

## **Excel Tutorial Part 2: Understanding your Microbiome "DataSummary" Document**

By Jimiane Ashe

Welcome!

This tutorial is brought to you by the Research Experiences in Microbiomes Network (or REMNet) and the City University of New York.

In this video you will learn about the contents of the Data Summary file provided by the Authentic Research Experience in Microbiology (or AREM) program of REMNet.

Prior to receiving this Data Summary file, you and your students collected samples in your environment and sequenced the microbiome for those samples using high throughput next generation sequencing technology. Raw data was generated for you using a microbiome bioinformatics pipeline.

Please open your Data Summary file in Microsoft Excel to follow along. To ensure that your tab-delimited document opens successfully in Excel, you can right click on the file name with your mouse cursor, select the second command option from the list called Open With, and select Microsoft Excel from the selection window that appears. As an alternative, you can open the Excel application, click on File, select Open from the pull down menu, and browse for your Data Summary document.

We will open and define the contents of each sheet within this document. Look toward the bottom of the screen as we move through each one.

The first sheet is the Mapping File. The Mapping File contains a table with specific and defining details about each of your environmental samples. The qiime bioinformatics pipeline used this information to analyze and organize your microbiome sequence data.

The second sheet is the Taxa Summary. The Taxa Summary contains your microbiome raw data. Column A lists the Operational Taxonomic Units (or OTUs) found within your project. Subsequent columns are your environmental samples with relative abundance reported for each OTU within that sample.

In this example, conditional formatting for the 10% most abundant OTUs in your project are highlighted in red.

The third sheet is the OTU Abundance-All. The OTU Abundance-All reports the number of DNA sequences from your environmental samples identified for each OTU in your project. The first column is a unique numerical identifier for each OTU. Columns B through H report available taxonomy for each OTU from kingdom to species.

Column I and on report the number of sequences for each sample. Again conditional formatting has been applied for the 10% most represented sequences in your sample. The final four columns of this sheet report statistics for your sequence data.

The fourth sheet is the OTU Abundance-LocationMean. The OTU Abundance-LocationMean reports statistically significant differences between the samples of your project, and it is organized using the same format as the OTU Abundance-All sheet.

The fifth sheet is the OTU rep seqs, which reports the representative nucleotide sequences for each OTU within your project. OTUs are listed by the numerical designations also found in the OTU Abundance sheets. This list may be quite large and sent as a separate data file. That is the case with this example.

The final two sheets report the biodiversity found within your project. The Alpha Diversity sheet reports within-sample OTU richness and abundance for your project using 5 different indices. Two column graphs are included. The first graph depicts richness using the chao1 and observed\_species indices, and the second graph displays abundance using the Shannon index.

The Beta Diversity sheet compares the community of OTUs for each of your samples. The values reported are generated by a principle coordinates analysis. Eigenvalues and the proportion explained for each component identified in the analysis for the entire project are reported at the top of this sheet.

The scatter plot displays the components that explain the most and second most variation between each of your individual samples. This plot is useful for identifying similarities between specific samples.

To summarize, you have learned how to open the Data Summary document in Excel and are now familiar with the contents contained within each sheet for your microbiome project associated with the AREM program.

Thank you for participating. If you found this helpful, we invite you to view our other tutorials guiding you through your microbiome data set.