# Progress Report

Charles Bond

11/29/2020

**Current Status: It finally ran! But this code is for cleaning the data, not producing the figures. I am producing the figure myself using the `autoplot` function.**

## Individual kingdoms

```
library(tidyverse)
```

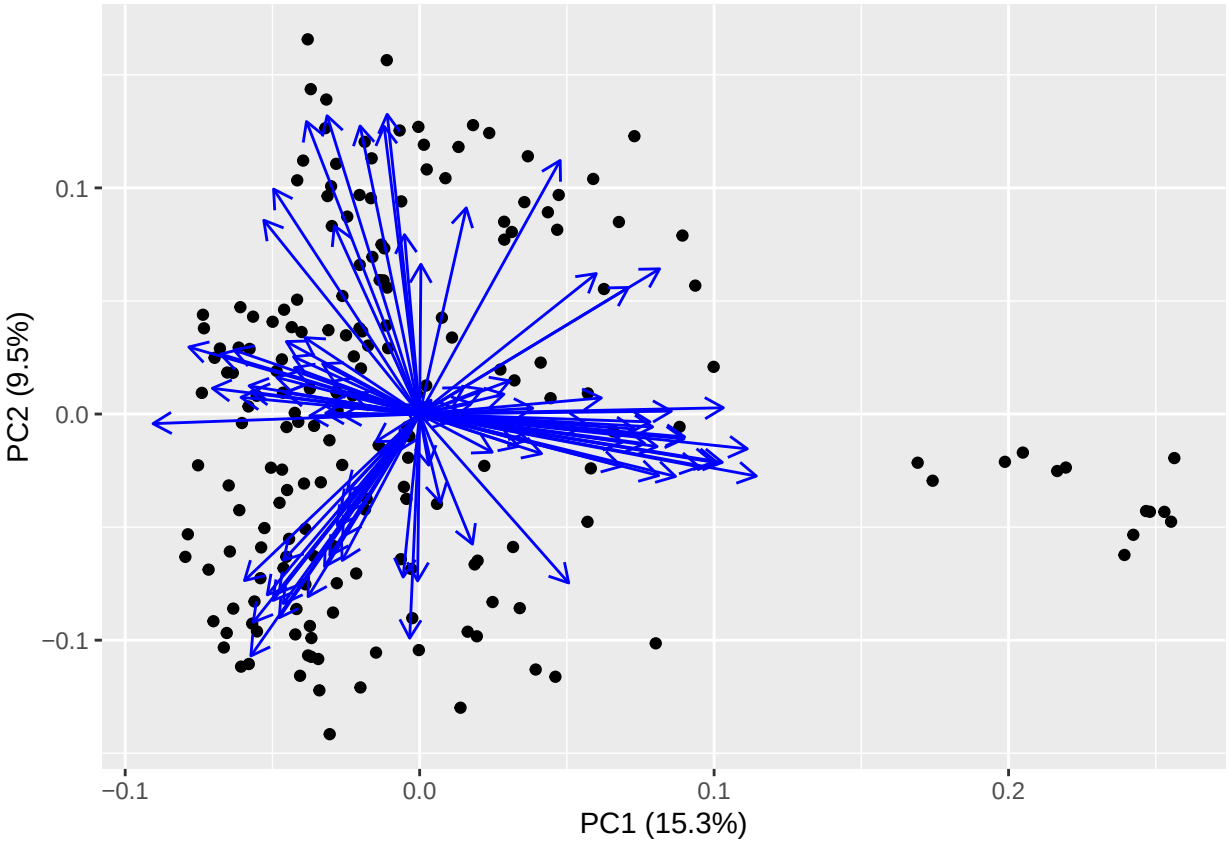
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x tidyr::pack()   masks Matrix::pack()
## x tidyr::unpack() masks Matrix::unpack()
```

```
library(ggfortify)

autoplot(pca.obj, loadings = 1, loadings.colour = 'blue')
```




Now, autoplot works, that's great, but a few things jump out to me at this point. First is minor and perhaps easy to fix: I'm not sure how to control the number of loadings in the plot. I've searched and tested things like `loadings.number =` but so far no dice.

More problematically, this first script here is supposed to produce two PCAs: one for fungi, one for bacteria. However, the chunk produces just one pca object, apparently fungi (as I explain below), which makes me think maybe it overwrites the results of bacteria pca when it makes the fungi one? But that wouldn't make sense. Maybe the fungi and bacteria are all in there and autoplot is just grabbing the fungi for some reason? I read through the script as best I can, it is definitely set up to run the analysis for each kingdom separately, and fungi are subsequeent to bacteria, so I think it must be overwriting it.

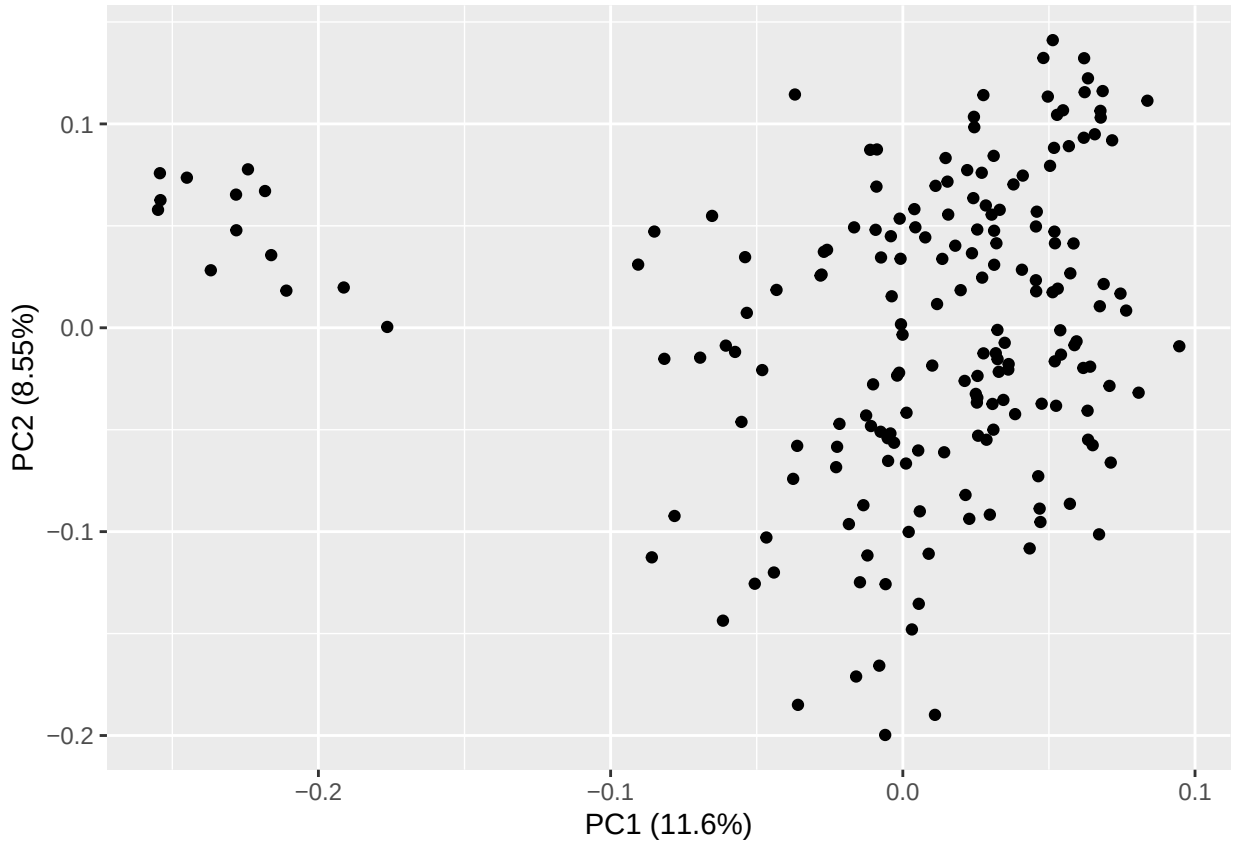
The plot shown looks like a 180 degree rotation of the published plot (Figure 3b.), made evident by the axis labels "PC1 (15.3%)" and "PC2 (9.5%)", just as in 3b. However, not only are both the x and y axes apparently flipped, but the scale is different, *an eerily similar problem to the NMDS I tried to replicate from my previous paper*. This makes me wonder if something about my system is flipping and scaling ordinations differently in general? Or maybe both papers plotted their ordinations using different functions, packages, or software than I am? These authors say they did it in R, and it looks like it was done in R... I know the scaling and orientation in ordinations are relative and not essential to the analysis, I just think is very curious that my system keeps doing this.

So, while I try to figure how to get 2 seperate plots for the individual kingdoms, I'm also going to go ahead and ...

## combined fungi and bacteria PCA

```
library(tidyverse)
library(ggfortify)

autoplot(compositeCommunity.pca, loadings = 0, loadings.colour = 'blue')
```




The percentages shown on the labels for PCC1 and PC2 here are identical to Figure 3c, the combined plot in the publication. Also, while the scaling here is again off by a factor of something like 70 or 80, it *is* oriented correctly, not rotated or flipped. So, that's something.

So, now what I have left to do is tinker with these plots to try and make them more consistent with the published plots. Yay!

# Previous updates:

**Before it worked, here was my troubleshooting on 11/25/20:**

After fixing the initial directory issues pointing out by Dr. Taylor, here is the error message for *both* the combined and individual kingdom PCAs: "`- Rescale variables?Error in setwd(paste0(Sys.getenv("ROOT_MICROBIOME_DI "/gwas/otus97", :... ... cannot change working directory`"

Now, I figure I need to change something to allow it to change directories mid-loop... I am deleting the `/` and leaving `"gwas/otus97"` as per Caz's earlier directory advice. Did not work in either chunk. Neither did

`"./gwas/otus97"`. I have double checked that the folder "otus97" is indeed within gwas, which is in turn within thee initial working directory which otherwise appears to work. So, is there some other reason that it cannot change working directories? I did some research on this error with the `setwd()` function, usually the error involves someone trying to set a file rather than a folder, doesn't apply to me. ### 11/29/20 Okay, I figured it out. The original code was `"gwas/otus"`, not `"gwas/otus97"`, but there is no `otus` folder in gwas, only `otus97`. So while I was still having other problems, I had added the `97` hoping that would fix it. Now I have no idea why this is working, but it works, despite the folder being `otus97` and not `otus`. ...

**11/15/20**

I've read and reread the methods section of the paper and all the READMEs and still don't know what I'm missing. I've run the code in the original scripts and in this RMarkdown document and get the same result: for some reason it can't source the methods files and then something in the loop fails but it won't give me the exact line, just the loop. This code is no simple PCA. In each of the two PCA scripts, the first section of code sorts the raw data in preparation for a loop that is supposed to generate the PCAs. I can get that first section of code to run, and then once it enters the loop I think something fails in the first iteration that gives me a long list of errors. This is after checking that the files being called up are in the working directory. I tried manually sourcing the methods code, doesn't help. Quadruple checked that the files aare in the working directory, but they still come up as errors when the script tries to access them.
The errors make me think that the folder system for organizing the scripts is perhaps too complicated for replicability? The code is only from 2019, would package updates or anything be making it not work? I don't think so. But maybe in the loops it's failing to access some file that it should be able to? I can't even load the data to run a prcomp on, since I require the code to clean up and spit out the actual data used in the original analyses. This was written for people who know R better than I do.

Looking for the tools to run the PCA analysis shown in figure 2 (Bergelson and Horton 2019), I have been unable to run the code so far. I appear to have all the files, code, and data from the two githubs and the data repository provided in the **Availability of Materials and Data** section of the article of interest. The code looks good, but I can't get past this setup error: the script seems to be having trouble accessing the files that came from the repository.

**Old Update, 11/08/2020**

Based on the availability of the data and the posting of the code on github ( https://github.com/bvilhjal/mixmogam and https://github.com/mahort/root_microbiome. ) I have selected a paper on the *Arabidopsis* root microbiome (Bergelson and Horton 2019) as the subject of my next replication attempt. If it goes well there are a lot of figures in this paper that would be worthwhile to replicate, but the one I am mmost interested in is the PCA shown in Figure 2 of the paper.

Once again, this paper uses the package `vegan` to do the PCA, so I've got that.

I downloaded whole github repository as a zip file and put it here, but I guess I need to alter relative paths or something to make it work? I tried running one of the PCA Rscript and get the error "Error in getData(cutoff = otuCutoff, combined = FALSE, marker = phylogeneticMarker, : unused arguments (cutoff = otuCutoff, combined = FALSE, marker = phylogeneticMarker, tissue = organ, minReadsPerSite = minimumReads, minSamplePerOTU = 1, whetherToGetTaxa = TRUE, rarefy = (minimumReads > 1), numberOfRarefactions = 1)", which idk what that means but I guess I need to tinker with this more.

## References

Bergelson, Joy, and Matthew Horton. 2019. "Characterizing Both Bacteria and Fungi Improves Understanding of the Arabidopsis Root Microbiome." Journal Article. *Scientific Reports (Nature Publisher Group)* 9 (1). https://doi.org/10.1038/s41598-018-37208-z.