

The gut microbiome is a community of bacteria, viruses, fungi and protozoa which produce metabolites and interact with the host to regulate gastrointestinal homeostasis, immune function, and systemic metabolism. The composition of this environment has been implicated in many aspects of nutrition and pathophysiology and is influenced by numerous factors, including diet, exercise, medications, sleep, and stress. This randomized controlled trial aims to isolate the effects of dietary protein sources on the gut microbiome by analyzing stool samples of fifty-one subjects ingesting pea vs. whey protein supplements in addition to their usual dietary intake and analyzing the effects on metabolites produced by gut microbes. The subjects completed a 24-hour food recall and provided a stool sample before and after seven days of incorporating the supplement into their usual diet. Bacterial composition was characterized using 16S rRNA sequencing. To account for other effects on microbiota composition, the subjects completed a lifestyle and International Physical Activity Questionnaire. Fecal nitrogen, metabolites, short chain fatty acids, and lipidomic profiles were also measured. With thousands of data points included in this experiment across twelve different datasets with variable structures and identifiers, it can be difficult to perform an in-depth analysis in a manner that is efficient. My role in this project is to use computer programming to develop a schema which will allow for a timely and organized analysis. The methods for this process include developing a data dictionary that is consistent across all files, discerning which data points are usable and which ones need to be transformed, and creating an architecture that allows for easy referencing and statistical computation between files. Ultimately, this framework will facilitate efficient integration and analysis of the datasets, allowing the research team to better characterize how dietary protein source influences the gut microbiome and its metabolic output.