Group Name: Khoa, Huong

Our question: What is the ideal Magnesium level so that the most number of bacterial species could exist? Are there more bacteria and bacterial diversity in forest soil than Corn soil (in 2014?)

Rationale for choosing the topic: Magnesium has been said to be involved in the process of cellular division, and involved in the synthesis of cellular protoplasm. Gram positive bacteria have been proved to require as ten times the level of magnesium as gram negative bacteria. The absence of magnesium have been proved to be lethal to bacterial species, and many bacteria have been developing a strategy to attract magnesium into themselves.

From looking at the Soil Chemical Data in 2014, I see that samples F1-F10 have seemingly higher levels of magnesium than samples C1-C10. We want to find out that how strong the correlation of concentration of magnesium against species richness. If we have time, we will look at other nutrients too.

Hypothesis: There is higher gram positive bacteria richness in places with more greater concentration of Mg within both forest and crop samples.

Data: SoilChemData2014.txt

File to give Mothur: 2014 bacteria sequences from <http://cluster.earlham.edu/metagenomes/FieldThree/>

Citations:

Webb, M., 1949. The Effect of Magnesium on the Growth of Bacteria in Simple Chemically Defined Media. *Journal of General Microbiology*, *3*(3), pp.418-24.

\_\_\_\_

Some prelim steps to take

1. Run mothur on all 2014’s Crop and Forest Pre Samples to get taxonomy info - done
2. Extract taxon files as tables on Desktop - done
3. Do some R code on those tables based on the R code worksheet that Chris gave us – in progress

Decide on what we want to focus on: bacterial richness in general, richness of bacteria in terms of gram positive, gram negative, bacterial diversity

Try to calculate bacterial diversity in Mothur

Mothur Documentation: <https://www.mothur.org/wiki/MiSeq_SOP>

Some exploratory analysis, using:

1. Correlation plots of Mg\_ppm against gram positive bacteria richness in each community Crop and Forest (Pre)
2. Boxplots of bacterial richness/diversity variation within each community Crop and Forest (Pre)

Questions to ask Chris:

After running mothur on all 20 samples of crop and forest, we download the .shared and .opti\_mcc.0.03.cons.tax.summary files into Desktop. Here are some questions/processes we are considering:

1. How does removing uchime affect our results?
2. Difference between “.wang.tax.summary” and “.opti\_mcc.0.03.cons.tax.summary” files (the first one lists all species, the latter one only lists bacteria?)
3. Workflow for calculation of species richness and diversity in Mothur:
4. find the number of sequences in the smallest sample of all crop and forest samples, by running for all samples: mothur > count.groups(shared=file.shared)
5. subsample them, size=x, x is the number of sequences in the smallest sample: sub.sample(shared=file.shared, size=x)d
6. Get a table of species richness and diversity in each sample: summary.single(shared=file.shared, calc=nseqs-coverage-sobs-invsimpson, subsample=x)

4. The table has the following headers

label group nseqs coverage sobs invsimpson invsimpson\_lci invsimpson\_hci

What is a good coverage percentage? And what is the difference among invsimpson, invsimpson\_lci, invsimpson\_hci?

5. Is there a way to calculate species richness and diversity using R?

Mothur further workflow

1. ssh into whedon, load all modules needed to run mothur
2. cd into folder that contains the files for each sample e.g F1Pre
3. Type mothur
4. Type count.groups(shared=fake.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.opti\_mcc.shared)
5. Get the number of sequences e.g C1Pre - 145325

Number of sequences in each sample

C1Pre - 145325

C2Pre - 168722

C3Pre - 129805

C4Pre - 138121

C5Pre - 120990

C6Pre - 132292

C7Pre - 149661

C8Pre - 143013

C9Pre - 149862

C10Pre - 119409

F1Pre - 124731

F2Pre - 149473

F3Pre - 147793

F4Pre - 176544

F5Pre - 162847

F6Pre - 124250

F7Pre - 193949

F8Pre – 114045 – smallest sample

F9Pre - 144717

F10Pre - 115916