Microbial Community Dynamics

Kris Sankaran

May 23, 2016

Problem background

- What is the problem domain?
  - Microbiomes are communities of microbes living in a shared environment (human gut, lake water, ...).
  - They are interesting for basic science as well as their clinical and industrial applications.
- Researchers care mostly about ecological and medical questions.
  - Ecological: What is the community structure of microbes, and how does it respond to environmental perturbations?
  - Medical: Are any subcommunities related to disease? Which change after treatment?

Information visualization tasks

- **Navigation** of hierarchical time series.
  - Each microbe contributes a time series of abudnances, and they can be arranged along an evolutionary tree.
  - For real-world use, need to view ≈ 500 microbes for ≈ 25 patients across ≈ 30 time points.
- **Comparisons** across sets of trees.
  - Which (sets of) microbes have different abundances in treatment and control?

Literature review

- Microbiome science has a few popular workflows.
  - QIIME, phyloseq, and ANVI'O [1, 2, 3].
- Visualization community has tackled many subtasks.
  - Degree-of-interest trees [4]: Navigation of large trees.
  - TreeVersity [5]: Comparison between pairs of trees.
  - Some methods for hierarchical time series visualization [6, 7].
  - Timeboxes and generalized selection[8, 9]: Exploring many time series, and relating to contextual information.
- What's my contribution?
  - I want to put myself out of work.

Linking time + tree

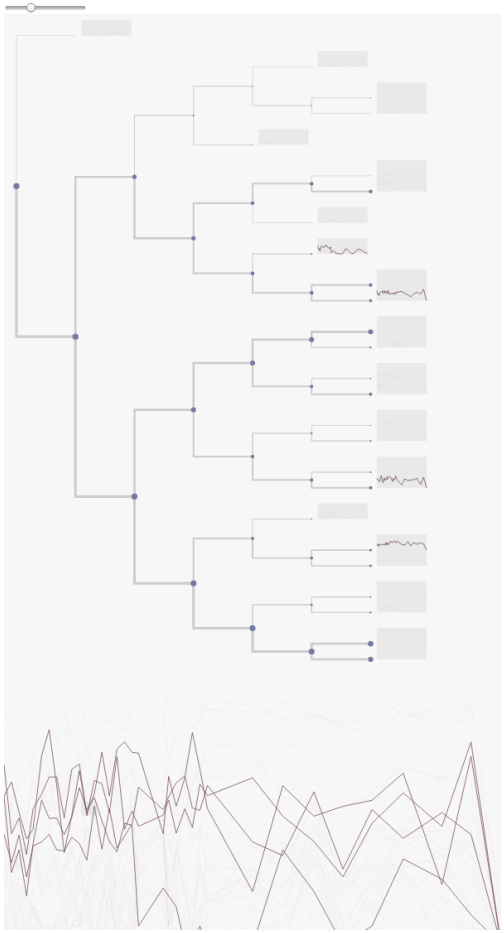

Figure: I've tried incorporating timeboxes and sparklines into visualization for a single sample.

Navigating trees

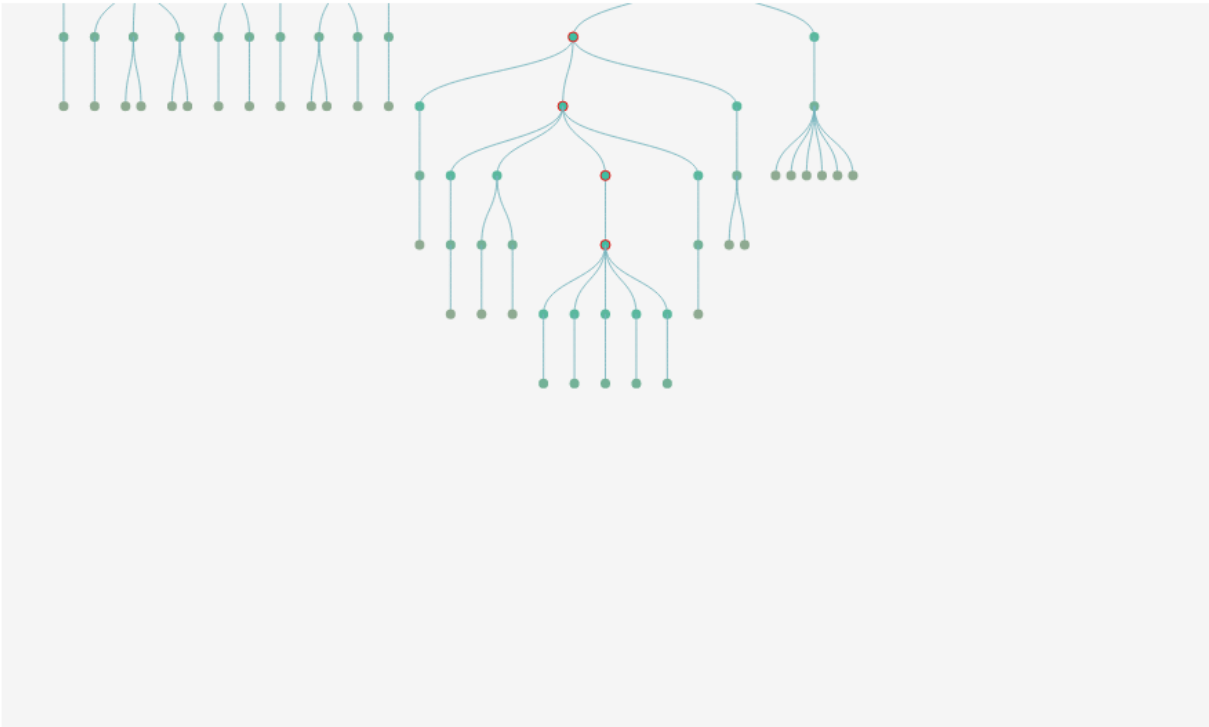

Figure: I've written a basic version of DOI trees, which doesn't include microbial abundance information.

Comparing trees

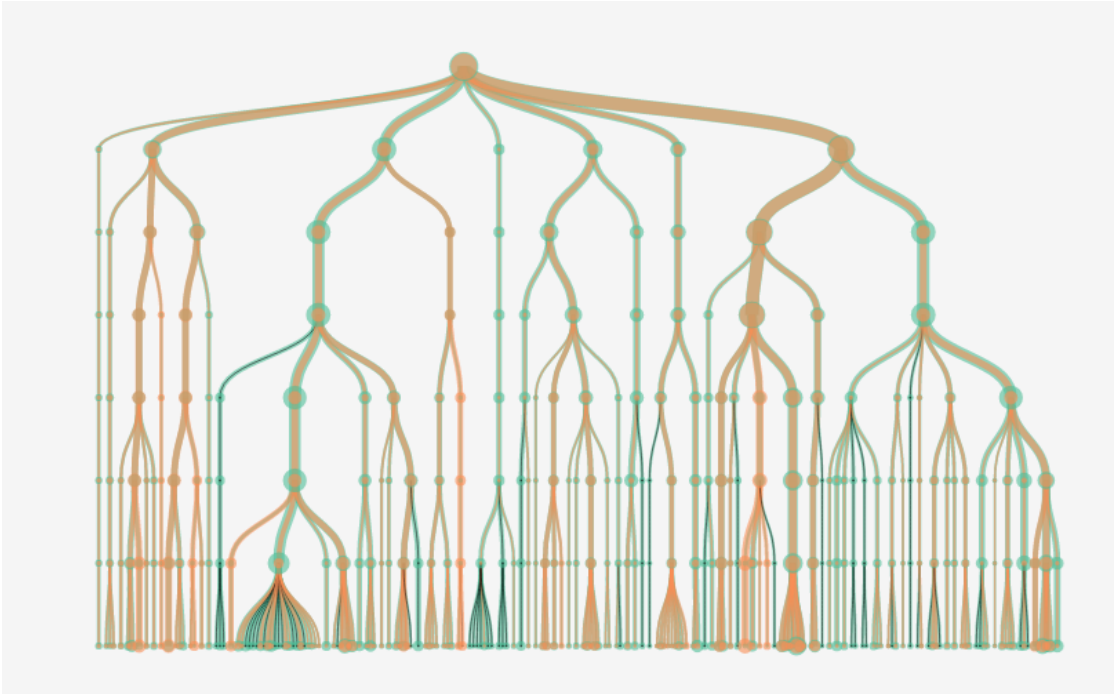

Figure: I've run preliminary experiments with TreeVersity-style comparison.

Feedback

- The next step is to synthesize different experiments into one visualization, how would you approach this?
- How would you compare more than 2-3 samples in a single tree?
- I actually collaborate with microbiome scientists, and it might be interesting to design a study – what would be the important points to keep in mind?
- Any ideas for names? (I don't really like the ones I came up with: phyloSparklines, or treelapse, or ChronoTree, ...)

J Gregory Caporaso, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D Bushman, Elizabeth K Costello, Noah Fierer, Antonio Gonzalez Pena, Julia K Goodrich, Jeffrey I Gordon, et al.

Qiime allows analysis of high-throughput community sequencing data.

Nature methods, 7(5):335–336, 2010.

Paul J McMurdie and Susan Holmes.

phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data.

PloS one, 8(4):e61217, 2013.

A Murat Eren, Özcan C Esen, Christopher Quince, Joseph H Vineis, Hilary G Morrison, Mitchell L Sogin, and Tom O Delmont.

Anvio: an advanced analysis and visualization platform for omics data.

PeerJ, 3:e1319, 2015.


Jeffrey Heer and Stuart K Card.
Doitrees revisited: scalable, space-constrained visualization of hierarchical data.
In *Proceedings of the working conference on Advanced visual interfaces*, pages 421–424. ACM, 2004.


John Alexis Guerra-Gómez, Audra Buck-Coleman, Catherine Plaisant, and Ben Shneiderman.
Treeversity: Interactive visualizations for comparing two trees with structure and node value changes.
In *Proc. Conference of the Design Research Society-DRS2012*, volume 1, page 10. Citeseer, 2012.


Spyros Makridakis, Steven C Wheelwright, and Rob J Hyndman.
Forecasting methods and applications.
John Wiley & Sons, 2008.


Yumiko Uchida and Takayuki Itoh.

A visualization and level-of-detail control technique for large scale time series data.
In *Information Visualisation, 2009 13th International Conference*, pages 80–85. IEEE, 2009.

Harry Hochheiser and Ben Shneiderman.
Interactive exploration of time series data.
In *Discovery Science*, pages 441–446. Springer, 2001.

Jeffrey Heer, Maneesh Agrawala, and Wesley Willett.
Generalized selection via interactive query relaxation.
In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 959–968. ACM, 2008.