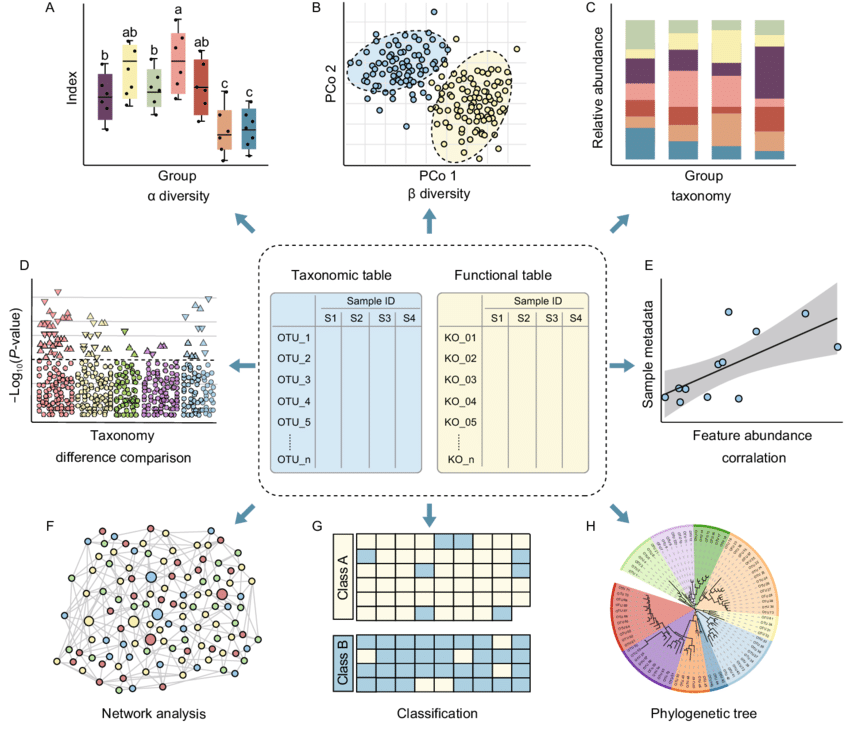
**WORK SAMPLE - SIWA**

Next-generation sequencing (NGS) technologies such as the **16S rRNA** gene sequencing yield useful data for describing microbial compositions in an ecosystem. These data support a core aspect of the work that SIWA does.

Following initial quality assessment/control steps, such as primer(s) removal, demultiplexing and quality filtering, the 16S amplicon sequences are clustered into **Operational Taxonomic Units (OTUs),** each OTU representing a unique sequence that may correspond to a specific bacterial group (strain, species, genus, etc). After the construction of OTUs or ASVs, these observed counts are typically organized into a large matrix referred to as the feature table.

The observed counts of features represent observed abundances of taxa in the sample. Since abundances in a feature table represent only relative information regarding each taxa, these are compositional data. These features and their associated counts can be grouped or subset into different taxonomic levels, transformed using various methods, and analyzed in a large number of ways.

****

\*functional table: features table

**Concepts:**

* **OTU/ASV/Features Table:** tabbed text files specifying feature vectors and categories for a set of observations. In OTU analysis, features are OTUs (rows) and observations are samples (columns) (QIIME2 convention).
* **Taxonomic table:** taxonomic classification assigned to the sequences.
* **Relative abundance of a taxa in a sample:** fraction of the taxon observed relative to the sum of all observed taxa corresponding to the sample in the feature table. Thus, the relative abundances sum to 1 (data might be given in an absolute count and transformation is required).

**Data:**

These data were generated from chickens in a research trial. Samples of intestinal contents were collected from 2 locations in the intestines, and the 16S gene of bacterial DNA isolated from these samples was sequenced. The metadata file allows us to associate OTU data with other traits of interest at the level of the sample (i.e. sample type), the animal (i.e. species), or even the farm (i.e. geographic location).

Otu\_taxa\_ws.txt: otu table including taxonomic classification.

* metadata\_ws.txt:
  + SampleID
  + AnimalType
  + SampleLocation: location of the animal’s body → Cecum, Ileum, Feces.
  + BWG: Body Weight Gain

**Tasks:**

The goal of this exercise is to provide you with a realistic sample of our work that includes aspects of data management, bioinformatics, and visualization. With this in mind, we have provided you with a small dataset that we would like you to explore and describe. If you have preferred tools or methods, you are free to answer these questions in your own way, but [Qiime2](https://qiime2.org/)is a popular suite of tools used for microbiome analysis, and their tutorials may be a good place to start for those unfamiliar with sequence data.

* Please provide a descriptive summary of the data: What does the dataset look like and how does the structure inform your decisions of how to manage it? How are the files connected? What features of the microbiome dataset are most abundant? Choose 1-2 figures that help summarize the data. You may choose any combination of figures to help summarize the data ([histogram, bar plots, tables](https://joey711.github.io/phyloseq/plot_bar-examples.html))
* There are several classic methods used to evaluate ecological diversity, and one of those is **beta diversity**. We would like to see a plot that explores this aspect of the microbiome. You can use any distance metric you would like (i.e. [Bray-Curtis, Jaccard, UniFrac](https://rdrr.io/cran/vegan/man/vegdist.html))
* Relating changes in the microbiome to changes in other factors of interest is an important part of our research. Please conduct a correlation analysis of microbiome features and the **BWG** data provided in the metadata file. You can use whichever method you prefer, and we recommend that you focus on a limited number of microbiome features (i.e. the top 10 most abundant).

**Deliverables:**

In addition to the figures and analyses we have asked for above, we would like a short (2-3) paragraph summary of your rationale for choosing certain methods, what you found in your analysis, and what questions you have about the data.

* Written summary of your exploration
* Visualization of your analysis

**Help:**

[**https://rstudio-pubs-static.s3.amazonaws.com/560496\_e117b605642847bfa5cfb6201e3bdb26.html**](https://rstudio-pubs-static.s3.amazonaws.com/560496_e117b605642847bfa5cfb6201e3bdb26.html)

[**https://www.onecodex.com/blog/2019/04/25/onecodex-jupyter-notebooks-for-data-viz/**](https://www.onecodex.com/blog/2019/04/25/onecodex-jupyter-notebooks-for-data-viz/)

**References**

[**https://www.nature.com/articles/s41522-020-00160-w/tables/1**](https://www.nature.com/articles/s41522-020-00160-w/tables/1)

[**https://qiime2.org/**](https://qiime2.org/)