Visualization of Microbial Community Dynamics

Kris Sankaran
Stanford University

Problem

Modern microbiome studies often revolve around the dynamics of microbial communities, attempting to characterize their response to environmental changes. For example, after an antibiotic shock, does the initial community rebound?

Further, it is useful to compare these dynamics across samples with different covariates. It would be useful to know whether certain patient populations have more or less stable microbiomes.

Motivation

As it is helpful to think of microbes according to their taxonomic organization, this problem motivated the study of techniques for the display and comparison of hierarchical time series. There is literature on several components of this problem, especially tree and time series visualization, but less on tree comparison, and almost none combining them [1, 3, 4, 5]. Further, there are limited implementations for practitioners.

Approaches

We have performed preliminary experiments adapting sparklines, Degree-of-Interest (DOI) trees, Timeboxes, and Sankey diagrams, each designed to facilitate a different kind of visual analysis. We describe these approaches below.

- Sparklines: We can view a hierarchical time series by placing each time series at tree tips, as in [5]. See Figure 1.
- DOI-trees: For displaying a single sample at one timepoint, we can apply the DOI principle to adjust tree layout based on user input [3]. We encode abundance by node and edge width. See Figures 2 and 3.
- Timeboxes: We can build timeboxes over the microbial time series, and incorporate generalized linking to identify associated taxonomic groups [1, 4].
- Treeboxes: We reverse the Timebox idea, highlighting series associated with the union of selected tree nodes. See Figure 5.
- DOI Sankey: For multiple samples at a single timepoint, we display trees with color representing sample. Comparing the proportion of different colors along edges gives the relative abundances across samples. See Figure 6.

Examples

For these experiments, we use data from [2], describing the microbiome during pregnancy. This study collected between 8 and 48 vaginal swabs up to a year before delivery. In sparkline, DOI-Tree, TimeBox, and TreeBox experiments, we focus on a single sample with the most samples, taking average abundance for node sizes. We also ignore non-vaginal measurements, though they are available.

For the Sankey Comparisons, we display the average value across samples grouped in “Community State Types” computed in the statistical analysis published in [2]. The fifth state type was associated with Preterm birth.

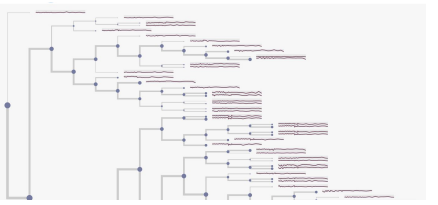

Figure 1. Sparklines make it easy to compare time series in adjacent taxonomic groups.

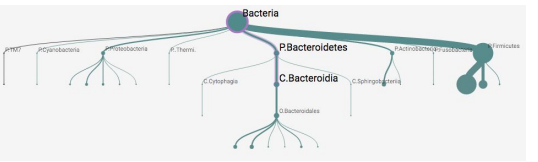

Figure 2. The DOI representation facilitates navigation of tree abundances.

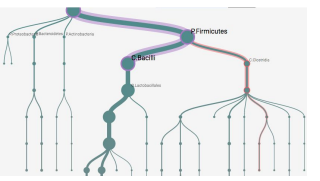

Figure 3. The DOI tree allows search -- here a path to the *Lachnospiraceae* is highlighted in red -- and different degrees of filtering.

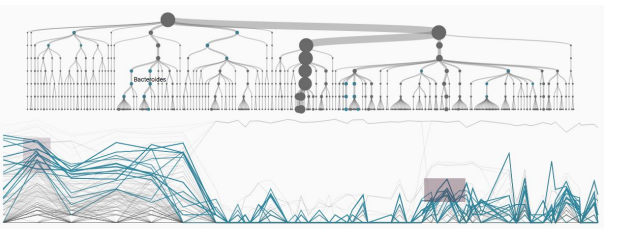

Figure 4. Timeboxes and generalized linking allow identification of taxonomic subgroups with specific time series shapes.

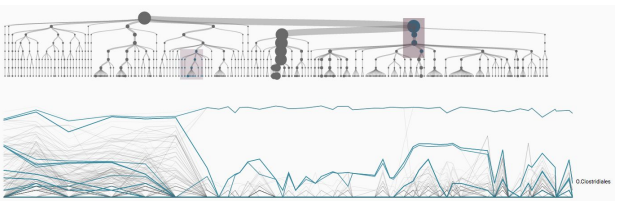

Figure 5. We define Treeboxes as an analogue to Timeboxes, where elements on the tree highlighted associated time series.

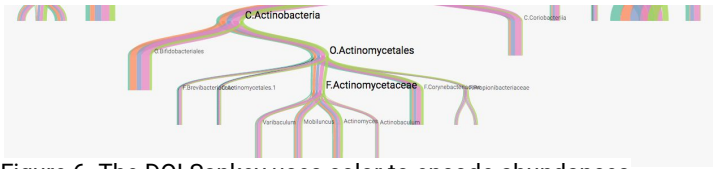

Figure 6. The DOI Sankey uses color to encode abundances across different samples.

References

[1] Richard A Becker and William S Cleveland. Brushing scatterplots. Technometrics, 29 (2):127–142, 1987.
[2] DiGiulio, Daniel B., et al. Temporal and spatial variation of the human microbiota during pregnancy. Proceedings of the National Academy of Sciences 112.35 (2015): 11060-11065.
[3] Jeffrey Heer and Stuart Card. Doitrees revisited: scalable, space-constrained visualization of hierarchical data. Proceedings of the Conference on Advanced Visual Interfaces, 421–424. ACM, 2004.
[4] Harry Hochheiser and Ben Shneiderman. Interactive exploration of time series data. In Discovery Science, 441–446. Springer, 2001.
[5] Edward R Tufte. The visual display of quantitative information. Graphics press Cheshire, CT, 1983.