Excel Tutorial Part 6: Making Microbiome Column Charts
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 how to create stacked column charts for your microbiome sequence data in Microsoft Excel.

In order to follow along with us today, you have opened and organized your taxa summary files, and you have converted your relative abundance data to number of sequences. If you need a refresher, please review our tutorials called Accessing your Microbiome Data, Understanding your Microbiome "DataSummary" Document, Organizing your Microbiome Data and Focusing on Taxonomy.

As you work with us today, you are welcome to use keystrokes or your mouse to utilize these commands and functions or to select necessary cells.

To orient you to this table, columns A through F contain the hierarchy of taxonomy for each OTU identified in the project. Columns G through I report the number of sequences for these samples.

For this demonstration we will focus on the top 20 most abundant Phyla in our Tree Canopy replicate samples.

#Select Phylum column > copy and paste to new column > Data > Remove Duplicates

Click Remove Duplicates.

We will use the SUM IF function to quantify the number of OTU sequences for each phylum.

Create new headers for each sample.

Type the = sign for access to the Excel Functions. Type SUMIF or select SUMIF from the list of functions.

Select your search range in column B, type a comma, select the cell with your phylum as your criteria, type a comma, and select the Tree Canopy 1 #Seq column as your sum range. Hit Enter, and apply this formula to the rest of the row using your favorite method.

Repeat these steps for Tree Canopy replicates 2 and 3.

Next we will calculate the AVERAGE number of sequences for our replicate samples by using the AVERAGE function.

#Type AVERAGE header
= > AVERAGE > Select sequence number cells in row > hit Enter
#Click on formula bar to review
#Fill Down

Now we will make a stacked column chart with the 20 most abundant phyla in our Tree Canopy samples.

First we will make some cosmetic adjustments to our data. Select your new Phylum and Tree Canopy columns.

#COPY > Select new column > PASTE SPECIAL > Select Values

Remove the p__ phylum prefix from your Phylum Column

#Select Phylum Column > Edit > Find > Replace… > type p__ in “Find what:” > leave “Replace with:” blank > Click Replace All

Excel will report the number of replacements to be made > Click OK

Next sort your samples by the Average Column

#Select the five new columns from Phylum to Tree Canopy Average
#Data > Sort… > Under Column select Tree Canopy Average > Under Order select largest to smallest > Click OK

We are now ready to make our stacked column chart

#Select the top 21 rows of the Phylum and Tree Canopy columns including your header row. Excel will incorporate phyla and sample names in your chart if you include them in your selection here.

#Insert > Charts > Stacked Column

The Stacked Column option is the most appropriate for our subset of data, and this method is commonly used to display microbiome data in publications.

Excel will display stacks for each sample as a default. We will manipulate this display to stack the Phylum for each sample group.

#Click on the chart > Switch Row/Column Tool

Depending on your version of Excel, you may need to right click on your chart, select Select Data from the menu, click the Switch Row/Column button, and click OK.

Now your data is displayed with your samples on the x-axis and number of sequences on the y-axis. Each Phylum is represented by a different color, and the legend lists those designations.

Note that these replicate samples differ slightly from one another. You can utilize a useful feature in Excel to explore the information displayed in these charts.

Hover your mouse cursor over a color segment of one sample.

#Hover mouse on eg. in TreeCanopy1 sample
#Hover mouse on eg. in TreeCanopy2 sample
#Hover mouse on eg. in TreeCanopy3 sample
#Hover mouse on eg. in TreeCanopyAVERAGE sample

Excel reports information from the original data cell.

Excel is also a useful platform to create charts like these because a number of additional customization options are available. Feel free to explore those features once you are comfortable building basic stacked column charts.

For example, we will add a title to our chart.

#In Charts? > Add Chart Element Tool > Chart Title
#Add an axis title or change legend as a second eg

You now have a graphical means of comparing and displaying your microbiome data.

To summarize, you learned how to apply the SUMIF and AVERAGE functions, how to create a stacked column chart, and some of the additional tools available to modify your charts.

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