Patient Microbiome

Visual Analytics for Exploring and Reporting Patient Microbiome

Colleen Bushell[1], Nick Chia[2], Lisa Gatzke[1], Peter Groves[1], Matthew Berry[1], Xiaoxia Liao[1], Stephen Johnson[2], Loretta Auvil[1], Michael Welge[1], Christian Followell[1], Nate Russell[1], Bryan White[1], Heidi Nelson[2] *1: University of Illinois, 2: Mayo Clinic*

Abstract

Microbial community compositions vary between individuals yet show commonalities among disease phenotypes.[1] Differences in the level of microbial diversity are apparent when comparing disease and non-disease states.[2] Understanding these microbial community profiles, and having the ability to compare an individual patient's profile to others with similar profiles, can lead to better understanding of a patient's condition. The Visual Analytics group at the University of Illinois, with input from Mayo's Microbiome Program researchers and clinicians, developed a prototype visual analytics pipeline and reporting tool that supports research and demonstrates feasibility for future use in lab analysis and clinical reporting. This prototype achieves the following:

A. Combines and transforms patient microbial and phenotypic data from multiple studies to establish quality cohorts and comparisons.

B. Provides query tools for building cohorts based on phenotype information.

C. Executes comparison analysis between cohorts and a single patient and visualizes the differences between their relative abundance, alpha diversity (evenness and richness) and beta diversities at multiple levels of detail can be deduced from the visualizations.

D. Important OTUs that differentiate cohorts are determined using the Random Forest algorithmic approach that handles small sample sizes.[3]

E. Multiple techniques for presenting these results are provided, including mapping information onto an interactive phylogenetic tree to identify specific branches of relevance.

F. Provides methods for indicating pathogenic microbes, as well as actionable single-species and complex biomarkers.

This system acts as both a tool for research, providing powerful analytics to study the underlying mechanisms of disease, as well as a visual tool for clinicians and patients to appreciate and comprehend the clues that their microbiome provides in understanding their health. This prototype helps doctors envision how future discoveries can be easily reported for clinical use. Sequences from multiple studies were used as an example and reprocessed together using IM-TORNADO[4]. In this poster we illustrate our design approach[5,6] and describe our solutions.

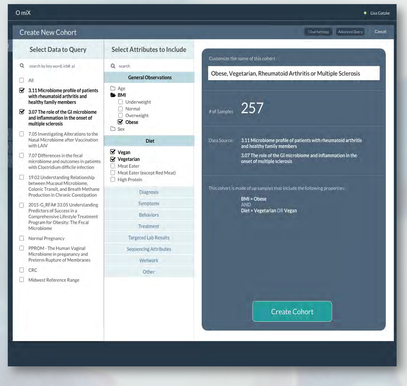
[1] Manimozhiyan, A. et. al. "Enterotypes of the Human Gut Microbiome", Nature vol 473, 174–180, May 2011.
[2] Kostic, A, et. al. "The Dynamics of the Human Infant Gut Microbiome in Development and in Progression Toward Type 1 Diabetes", Cell Host & Microbe, Feb 2015.
[3] Welge, M., C. Bushell, N. Russell, M. Berry, B. White, et. al. "Feature Interaction Detection with Random Forest in a High Dimensional Setting," Poster Presentation, Individualizing Medicine Conference, Mayo Clinic, MN, Oct 2014.
[4] Jeraldo, P., et. al. "IM-TORNADO: A Tool for Comparison of 16S Reads from Paired-End Libraries", PLoS One, 9(12): e114804, December, 2014.
[5] Bushell, C. "Design Approaches for the Display of Genetic Test Results", Presentation, Individualizing Medicine Conference, Mayo Clinic, MN, 2012.
[6] Schahl, K., C. Bushell, M. Ferber. "Data Visualization in Reporting Next Generation Sequencing Results." Poster Session, American College of Medical Genetics and Genomics *ACMG Annual Clinical Genetics Conference*, Phoenix, AZ (March 19-23, 2013).

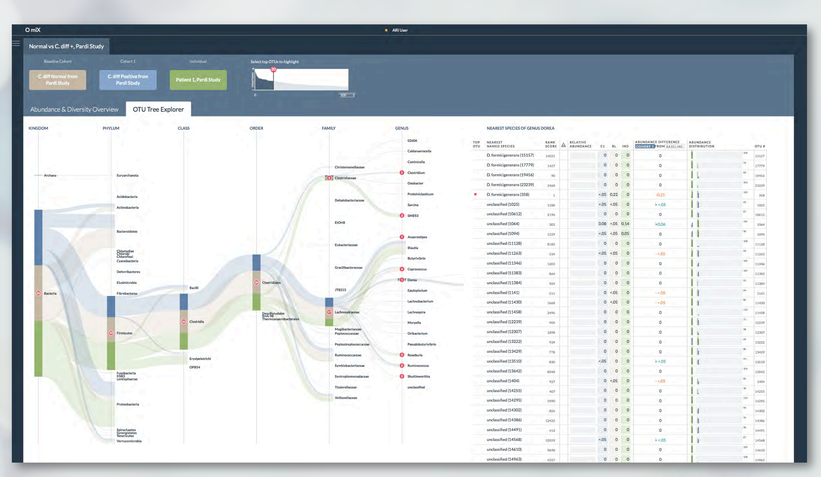
Acknowledgements

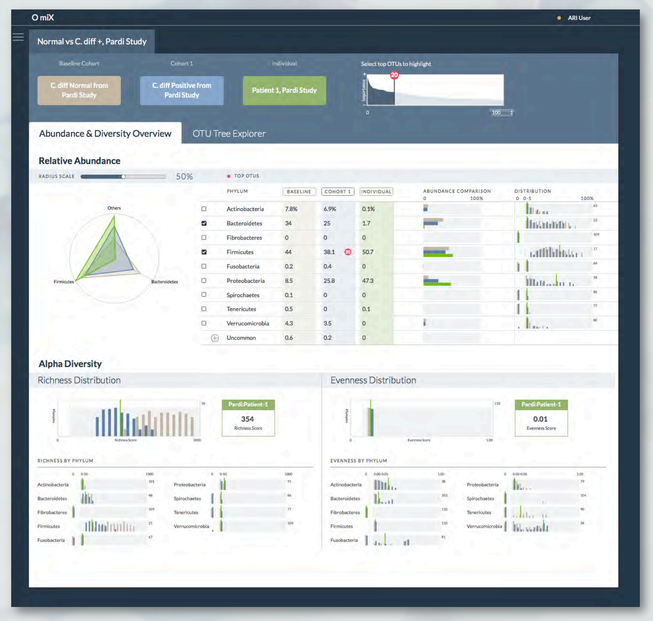
Funding provided by the University of Illinois ARI Visual Analytics group, the Mayo Clinic CIM Microbiome Group and the Mayo Illinois Alliance for Technology Based Healthcare

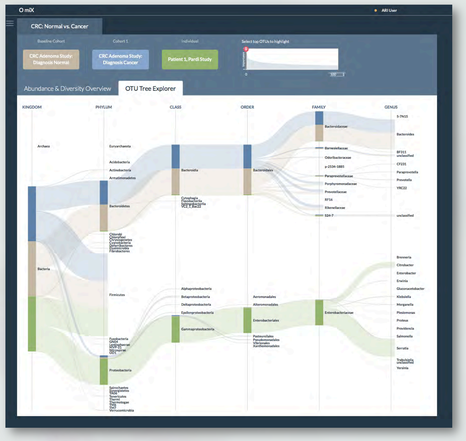
Cohorts and Important OTUs

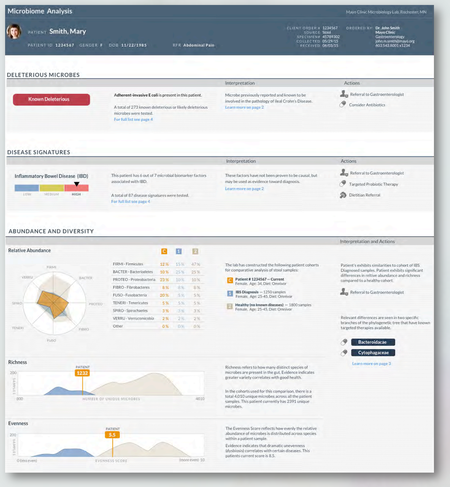
The researcher, or the lab analyst, creates cohorts using phenotype information. These cohorts are characterized by calculating relative abundance and diversity of the microbial community at all levels of the phylogenetic tree. Comparative analysis is executed using the computationally intensive Random Forest algorithmic approach to identify discriminating microbes (microbial OTUs) between two cohorts - typically a "healthy" cohort to a phenotype of interest. These important OTUs are indicated with red dots throughout the visualizations so researchers can study potential mechanisms that describe a phenotype. A single patient can be compared to two cohorts helping the clinician and patient understand their microbiome within the context of others. In these examples, the tan cohort consists of healthy patients, the blue cohort consists of patients with C. diff infections, and the green represents a single C. diff patient.

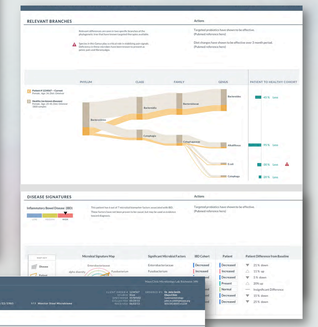


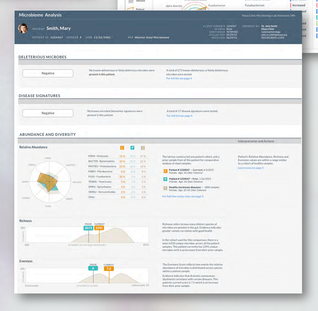












Relative Abundance

Relative abundance is the percent composition of an organism of a particular kind *relative* to the total number of organisms in the patient's sample. These are calculated at all levels of the phylogenetic tree so that the microbial community can be characterized at various levels of detail. When describing the relative abundance of an entire cohort, an average for the group is used. This single value can be misleading if people within the cohort vary significantly. To help elevate misunderstanding, we show the distribution of the entire group, and place a vertical green line indicating where the specific patient falls. An interactive radar graph is used to compare phyla of interest, visually overlapping both cohorts and the single patient.

Diversity

Richness is a function of how many unique species of bacteria are in a sample. The metric used is the number of unique Organizational Taxonomic Units (OTUs) in a sample. OTUs are roughly defined as groups of r16s genetic sequences that are 97% similar. OTUs that are 97% similar to a known species of bacteria are considered to be of that species, with lower thresholds of similarity needed to classify Genus, Family, etc. when a Species match is not found. In the case of Phyla, the Richness is the number of unique OTUs that have been classified as being part of that Phylum (and the OTU has a non-zero abundance in the particular sample).

Evenness indicates if the species that are present have similar levels of abundance in a sample, or if a small number of species dominate the population. A low Evenness score presumably means that some kind of selective pressure exists that benefits a small number of species. A high Evenness score likely indicates that either 1) the distribution of species is fairly random, with little selective pressures to prefer one species over another, or 2) some kind of feedback system exists that keeps the population balanced. The implication is that a low Evenness score is evidence of strong selective pressures benefiting a small number of species, but a high Evenness score provides little evidence to indicate if there are strong selective pressures or randomness at work.

Phylogenetic Tree Explorer

It is valuable for researchers to understand relative abundance and important OTUs in the context of the phylogenetic tree. This makes it easier to identify specific branches that show microbial imbalance, or include discriminating OTUs. In this view, branches can be clicked on and off to display levels within the tree. Each color band represents the relative abundance for each cohort (tan: healthy, blue: C. diff patients, green: single C. diff patient) for easy comparison. A table is provided at the species level. For example, in the tree view to the left, braches are expanded that contain significant quantities of top OTU's (indicated with red dots). In the other tree view, branches illustrating dramatic differences between the single patient and the cohorts are expanded.

Reporting

These images show hypothetical reports illustrating concepts for how microbiome data might be used in a clinical setting in the future. In addition to characterizing the patient's microbiome in the context of cohorts, attempts are made to show how bacteria of known clinical relevance might be included, as well as biomarker signatures that explain a phenotype of clinical relevance. Our design approach focuses on providing the information at a summary level on the first page. Here, a short description, brief lab interpretation and possible actions are provided. Then more detail, evidence and citations are provided on subsequent pages.