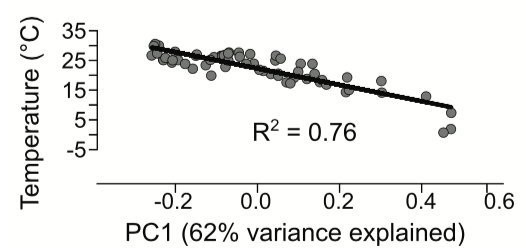
Lab 8 Write-up

OneCard:

# 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

|  |
| --- |
| a.taxonomy_stacked_bar_chart.png |
| b.taxonomy_area_chart.png |

Figure . 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> .

Figure 2

Table 1. A table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

# Check for Understanding

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

Some text.

**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.**

Some text.

**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.)**

Some text.

# Mini-Research Conclusion