**­­Microbial Diversity and Ecology Lab, Fall 2020**

**Week 11 Lab 19/20: On Your Own Exercise**

**Due**: Friday November 6, 11:59 pm

**Task**: upload this document in the assignment on WyoCourse; Week 11 Lab 19/20 OnYourOwn, with your answers embedded and rename the file by adding your last name at the end of the file name, e.g. Week 11\_Lab19/20\_OnYourOwn\_Assignment\_vanDiepen. Please save your file as a Word document.

**Points**: 4.0

**Objective**: In this exercise you will put the skills we have covered in the last few classes into action. Due to time constraints we have run through all of the raw sequence processing and provided you with a new phyloseq object. This phyloseq object comes from a study which examines the bulk and rhizosphere soil microbiomes of potato (Solanum tuberosum). These samples were collected in Lingle, WY at two points during the summer. The metadata includes a soil origin (bulk (B) or rhizosphere (A)), sampling time (early (1) or late (2)), pH, EC, and gravimetric water content. The sample names reflect the plant species potato (indicated with P in the first character), plot replicate (indicated with number 1-9 in the second character), sampling time (indicated with 1 or 2 in the third character; early (1) or late (2)), and soil origin (indicated with in the fourth character; A for rhizosphere, and B for bulk).

**Instructions:** Download the following files from WyoCourse Week 11 Module “Lab Materials”; On\_Your\_Own\_Student\_Version.Rmd and On\_Your\_Own.RData. Copy these files to your Desktop in your course folder, in which you could create a separate folder named “OnYourOwn” to keep everything together.

Open the Rmd file “On\_Your\_Own\_Student\_Version” and you can begin reading in the R tutorial and refer back to this document when you are ready to answer a question or complete one of the exercises. Make sure you save your plots and summaries to use for exercise 9!

Exercise 1: How many taxa are included in the original phyloseq object? Record here:

Exercise 2: Remove all non-bacterial taxa and name this new object ps. How many taxa are now included in this phyloseq object? Record here:

Exercise 3: Look at the number of reads per sample. Should we remove any samples prior to rarefying? If you chose to remove any samples, what sample(s) did you decide to remove and why? If you chose not to remove any, why did you decide this?

Exercise 4: Modify your dataset in two ways; 1) make a rarefied ps object and a within sample standardized/transformed ps object. Make sure to make each new dataset a different phyloseq object. Check how many taxa are remaining in your rarefied dataset, and record here:

Exercise 5: Using the rarefied ps phyloseq object you made in exercise 4, make a box and whisker plot to show the differences in Shannon diversity, Simpson diversity,and species richness (remember this metric is called "Observed" in phyloseq) between the two soil origins. Note: Remember that for the richness functions you can not use the transformed dataset. Make sure to add an appropriate title to the graph, and label the axes. You can check what the official names of your columns are in the phyloseq objects you have created using the code: sample\_data(PutTheNameOfYourPhyloseqObjectHere). NOTE: remember to load the package that is needed to make your plots.

Exercise 6: Run an ANOVA or Kruskal-Wallis test on the richness output to determine whether or not the differences in Shannon diversity, Simpson diversity, and species richness between the two soil origins are significant. Start by using the correct function to have phyloseq calculate and make the dataframe with the alpha diversity metrics. Then, you will need to add a column called Soil\_Origin to this new dataframe based on the sample name that is reflected in the rownames of this new dataframe (see line 7 for information on the sample names). This column will be made the same way you have done for the metadata and the richness dataframe in the phyloseq tutorial. Go back and check how you did it!

Exercise 7:

A) Ordinate your phyloseq object using "NMDS" & Bray-Curtis distance. Once you have done this, plot your ordination. Color the points based on soil origin, and base the shape of the points on sampling time. Give the plot a descriptive title.

B) Using the Adonis function, test to see if the bacterial community composition (based on the Bray-Curtis distance matrix) is significantly different between your soil origins. Remember Adonis testing is a permutation based MANOVA (lacks the assumptions of the normal multivariate ANOVA (MANOVA)). This test allows researchers to test for significance of treatments or uncontrolled covariates. MAKE SURE to load the correct packages to be able to run these codes.

Exercise 8: Create a plot which shows the phylum level abundances of both soil origins.

Exercise 9: Write a brief summary (1 - 2 paragraphs) that summarizes your findings. That is, can you interpret what you analyzed and graphed today? Similar to a publication, you will need to reference your figures. In addition, your figures should have captions.