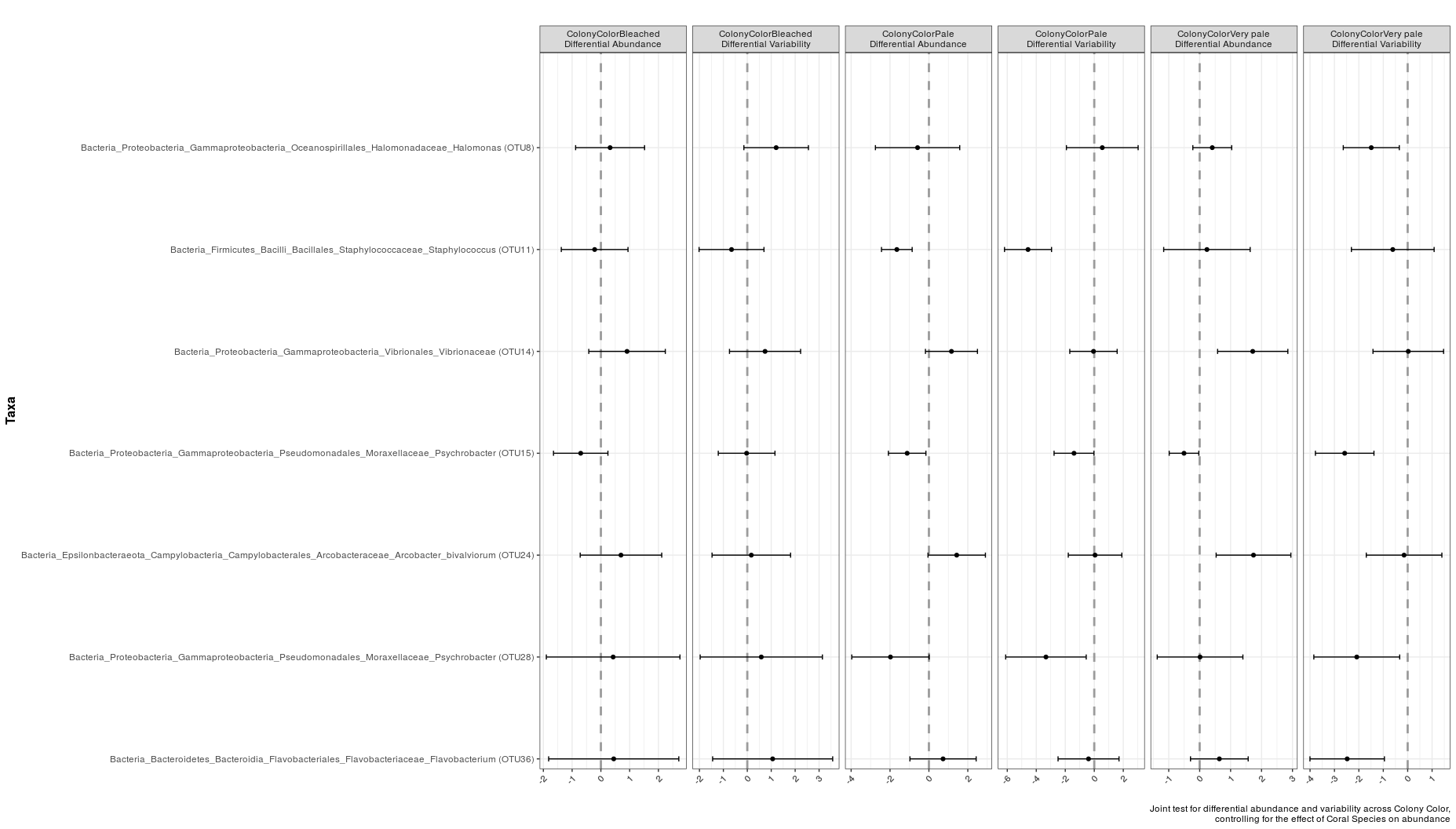
Okay, I think the figures and analyses are ready, at least for a first-pass.

<https://github.com/gzahn/Chagos>

Standard ordinations by island, coral species, temperature, etc are all finished. No great separation in overall community using those methods, as you'll see.

The new method I used was related to finding taxa that are differentially abundant between groups. I looked at this for ColonyColor, TemperatureGroup, and CoralSpecies. There wasn't enough replication between islands to do this by island though.

This image is an example of the analysis (Full\_da\_test.png):



This is showing all of the taxa that have significant differential abundance between corals with different color. It's showing these 7 taxa relative to the baseline "Healthy" corals. These taxa were identified using the corncob R package (**Co**unt **R**egressio**n** for **C**orrelated **O**bservations with the **B**eta-binomial). Here's the pre-print describing the theory <https://arxiv.org/pdf/1902.02776.pdf>

Basically, this method let's us test for differential relative abundance but also for differential *variability*. The idea is that disbiosis (disease) can often manifest itself not so much in just different relative abundance, but also in the *variability* of species abundances. This figure is showing both differential abundance and differential variability. It's a bit cluttered perhaps, so I made several plots just showing differential abundance. All found in the output/figs/ directory of the github site.

The stats tests wind up being the same as a standard glm model. I did a separate test for each taxon that was identified as being differentially abundant, p-values are corrected for false discovery rate. The output below is from one of these tests. All tests are grouped together in what I hope are intuitively named files (like ./output/stats/corncob\_TempGroup\_tables.txt). Did it for ColonyColor, TempGroup, and CoralSpecies.

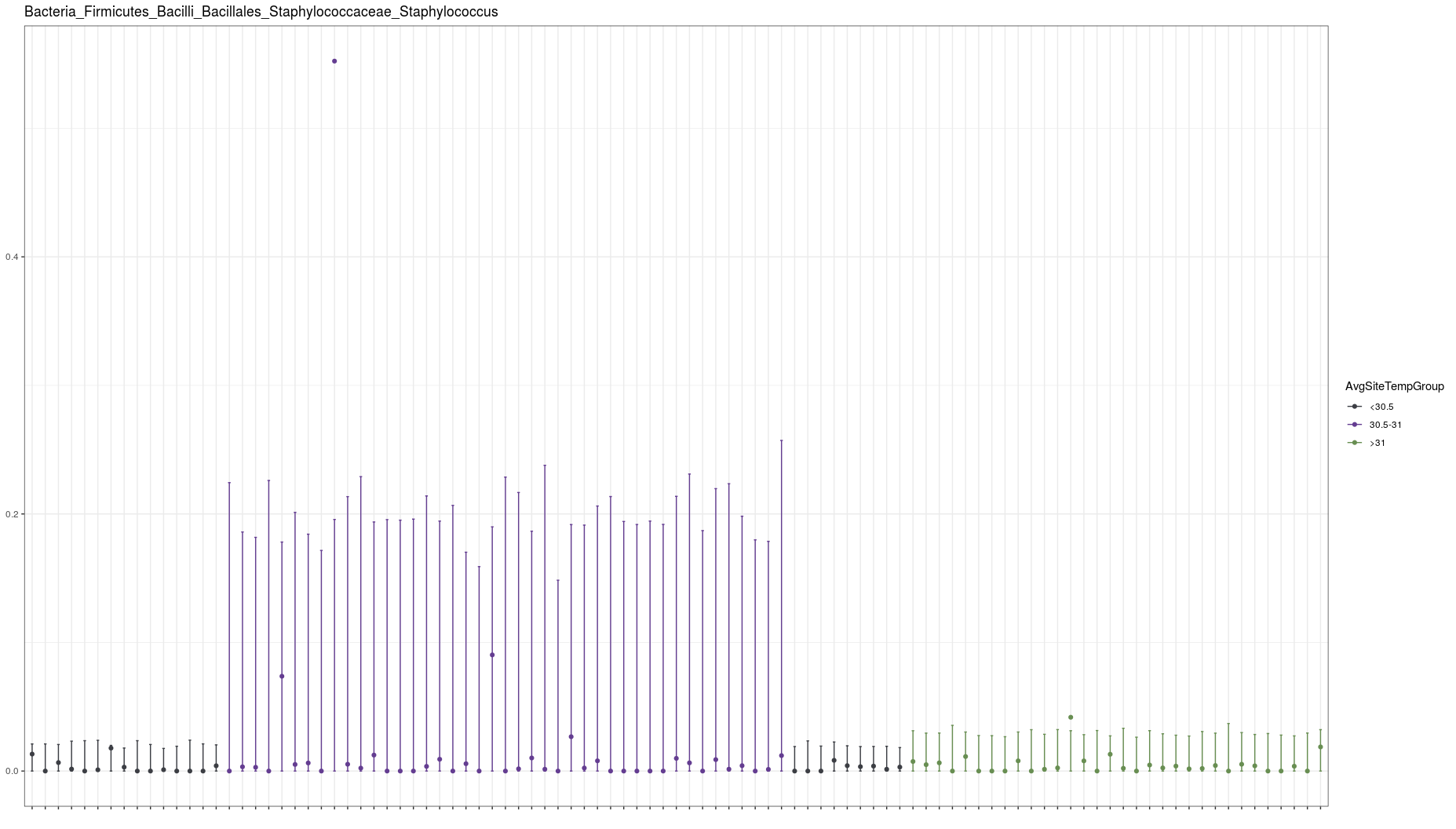
Call:  
bbdml(formula = OTU11 ~ AvgSiteTempGroup, phi.formula = ~AvgSiteTempGroup,   
    data = subset\_samples(ps, AvgSiteTempGroup %in% levels(ps@sam\_data$AvgSiteTempGroup)))  
  
  
Coefficients associated with abundance:  
                        Estimate Std. Error t value Pr(>|t|)      
(Intercept)              -5.8089     0.4213 -13.788  < 2e-16 \*\*\*  
AvgSiteTempGroup30.5-31   2.0928     0.5664   3.695 0.000371 \*\*\*  
AvgSiteTempGroup>31       0.4912     0.5245   0.936 0.351455      
---  
Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
  
Coefficients associated with dispersion:  
                        Estimate Std. Error t value Pr(>|t|)      
(Intercept)              -4.3738     0.5333  -8.202 1.28e-12 \*\*\*  
AvgSiteTempGroup30.5-31   2.5946     0.6795   3.818 0.000242 \*\*\*  
AvgSiteTempGroup>31       0.1734     0.6765   0.256 0.798314      
---  
Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Interpretation:

We get two tables for each OTU/ESV. This one is a test showing just OTU11 and the effect of temperature on its abundance and variability (dispersion). The "Intercept" has been set as what I considered to be the baseline. For colony color, the intercept is "Healthy" and for temperature group, the intercept is <30.5 degrees.

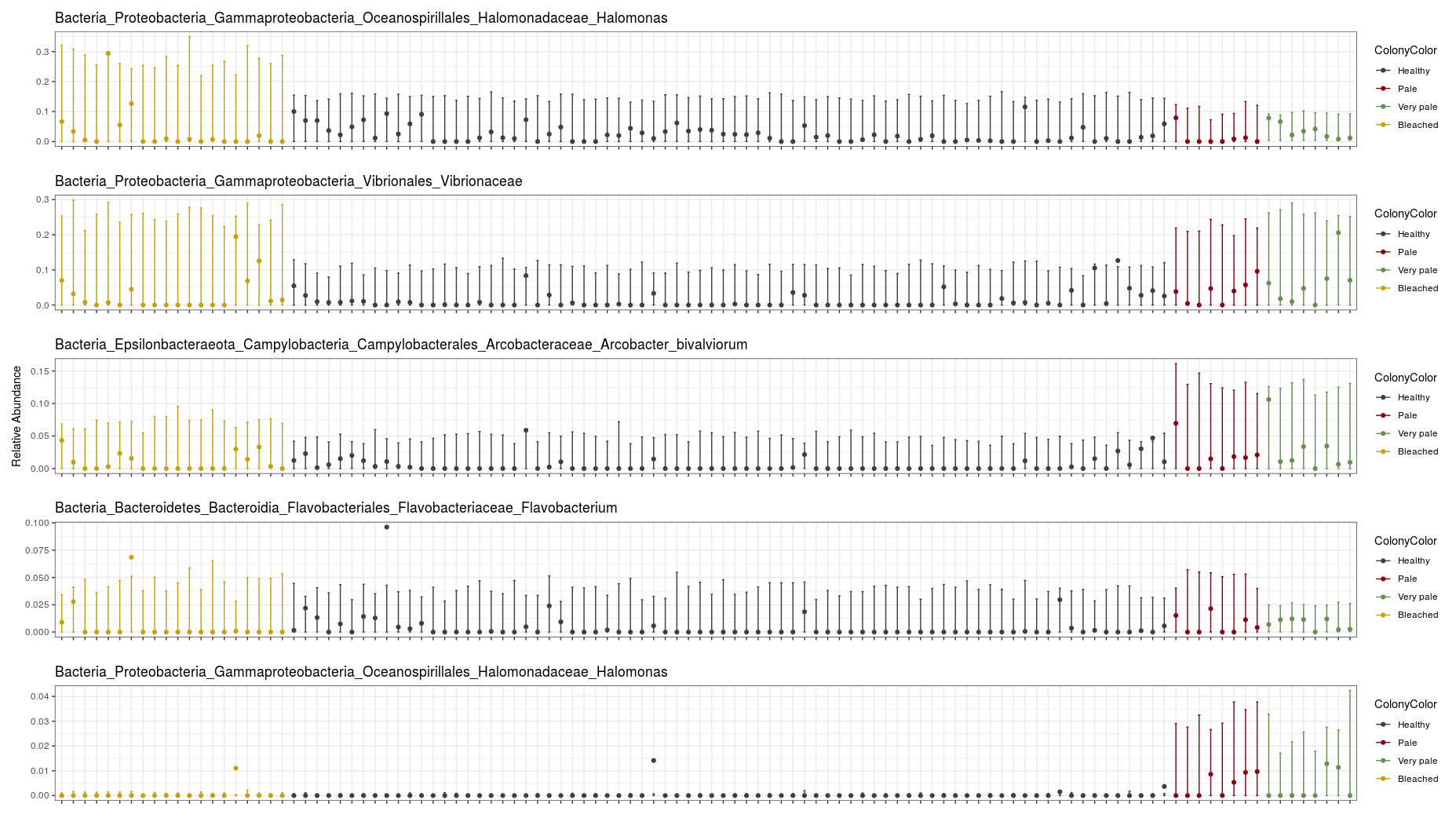
Each of these tables (for each detected OTU) has also been associated with a figure. I stuck all the figures together for each type of test (such as the TempGroup test found in /output/figs/TempGroup\_diff\_abund\_taxa.png)

But here's a slice of that big image:



This is just showing OTU11, from that previous statistical table. This shows what the table confirms: Relative abundance of OTU11 (Staphylococcus) is higher in samples from temperature group 30.5-31 degrees, and dispersion (variability) is higher in that grouping as well. The bars here are 95% CI around the mean.

Stuck all together it looks like this:



This gives some pretty interesting results, in my opinion!

For example, Halomonas (there were two species found, but only one of them follows this pattern...bottom of the previous plot) which has been associated with coral disease in other papers, is increased in pale and very pale corals but NOT in fully bleached corals. Also, it's variability is greater in those two categories. Pretty cool!

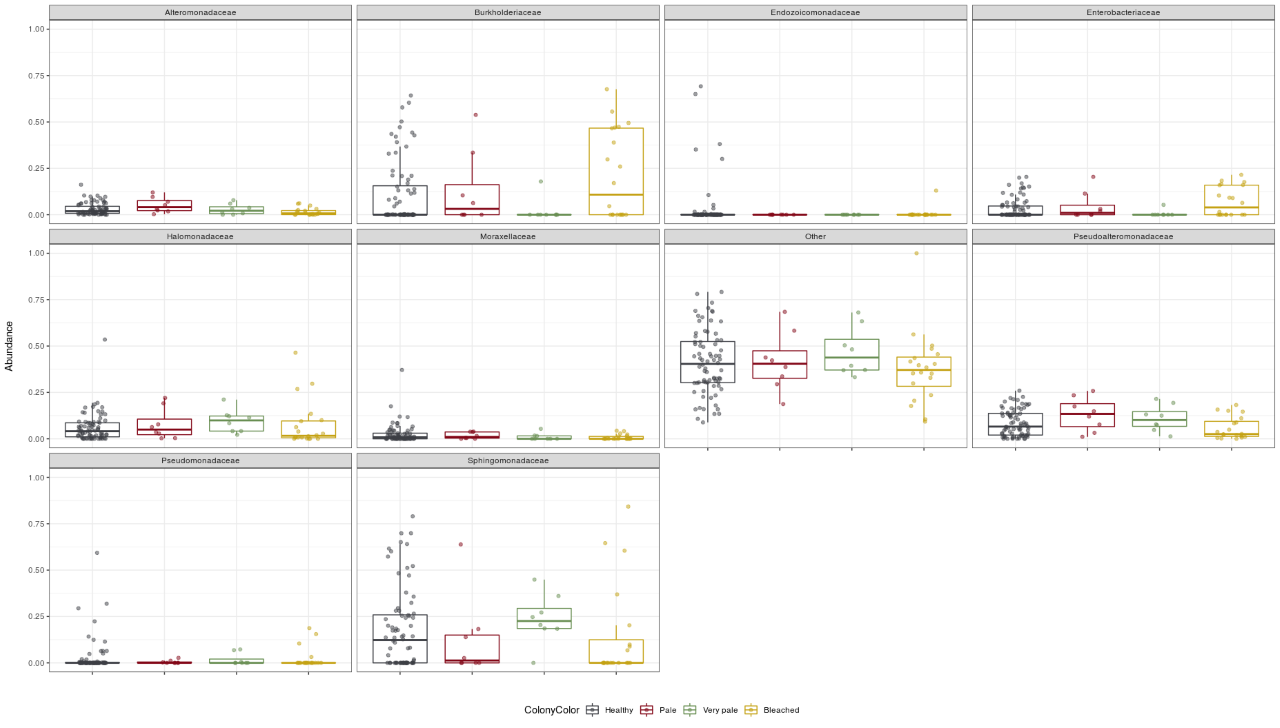
Note that the y-axis for each section of the plot can vary.

I messed about with trying to put all the sites on a map, but it wasn't looking very good. Hopefully you can do a better job. I'm willing to try harder if you prefer.

I also modeled alpha diversity (richness and Shannon diversity) against all the metadata variables. Those results are in output/stats/alpha\_diversity\_models.txt

Finally, I make a cute plot of the most abundant families by colony color:

output/figs/most\_abundant\_families\_by\_colonycolor.png



If you want, I can do this for tempgroup or coral species as well. Or I could do it at a higher taxonomic rank. Don't know if it's useful.

This is probably a lot to digest so I'll stop there. As always, let me know if I'm on the right track or if you'd like alterations to anything. The ordinations aren't very interesting, but they're up on github. I think this corncob analysis really gives a lot of potential meat to the stories we can spin from the data.

Geoff