Preliminary Data

7.16.21

[DADA2 Tutorial on GitHub](https://github.com/amoliverio/dada2_fiererlab)

Cliff’s Alpha & Beta Diversity Tutorial in R...code in folder (caro\_16S.R)

Notes on sample prep

When preparing samples for sequencing, the tundra samples (NRt0) had to be diluted substantially to recover DNA. This was done following Jessica Henley’s recommendations in the Fierer Lab and after a single dilution, the DNA was recovered.

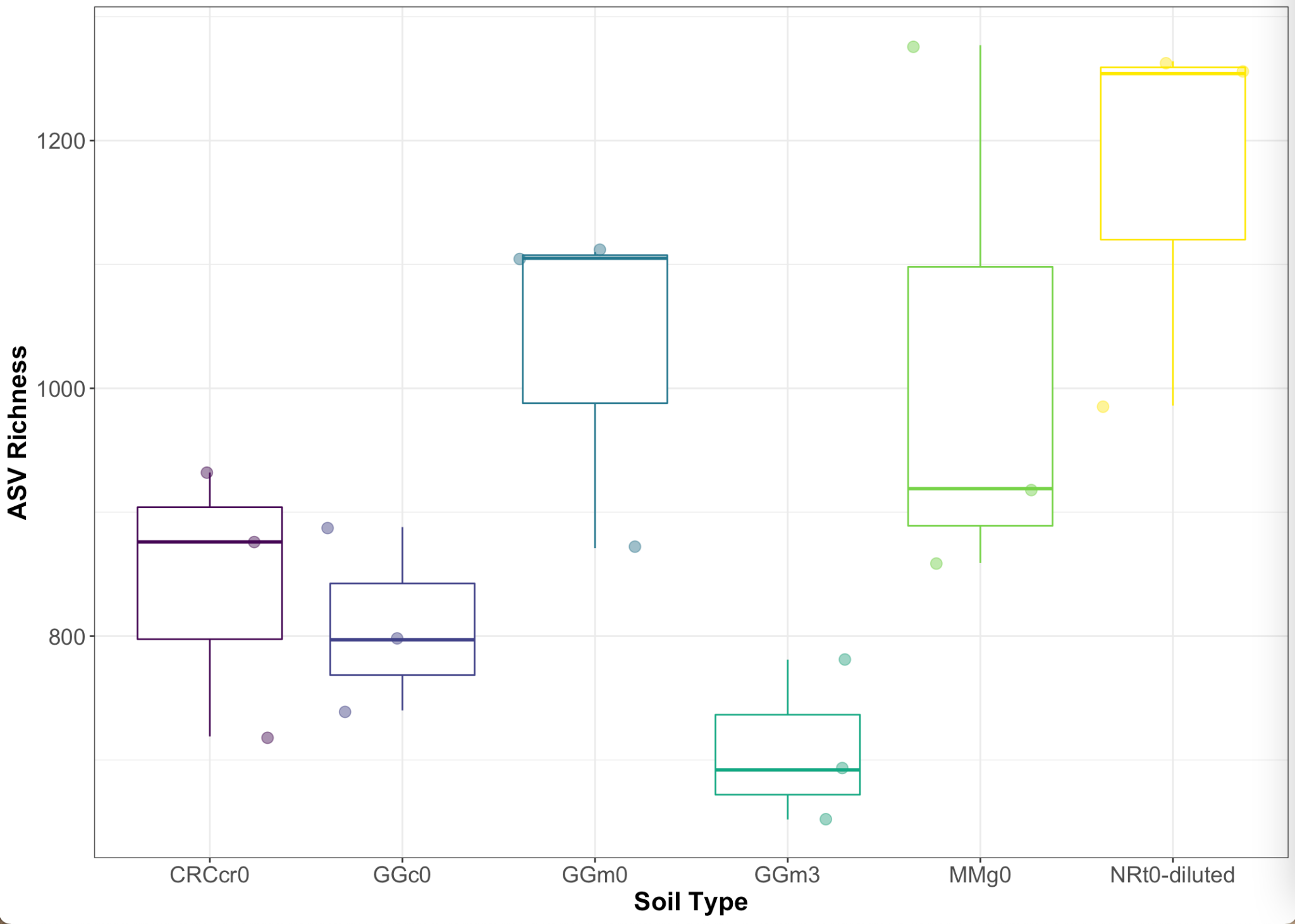
Notes on pre-processing

Took raw seqtab file and created a sum column. Sorted the sheet by the sums column and deleted all ASVs with zero reads. Deleted the sum column. Saved sheet as .txt

Created a mapping file called caro\_mapping.txt with basic sample information for R analysis.

**Findings**

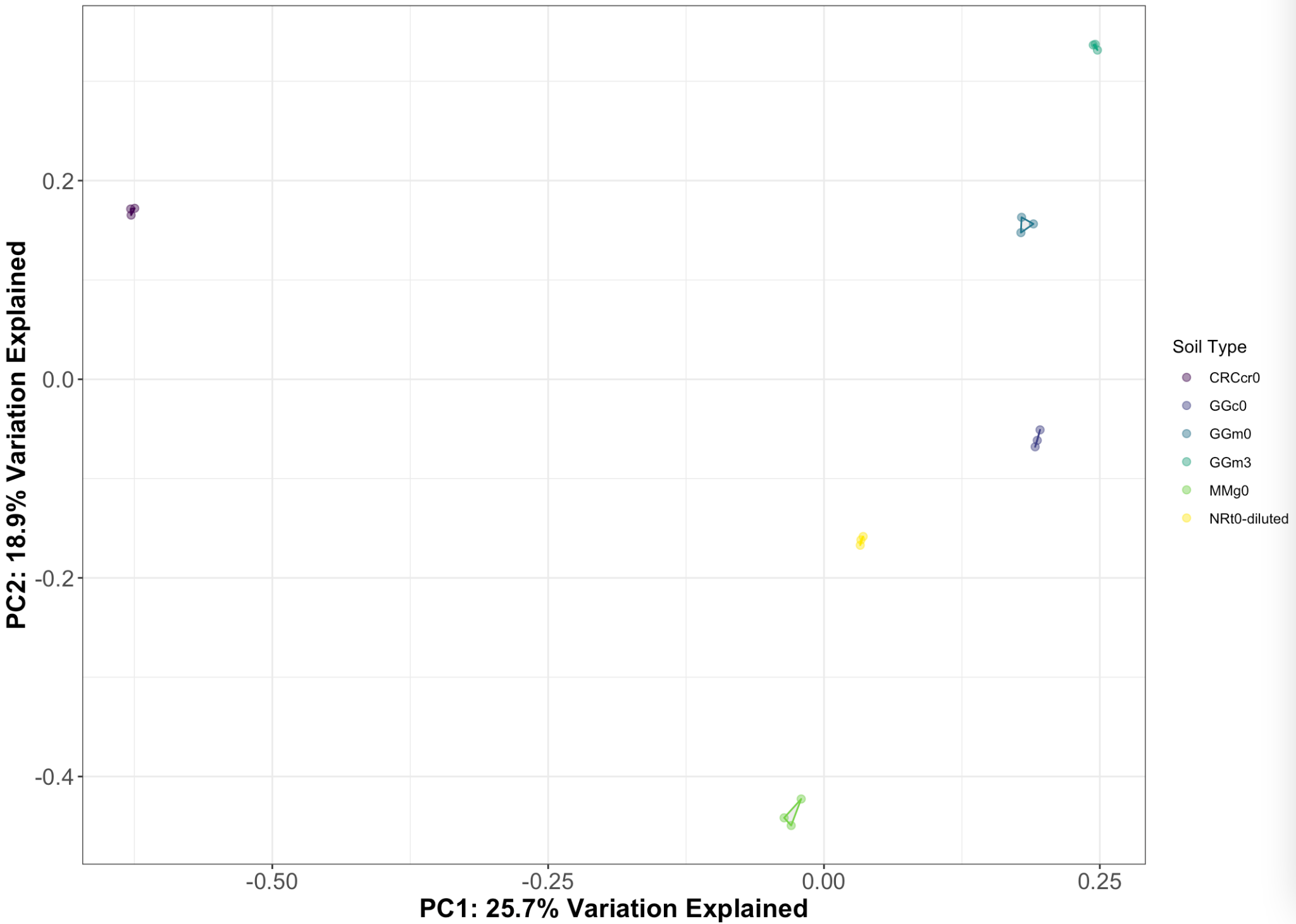
1. Richness



NRt0 is significantly greater than GGm3 (ANOVA, p<0.05). The rest are not different from one another.

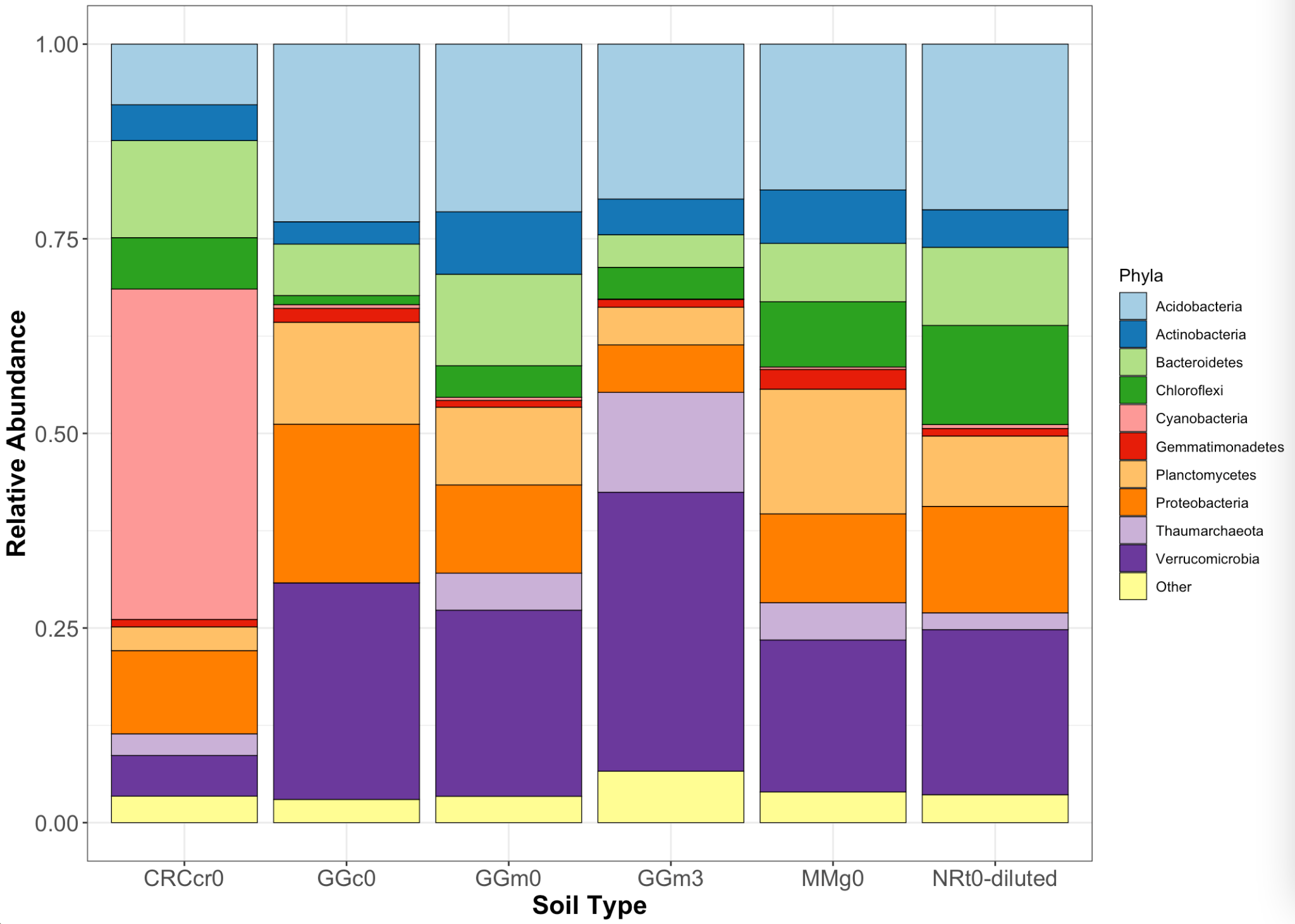
ASV richness (high to low): tundra, meadow surface, grassland, desert, conifer, meadow subsurface

1. PCA

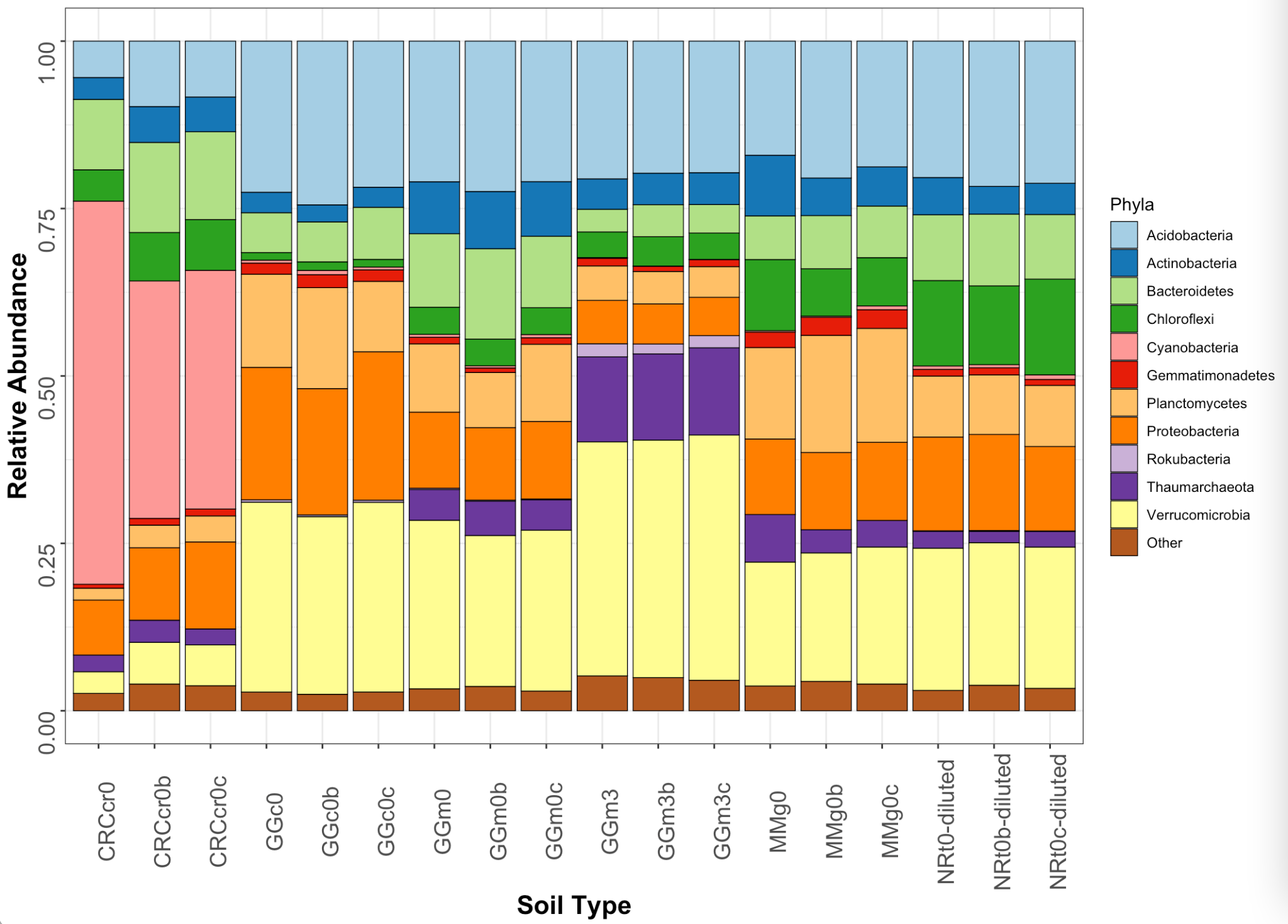


The first axis is largely controlled by the desert samples. This separation is also apparent when you see the heatmaps below. The second axis from top to bottom: meadow subsurface, meadow, conifer, tundra, and grassland. I am curious about abiotic controls on this trend in community composition (temperature and precipitation). If this is true, we may be seeing a left-right gradient in precipitation. I am not sure what the top-bottom one would be.

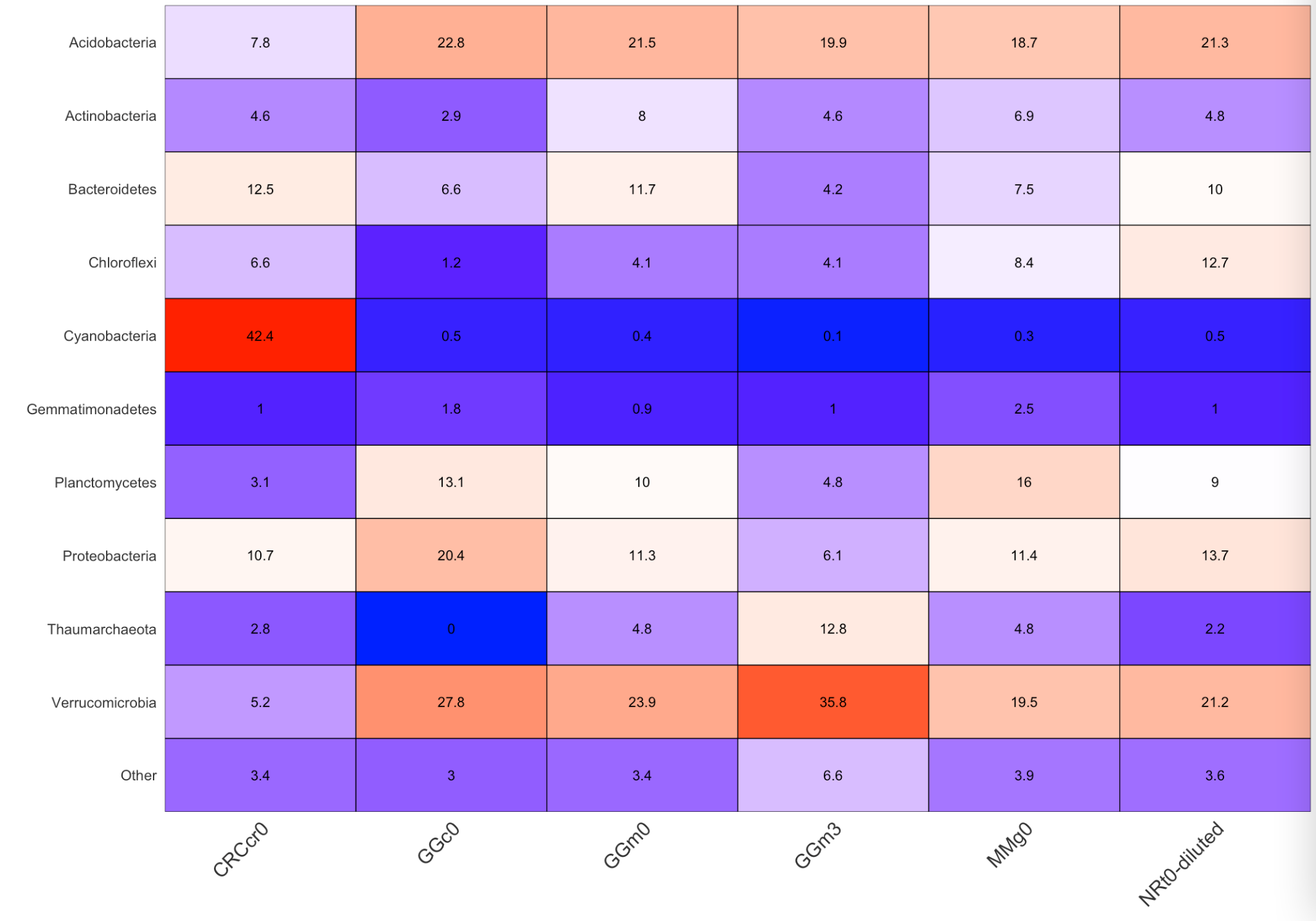
1. Stacked bar chart of 10 dominant phyla with samples averaged (n=3).



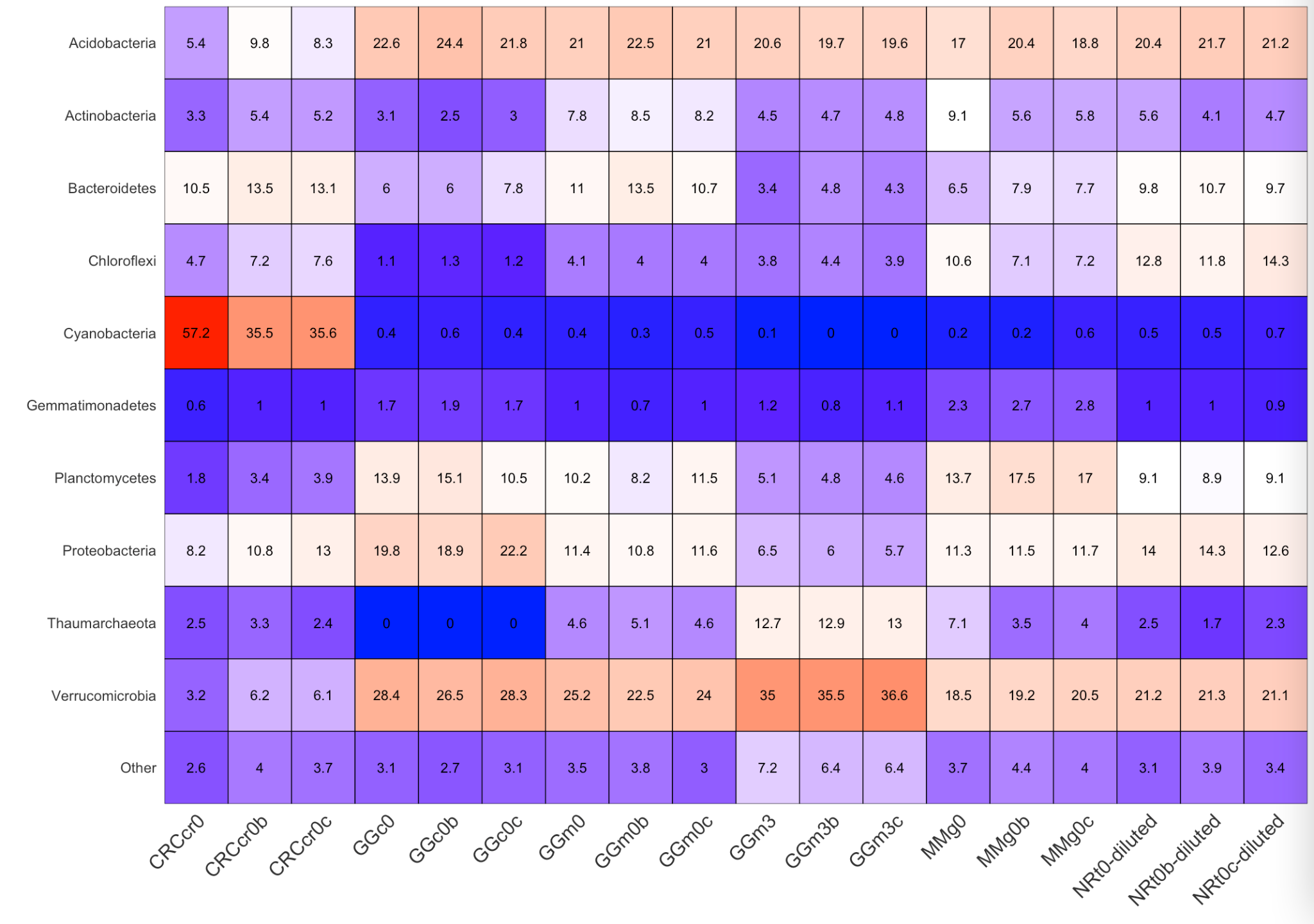
Stacked bar chart of 11 dominant phyla with samples kept separate (not averaged)



1. Heatmap showing similar information as is #3 above. Here are the 10 dominant phyla for the different soil types (n=3). Note: Cyanobacteria in desert samples, Verrucomicrobia in subsurface meadow samples.



Heatmap showing similar information as in #3 above. Here are the 11 dominant phyla for the different soil types (not averaged)

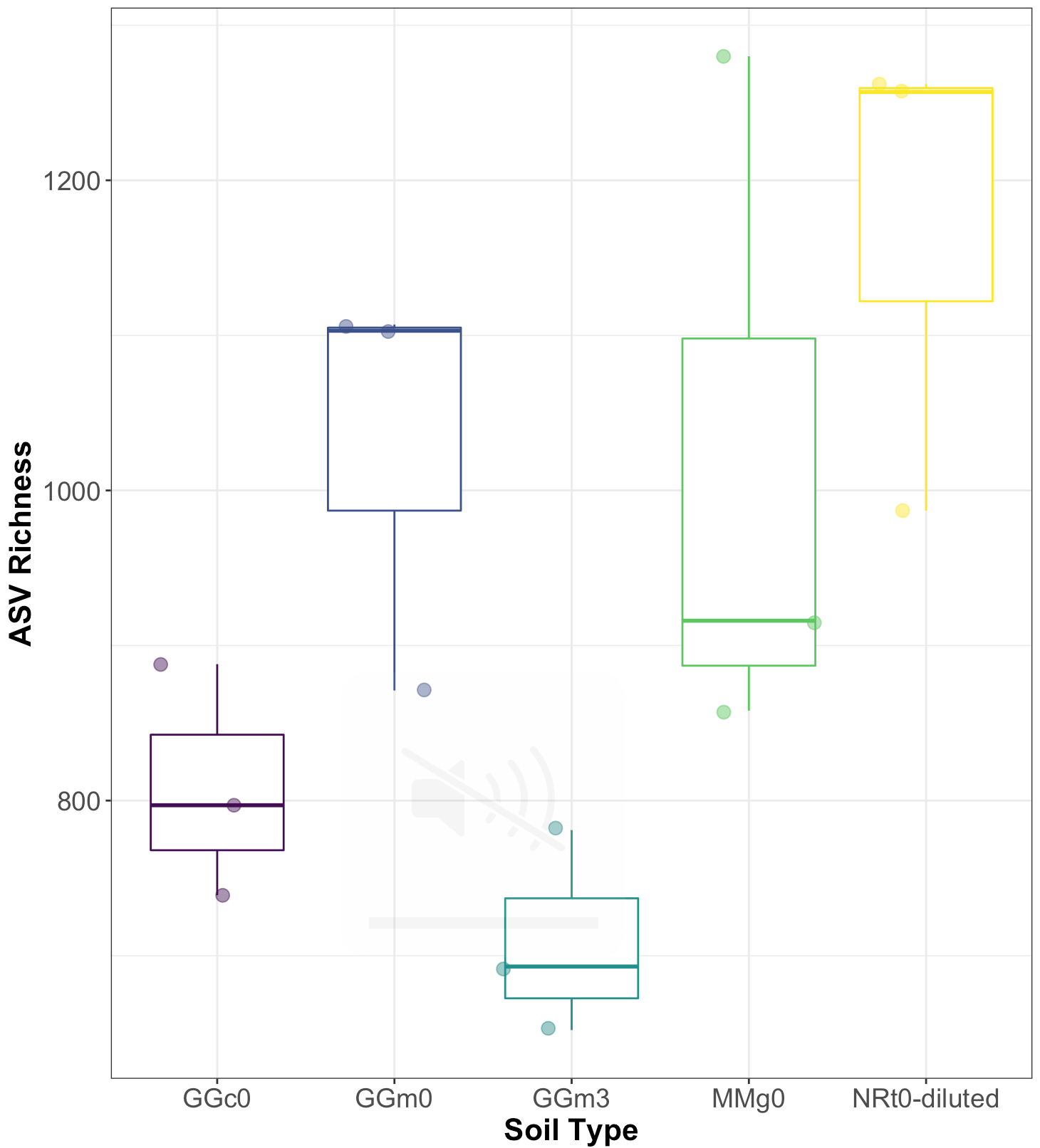


Thoughts for next steps. Obviously, the biocrust samples are very different from the others. There are other interesting trends to consider in the non-desert samples. For instance, the differences between the surface and subsurface meadow samples. It would be worth investigating other papers with microbial community data for each of these general soil types and seeing if we match reasonably well.

**7.20.21**

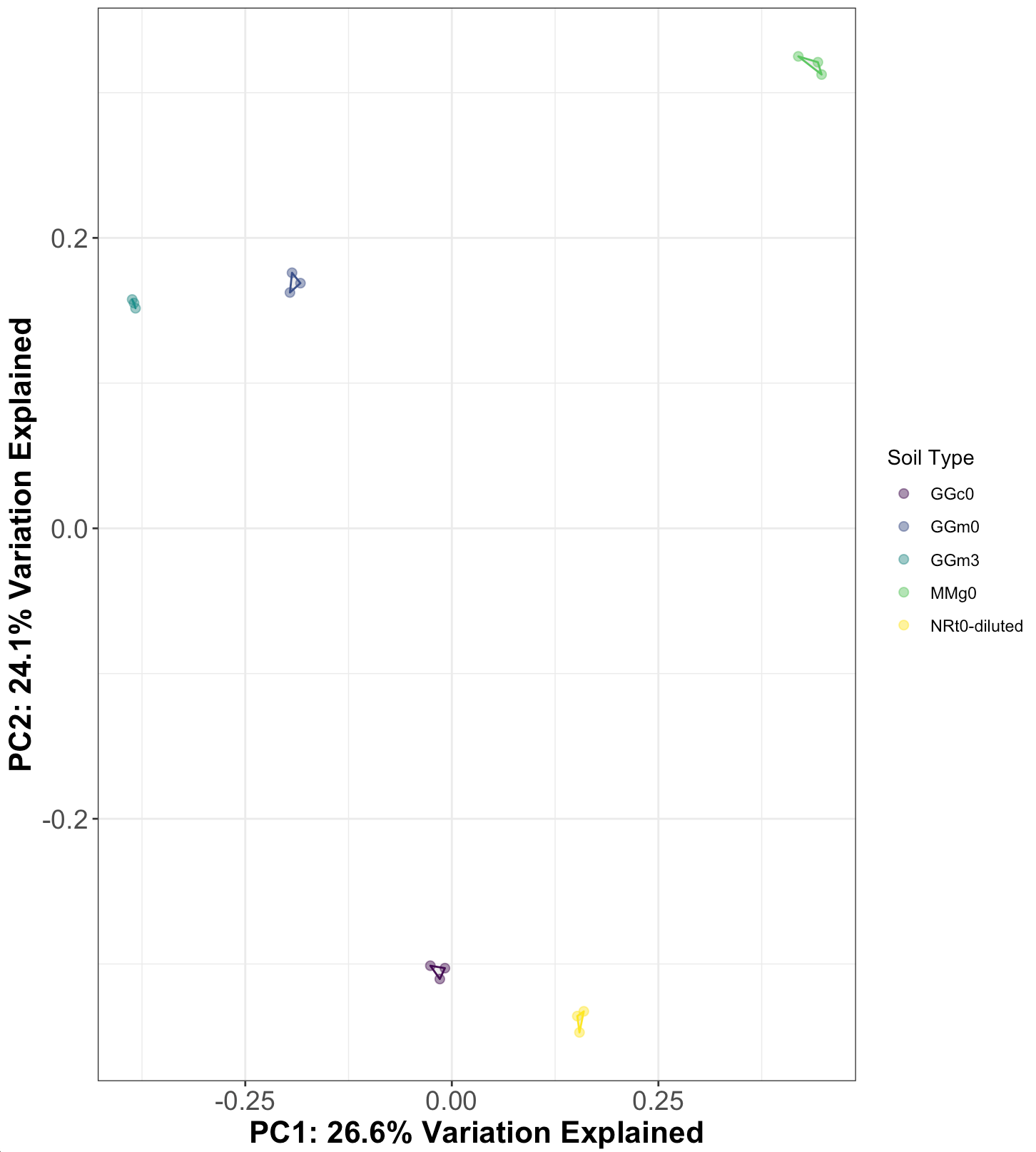
**Biocrusts removed**

Richness (number of ASVs)

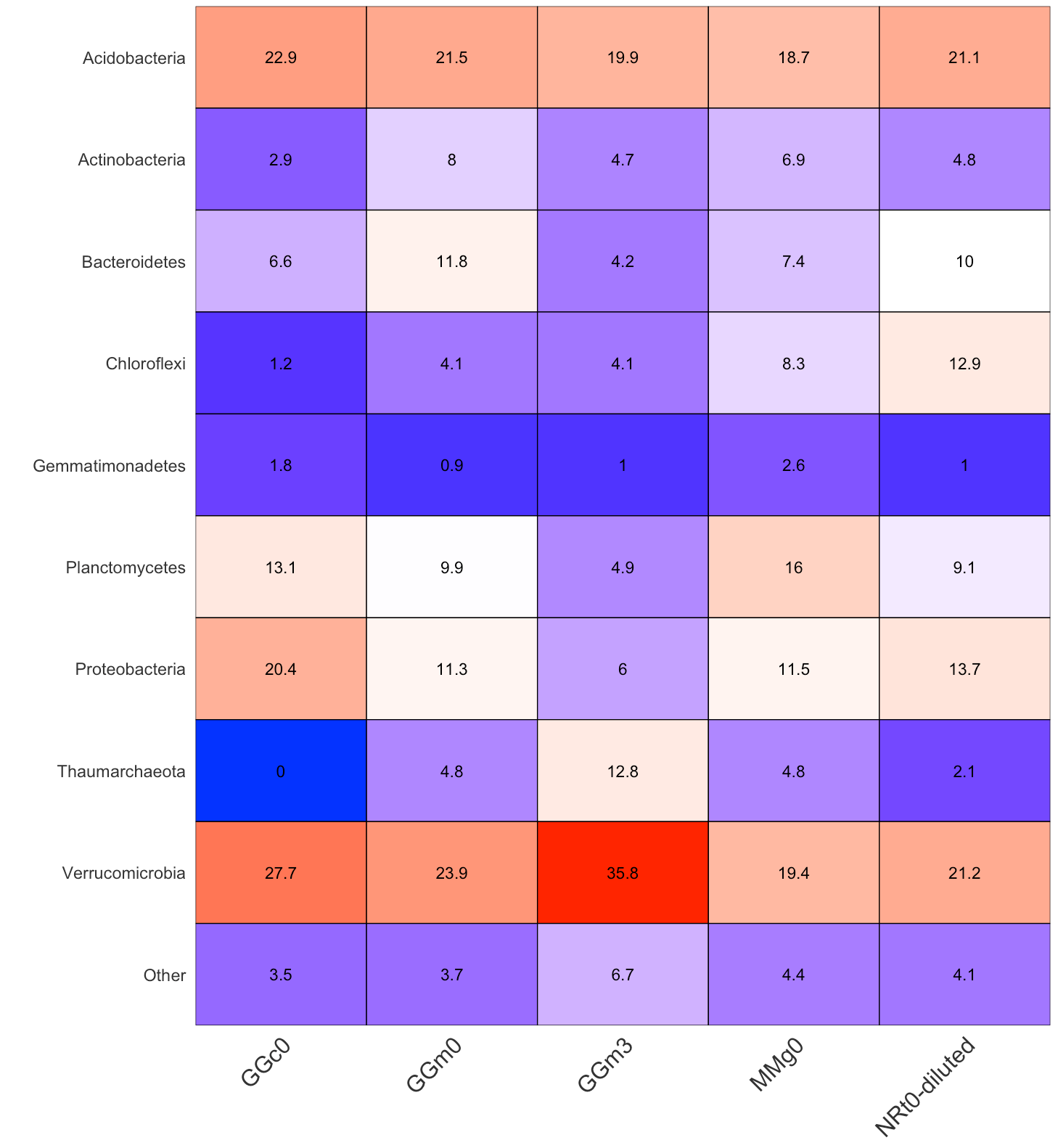


ASV richness by soil type (ANOVA, p=0.0194). The only significant difference is between the NRt0-diluted samples and GGm3 (TukeyHSD, p=0.0200).

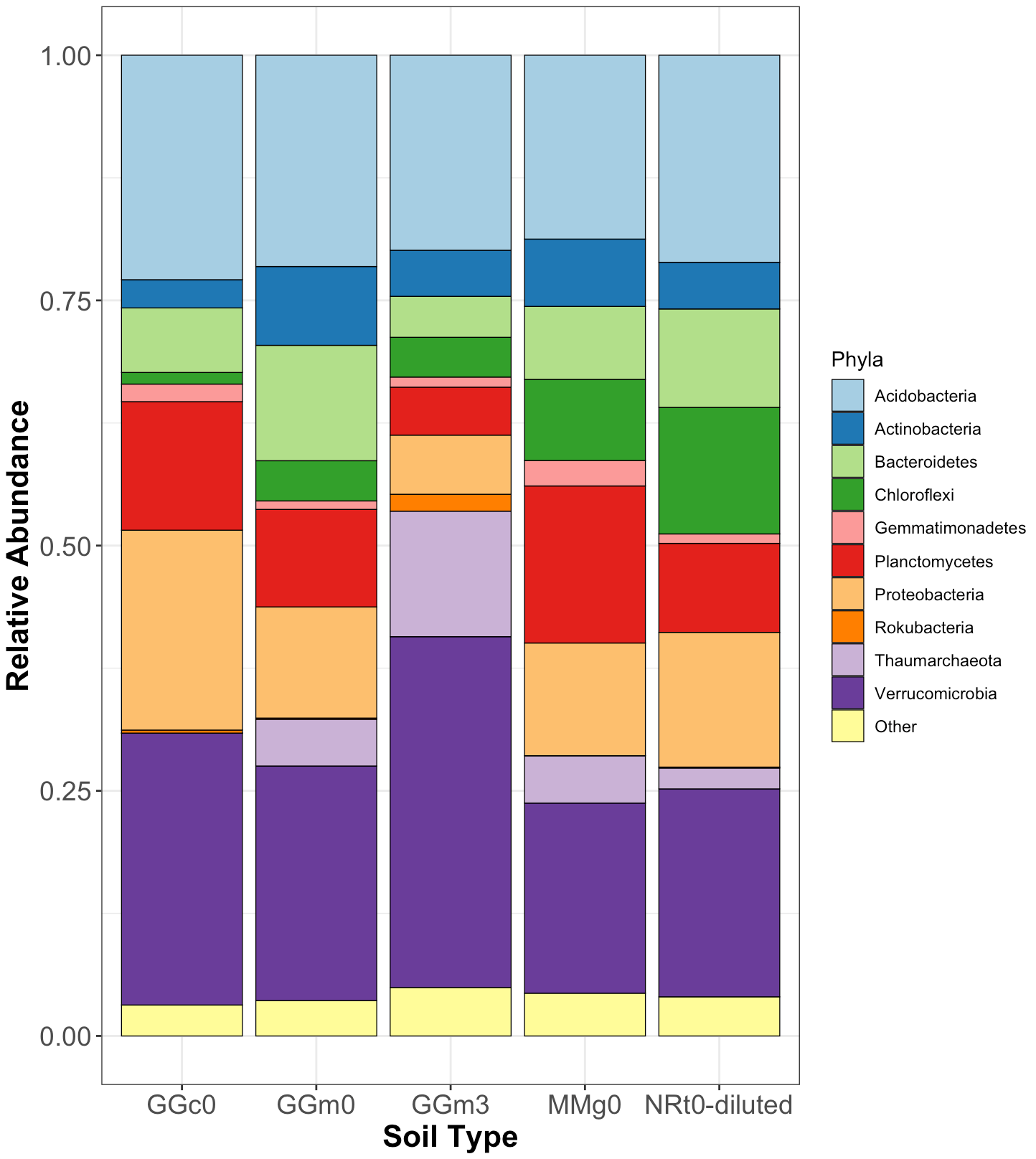
PCA:



Tundra and conifer forest separate from meadow and grassland samples in the y direction. Again curious about moisture and temperature for these axes. The centroids of the hulls are significantly different from one another (PERMANOVA with adonis, F = 11.307, p=0.001). Dispersion is not different (PERMDISP, F= 0.777, p=0.5644).

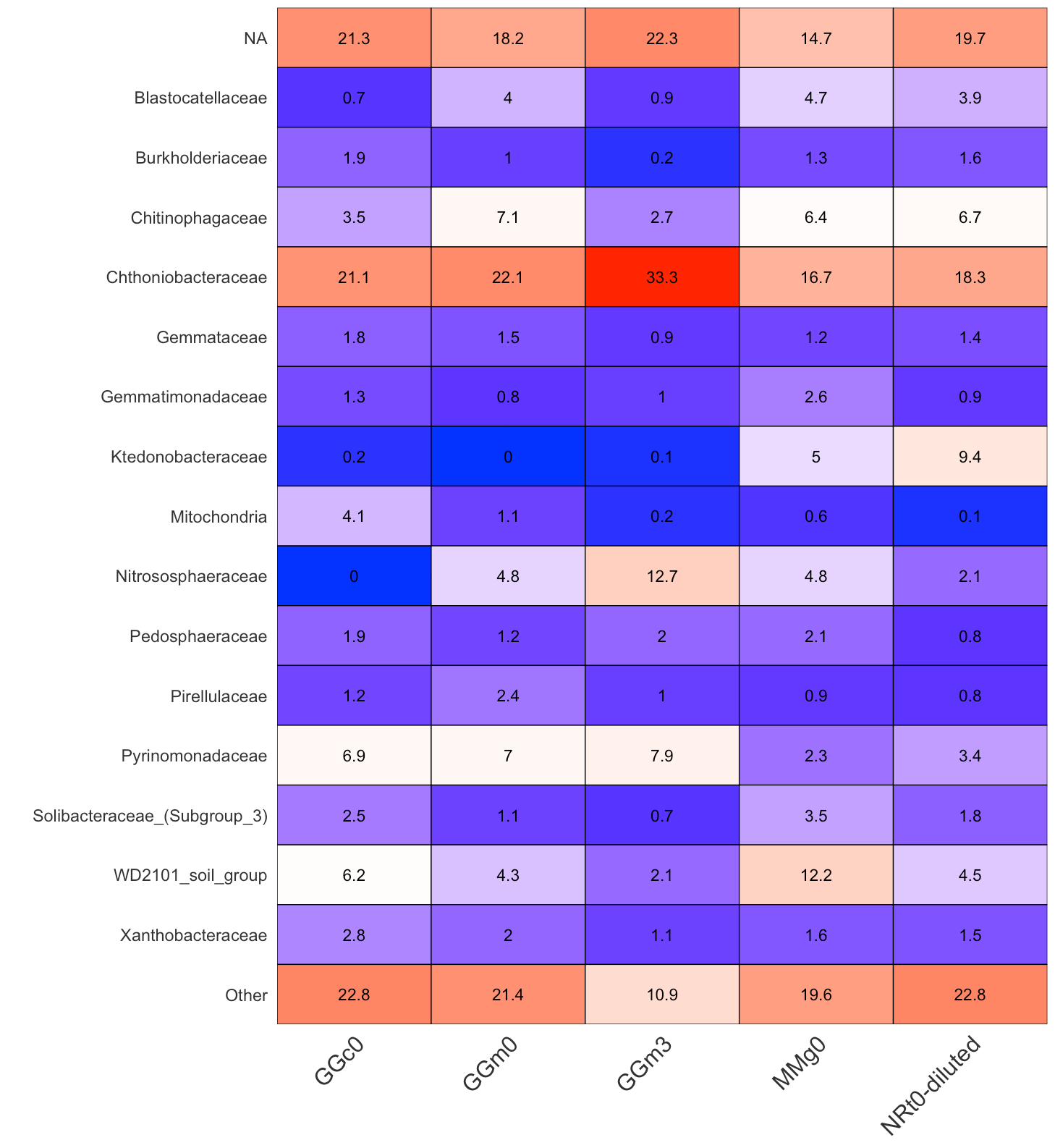


Verrucomicrobia is ubiquitous across samples, but more important in the subsurface. Planctomycetes are more common in the grassland and conifer forest. Chloroflexi are more common in the tundra. Proteobacteria are more common in the conifer forest.

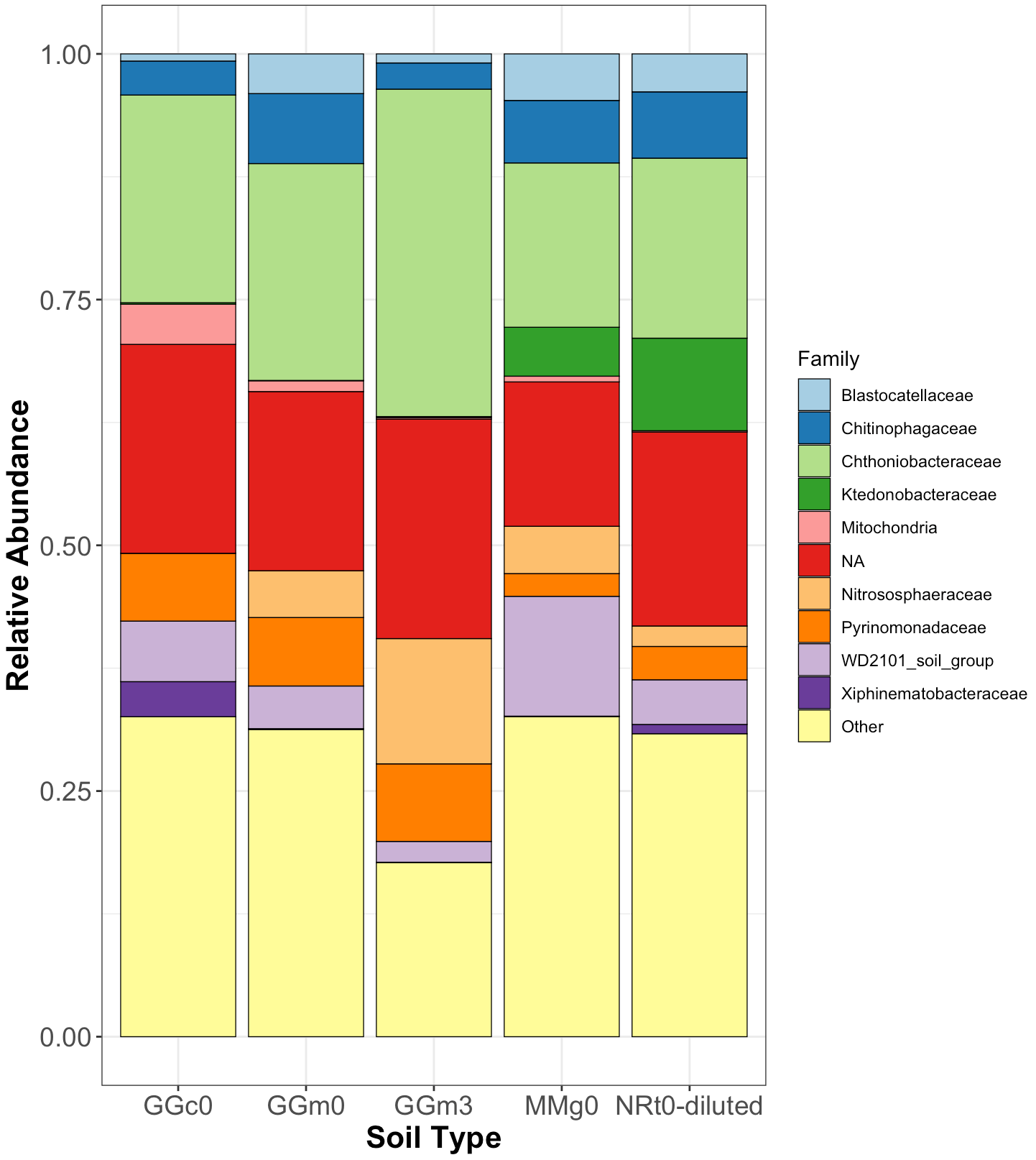


Kruskal-Wallis test on all phyla with mean relative abundance greater than 0.05 and Bonferroni correction on the p-value, no significant differences in the dominant phyla by soil type.

Family level



Mitochondria seem to still be in the data set...need to remove



At the family level, Kruskal-Wallis insignificant as well. Need to deal with the NA values here.