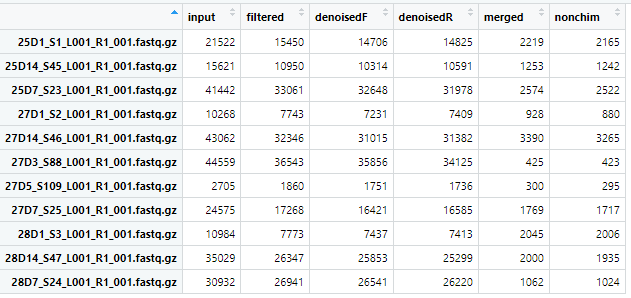
Michelle ITS notes

# Pre-processing

* Seqs are 251 bp in length, unexpected for ITS which is highly variable in length
  + Did the sequencing center do something wrong?
* Reverse reads are of very poor quality – truncating at 100bp produces decent merging results, but is it appropriate? Limits the certainty of identifying the correct ASV.
  + Could blast ambiguous IDs. Some fungal groups are more conserved than others, so at least a family or genus identification could be reliable.
  + Reverse reads are often worse, but typically more so in ITS (according to William King, Terry Bell’s old postdoc)
  + This would again point to sequencing center failure
  + Ends of rev reads are often very homogeneous – long seqs of the same bp. What could cause this? Too short of reads??
* The primers used are a bit unusual. I didn’t find them in any of the studies about efficacy of different ITS primers. What system did they come from?
* Changing maxN and maxEE parameters doesn’t significantly impact merge success, so it’s not a case of ambiguous bases scattered through the seqs. Errors are definitely coming from the rev reads.
* Show michelle how to successfully transform data into %s
* 100% transformation graph
* Code to only show the most prevalent, group others into other
* Filter within phylum for species
* pH script
* filtering, which statement

With c(250,100)



# ANOVA

When looking at Shannon diversity scores for all bacterial samples, samples from the Greenhouse have noticeably more range in value.

A graph with green and red dots

Description automatically generated

This is probably making it harder to see relationships between diversity and other variables.

I got rid of the Greenhouse samples and ran the alpha diversity analysis again.

There may be a significant differences between treatments (FSUW are about the same, and CHL are about the same), soil type (although since Growth room and Field don’t share the same soils, this may be tied to differences in experiment locations).

There is a significant difference in the Shannon scores associated with shoot\_dis category.

The next steps are to look for pairwise differences (i.e., F vs. C) when the variable in questions has more than two levels.