.
Mann, H.B. and Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* 1947:50-60.
McMurdie, P.J. and Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* 2013;8(4):e61217.
Moon, K.R., *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology* 2019;37(12):1482-1492.
Ondov, B.D., Bergman, N.H. and Phillippy, A.M. Interactive metagenomic visualization in a Web browser. *BMC bioinformatics* 2011;12(1):385.
Parks, D.H., *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology* 2018;36(10):996-1004.
Quince, C., *et al.* Shotgun metagenomics, from sampling to analysis. *Nature biotechnology* 2017;35(9):833-844.
Ward, D.M., Weller, R. and Bateson, M.M. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 1990;345(6270):63-65.
Welch, B.L. The generalization of ‘STUDENT'S’problem when several different population varlances are involved. *Biometrika* 1947;34(1-2):28-35.
Wickham, H. ggplot2. *Wiley interdisciplinary reviews: computational statistics* 2011;3(2):180-185.
Wilke, C.O. Fundamentals of data visualization: a primer on making informative and compelling figures. O'Reilly Media; 2019.
Wilmes, P. and Bond, P.L. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends in microbiology* 2006;14(2):92-97.
Wu, Q., *et al.* Operational taxon-function framework in MetaX: Unveiling taxonomic and functional associations in metaproteomics. *Analytical Chemistry* 2025;97(18):9739-9747.
Zhang, A., *et al.* Microbiome-Dependent Functional Responses to Structurally Distinct Oligosaccharides Revealed by Metaproteomics. *bioRxiv* 2025:2025.2006. 2026.661555.

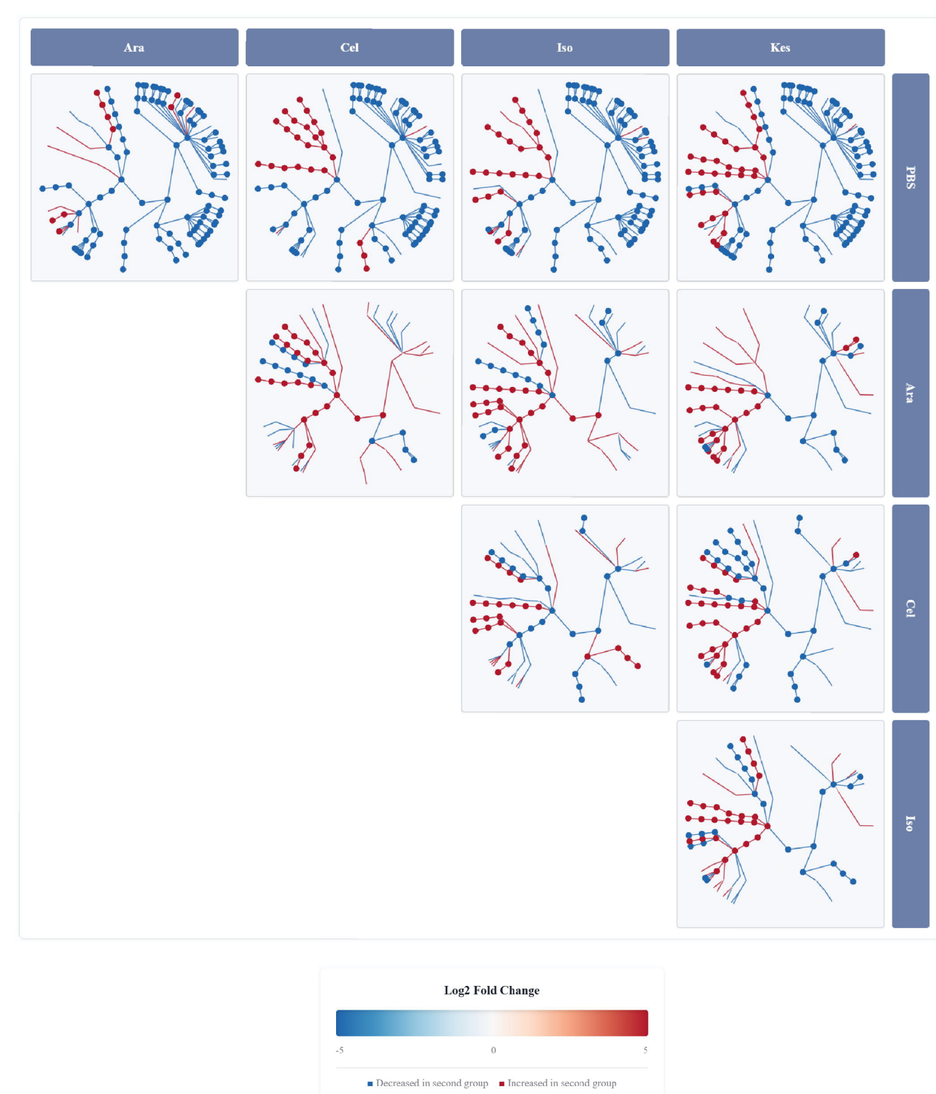
A


MetaTree
Interactive platform for hierarchical data visualization and multi-group comparison
Data & Metadata
Expand
Analysis Mode
Sample View
Group View
Pairwise Diff
Diff Matrix
Aggregation:
Mean
Auto-group by meta:
Oligosaccharide
(7 groups)
Groups:
Define groups...
Select groups to visualize:
All
None
Invert
PBS (9)
Ara (9)
Cel (9)
Transform:
Square root
Sync zoom across panels
Quantile range:
0
100
%
Show legends
Labels & Sizing
Collapse
PBS
Ara
Cel
Iso
Kes
Raf

B


PBS_V1_1
A_V1_1
C_V1_1
Color scale
0.00
sqrt (q0%-q100%)
50000.00

C


Ara
Cel
Iso
Kes
PBS
Ara
Cel
Iso
Log2 Fold Change
-5
0
5
Decreased in second group
Increased in second group