REMNet Tutorial, R Part 5: Normalizing Microbiome Data in R

By Jessica Lee Joyner

Hello, my name is Jessica Lee Joyner, and I teach and do research at the City University of New York.

One important consideration when you're working with microbiome data is: how do you control for sequencing variability? One common way is to do what is called normalizing your data, or in some cases in different processes it would be called rarifying your data.

What I'm going to do next is normalize using a relative method to see what could be a different perspective on a more even scale of comparing across sites because sometimes the number of sequences we get for each sample could buy us what we see as far as the taxa present and the diversity. So I'm just going to keep working with this taxa bar chart, but I'm going to now normalize the data and present the taxa bar chart with a more even or relative abundance of those taxa present.

The first step is to create that transformed Phyloseq object again. So we have our BioBlitzdata, and I'm going to transform it with a "t", and save that as a new object called BioBlitzdata now "_MseqT". Then I'm going to take that new Phyloseq object, and put it through the function called "normalise_data", and the method that I wanted use is "relative".

You can find other ways that you can normalize by looking up the function and going to the, um, Internet. So here I can do a quick search in the help database, so this is a good point to show that if there's ever a function that you want to look a little bit more into in R, you can do a question mark and then the name of that function. And then that came up on our screen here, and gives us all of the information in this bottom right as far as what is going on with this function. So we're going to normalize the taxa abundance. It gives us our usage and then the arguments, so what information goes into each. So our Phyloseq object and then the method, and here you see where it'll list out the different methods you can do.

And I did a little bit more digging to learn that we can do relative as well. So that's how I'm going to use the function, and you see some other examples here of how the function can be used. OK, I'm going to bring that back down, and run these lines of code.

So I have one old note to myself. We're going to go to this next piece where again I want to summarize our data at the order level. Now in this case the data is already transformed so we don't need to have the "t" parentheses this time.

OK, so now let's see what this data looked like using the same process that we did earlier to create the bar charts, but this time instead of saying what we had in line 40 as Mseq.Order, we're going to have our new object, our Mseq normalized or .normOrder.

All right, adjust the screen so again we have the best possible resolution. I’m going to copy the name of the new taxa plot using command c as a shortcut, and I’m going to paste in using command v as another keyboard shortcut and hit enter.

Again needing to still do a little bit of adjusting, but we have our five sites and you can immediately see how by normalizing, they’re more equally representing as far as the taxa that are present and the abundances. We can look a little bit better into this data, and see what is driving differences between our samples.

Again we have the circles on the bottom, which the size of that circle is going to tell us which ones are more unique than the other samples, and we can look through and try to identify some of the taxa that we would want to follow up more with. So for me I’m definitely going to look more into what this blue taxa is that defines these sites, or is really distinct in those sites, and then this pink taxa for the Inwood Hill and New York Botanical Gardens.

But this is unique to my dataset. For your dataset and as you’re going through these, you might see different patterns, and I would certainly take a note of that on the side. And now at this point you should have plenty of ways that you can dig a little bit more into this data. We’ve looked at the alpha diversity using plot1, so type in plot1, and now we had the taxa that we could look into. So taking these different perspectives of biodiversity and the taxa that are present is our first glance at summarizing and understanding what is going on in these microbiome—what is going on in these microbiome sequences and understand differences between our sites and the projects.