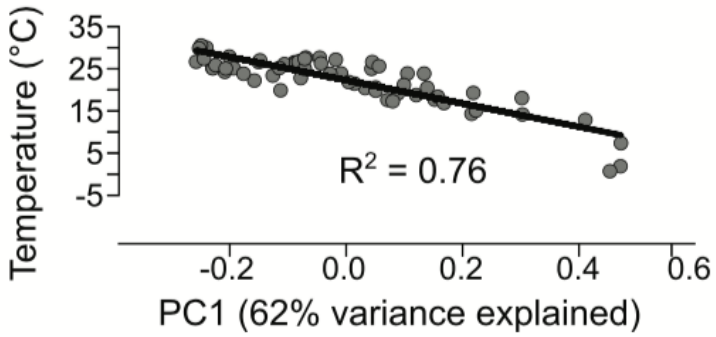
Lab 8 Write-up

OneCard: 1955791
7 Nov 2017

Mini-Research Question

In the paper "Structure and function of the global ocean microbiome," Sunagawa et al. found that overall variability in a community (as measured by principle component analysis) was best explained by temperate. (Sunagawa et al., 2015)



Item 1. The Figure 5a from the Sunagawa et al. paper.

Thus, one question might be**:** *is there is greater species richness and species diversity (as measured by Shannon-Weiner Index) in warmer water samples (15-30 °C) than in colder water samples (0-10 °C)?*

For analysis, six of the eleven available sample regions were chosen, which span almost the full range of temperatures mentioned in the Sunagawa et al. paper (-0.7 °C – 26.54 °C).

Label	Run ID	Region	Sample details	Depth	Temp °C
01_dcm_SernOcean	ERR599104	Southern Ocean (near Antarctica)	deep chlorophyll maximum layer	90 m	-0.78154
02_surface_SernOcean	ERR599090	Southern Ocean (near Antarctica)	surface water layer	5 m	0.67108
03_meso_SPacific	ERR598999	South Pacific (near the Marquesas)	mesopelagic zone	600 m	7.212238
04_surface_NAtlantic	ERR599078	North Atlantic (off the coast of Portugal)	surface water layer	5 m	14.28065
05_dcm_SPacific	ERR598948	South Pacific (near the Marquesas)	deep chlorophyll maximum layer	115 m	24.69625
06_surface_SPacific	ERR598992	South Pacific (near the Marquesas)	surface water layer	5 m	26.54413

Item 2. Metadata for selected samples. Coloring is to indicate temperature range (cold, medium, warm).

Data & Analysis

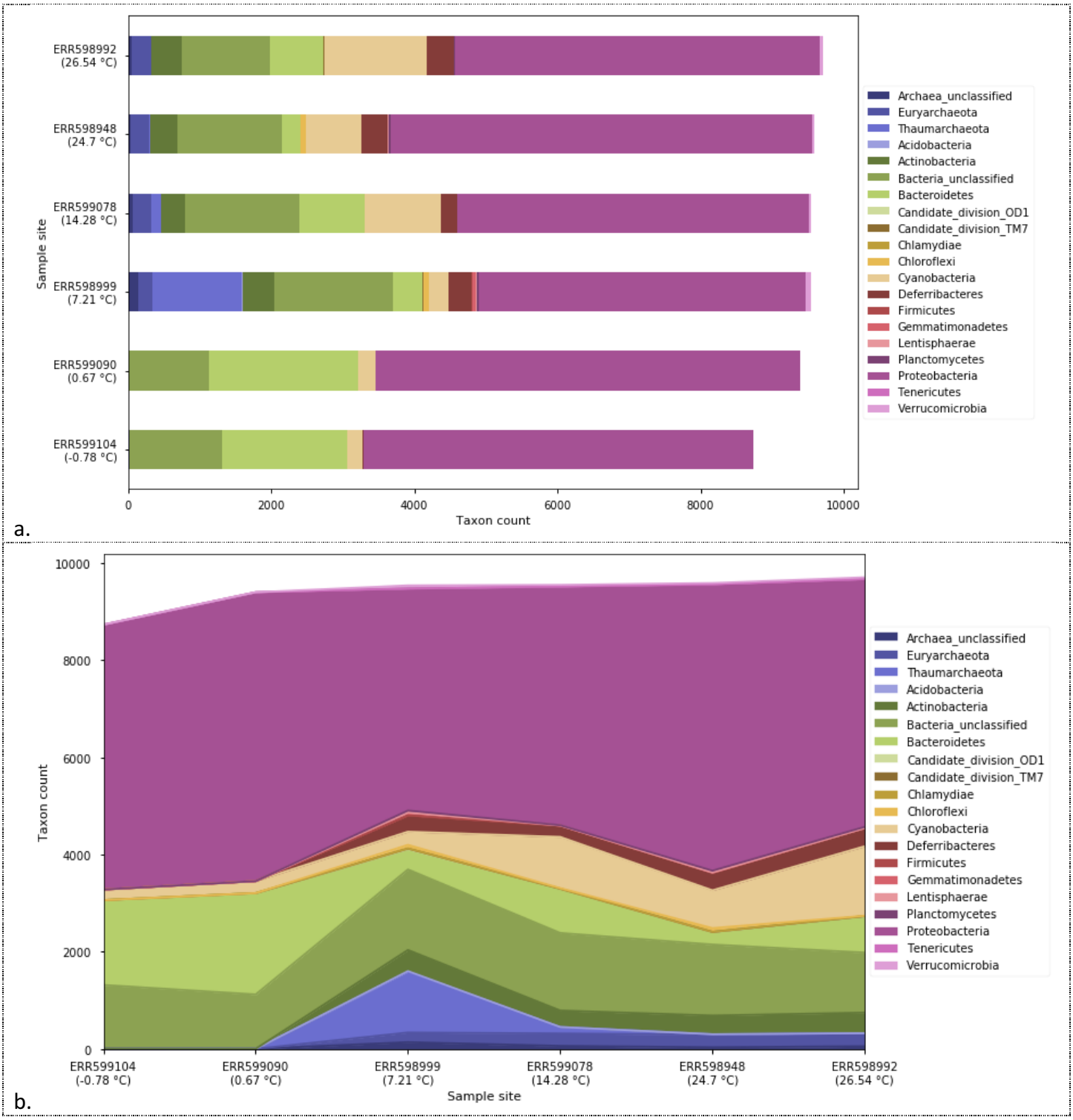
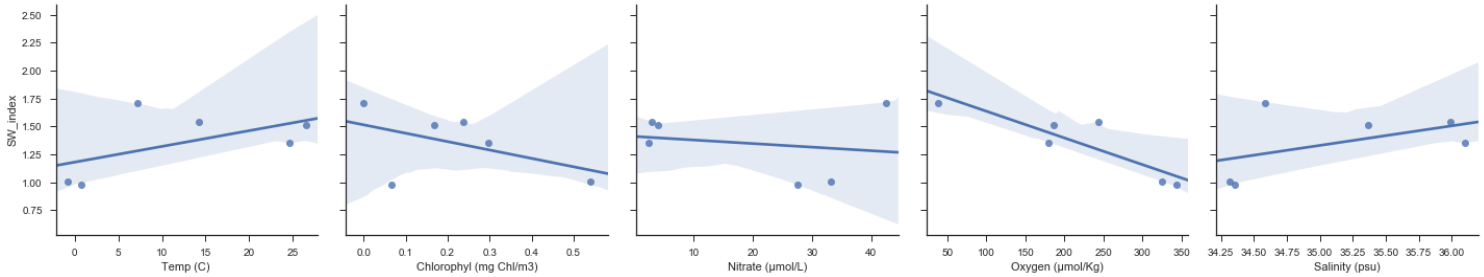


Figure 1. Two different ways-- (a) stacked bar plot, (b) area plot-- of visualizing the taxonomic distribution at taxon level 2, between Tara Ocean Samples, as identified by "Mothur." These plots were lovingly crafted with Python (using Pandas and Matplotlib.) See: https://nbviewer.jupyter.org/github/dustinmichels/biol338-genomics/blob/master/lab-8/analysis/dustin_matplot_charts.ipynb .

Table 1. Summary of diversity statistics calculated for each of the six Tara Ocean samples.

Sample: Label	Sample: Run Id	Total Number Sequences	Species Richness (r)	Shannon-Weiner index (H')
01_dcm_SernOcean	ERR599104	8732	9	1.003610566
02_surface_SernOcean	ERR599090	9399	5	0.974619646
03_meso_SPacific	ERR598999	9539	20	1.714318417
04_surface_NAtlantic	ERR599078	9551	15	1.544702156
05_dcm_SPacific	ERR598948	9588	17	1.349562238
06_surface_SPacific	ERR598992	9706	13	1.513954316

a)



b)

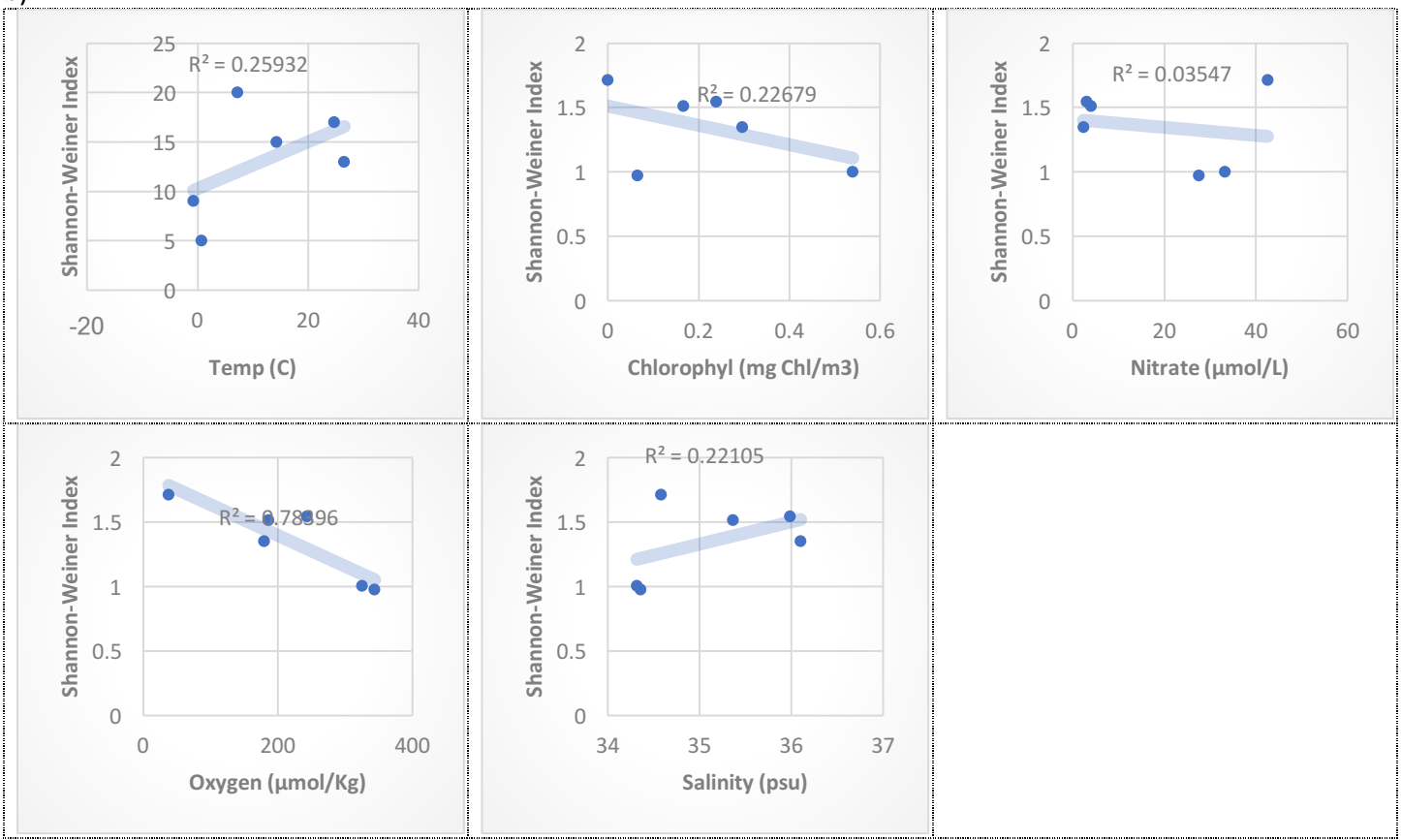


Figure 2. Two slightly different visualizations of the relationship between five metadata attributes pertaining to the Tarra Ocean sample sites (temp, chlorophyll, nitrate, oxygen, and salinity) and the Shannon-Weiner Index of that site, based on taxonomy level 2. (a) I attempted to generating plots programmatically with Python/Jupyter/Matplotlib. The regression lines are drawn by 'Seaborn,' but unfortunately, I could not figure out how to display or output the R2 values, and wound up (b) duplicating this work in Excel. The Jupyter Notebook for the first plots can be viewed here: https://nbviewer.jupyter.org/github/dustinmichels/biol338-genomics/blob/master/lab-8/analysis/dustin_scatterplots.ipynb.

Check for Understanding

1. Many of your sequences were unclassifiable. How would this likely affect your richness calculations for each sample? Explain why.

If we simply ignored unclassifiable samples as I believe we did, this cause us to underestimate species richness.

2. What is the difference between richness and the Shannon-Weiner index? Describe a situation in which you might have a high richness but a relatively low Shannon-Weiner index.

Species richness merely relays the number of different species present in a sample, while Shannon-Weiner index takes into account the abundance and evenness of species present. In a sample where there are many different species but the sample is dominated by one or two species, you could get a high richness but low Shannon-Weiner index.

3. Does your taxonomic diversity, as calculated by the Shannon-Weiner index, correlate with any of the metadata for your sample (temperature, chlorophyll, nitrate, oxygen, salinity)? (The R squared value should vary between 0 and 1; the stronger the correlation, the closer the R-squared value is to 1. We did not calculate p-values or conduct a more rigorous statistical analysis, but the R-squared value will tell you how closely the variables are correlated.) Write a short paragraph speculating on any correlations you find. (It's possible the correlations will be terrible.)

The highest correlations between a metadata attribute and Shannon-Weiner index was for oxygen (R^2 = 0.78). Next best, at roughly the same strength of correlation, are temperature, chlorophyll, and salinity (R^2 = 0.26, 0.23, and 0.22, respectively). These are not particularly strong, but much stronger than the correlation between Shannon-Weiner index and nitrate (0.04).

Mini-Research Conclusion

I wanted to investigate how species richness and diversity (as measured by the Shannon-Weiner index) compared to temperature. Based on the paper the "Structure and function of the global ocean microbiome" by Sunagawa et al., I expected that richness and diversity would both decrease as temperature decreased (from about 30 °C to -5 °C). There does indeed to be a slight positive correlation between temperature/richness and temperature/diversity in that temperature range (Figure 3). However, I get an R^2 value between 2.5 and 3.0 in both cases, whereas Sunagawa et al., reported an R^2 of 0.76 between temperature and "variance." (I'm not entirely sure which factors went into the variable of "variance" in their principal component analysis.) The correlation I found between temperature and Shannon-Weiner index is not noticeably stronger than that between Shannon-Weiner index and salinity or chlorophyll (Figure 2).

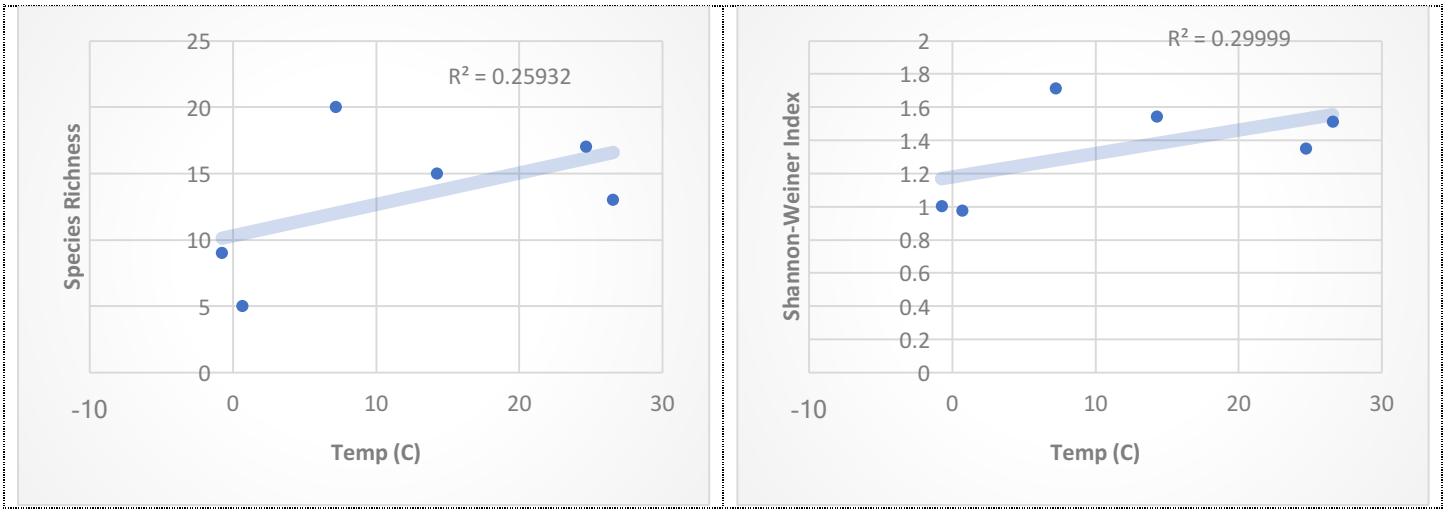


Figure 3. Species richness and Shannon-Weiner index compared to temperature for six Tara Ocean samples.

References

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., et al. (2015). Structure and function of the global ocean microbiome. Science *348*, 1261359.