

# Tutorial Assignment 1: Atacama soil microbiome analysis

Due: 10am (before class), 10/8

For this homework, we will be working with soil samples from the Atacama Desert in northern Chile. The Atacama Desert is one of the most arid locations on Earth, with some areas receiving less than a millimeter of rain per decade. Despite this extreme aridity, there are microbes living in the soil. The soil microbiomes profiled in this study follow two east-west transects, Baquedano and Yungay, across which average soil relative humidity is positively correlated with elevation (higher elevations are less arid and thus have higher average soil relative humidity). Along these transects, pits were dug at each site and soil samples were collected from three depths in each pit.

## Additional Notes:

- Continue to use the QIIME2 kernel at bio-datahub. As in the tutorial, use git-clone to clone the tutorial 1 repo. Work in the "TutorialHW1.ipynb" notebook.
- Remember to select the "Python [conda env:qiime2]" kernel
- Some questions below will require you to add your own cell blocks to this file and write a few commands.
- For this homework, upload both a PDF of answers as well as a PDF of your completed .ipynb.

## Questions

1. Calculate beta diversity with both weighted UniFrac and unweighted UniFrac. Then, create PcoA plots. Which beta diversity metric explains more variation in the first two principle components?
2. Do the samples cluster more clearly by "vegetation" or "transect name" on the weighted UniFrac PcoA? Provide a screenshot of your PCoA plot colored by the more explanatory variable.
3. Calculate alpha diversity. Is anything in the metadata associated with shannon diversity?

4. Test for differential abundance with ANCOM. Look at class (3) and genus (6) level taxonomic groupings. Is anything associated with vegetation?
5. Generate a taxa barplot of bacterial phyla sorted by % relative humidity. Which phyla are more abundant at low vs high humidity?