Representing Diet in a Tree-Based Format for Interactive and Exploratory Assessment of Dietary Intake Data

Nicole Litwin,[1] Kalen Cantrell,[1] Aurelie Cotillard,[2] Muriel Derrien,[2] Abigail Johnson,[3] Rob Knight,[1] Franck Lejzerowicz,[1] Daniel McDonald,[4] Brent Nowinski,[1] Se Jin Song,[1] Julien Tap,[2] and Patrick Veiga[2]

[1]Center for Microbiome Innovation, University of California, San Diego; [2]Danone Nutricia Research; [3]University of Minnesota; and [4]University of California San Diego

Objectives: We assessed the utility of representing dietary intake data in hierarchical tree structures that consider relationships among foods.

Methods: Dietary intake was collected from 1909 adults ($\geq$18 years) using a food frequency questionnaire (FFQ; VioScreen) from the American Gut Project. FFQ food items were formatted into hierarchical tree structures based on 1) USDA's Food Nutrient and Database for Dietary Studies (FNDDS) classifications, 2) nutrient content, and 3) molecular compound information detected via mass spectrometry to capture the non-nutrient composition of foods. Next, we compared how well representing dissimilarities (or distances) between individuals based on their diet corresponded with indices such as the Healthy Eating Index (HEI-2015), when those distances are calculated using tree-based versus non-tree-based metrics. We performed an Adonis test (PERMANOVA) to measure the amount of variation explained (R^2) in these diet-based distances by HEI-2015.

Results: We observed that dietary ordinations generated using tree-based relationships between foods have better agreement with HEI than ordinations generated without considering relatedness between foods. The variation explained by HEI-2015 increased by 35% when using the FNDDS tree compared to using a non-tree based quantitative metric (Bray-Curtis (not tree-based) $R^2 = 0.02931$ vs. Weighted UniFrac (tree-based) $R^2 = 0.03969$), by >20% when using the nutrient tree (vs. Weighted UniFrac $R^2 = 0.03627$), and only marginally (6%) when using the molecular compound tree (vs. Weighted UniFrac $R^2 = 0.03116$).

Conclusions: We show that tree-based measurements of dietary similarity lead to better agreement with diet indices (e.g., HEI) than when relationships among foods are not considered. We also show that representing dietary intake in a tree-like structure can offer interactive visualizations of data that can be used to inform hypotheses regarding dietary characteristics.