

Assignment 1 – Ana Penedos

UID: 1932; student n.: 180981575

Dataset 1 Results Report

In order to investigate the impact of season and latitude on microbial diversity in oceans, seawater samples were collected from 10 separate locations at two times of the year (January and August). Sequences obtained through enrichment and whole genome sequencing (WGS) were identified using BLASTN (Altschul et al., 1990). UniFrac (Lozupone and Knight, 2005) was used as an indicator of microbial diversity, which was expressed in relation to a reference sample.

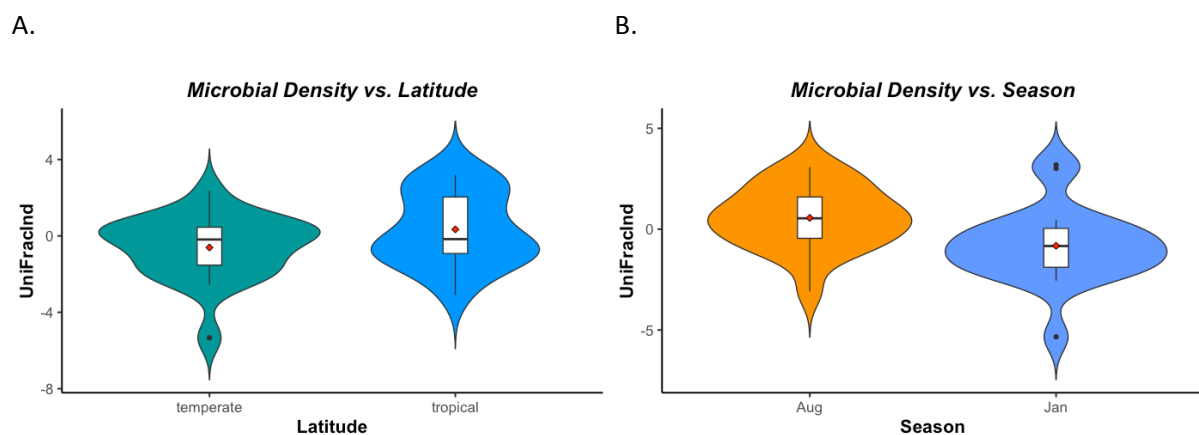


Figure 1. Distribution of observed microbial diversity (UniFracInd) values with latitude (A) and season (B). 20 observations were made for each group. Means of the distributions are indicated by the red losanges (♦).

We find that there is no statistically significant difference in microbial diversity at different latitudes ($t=-1.672$, $df=37$, $P=0.1029$) (figure 1A). A significance level of $\alpha=0.05$ was used for all statistical tests. A previous study reported higher microbial diversity was found in seawater at higher latitudes (Ladau et al., 2013). When microbial diversity is measured relatively to the season at which the samples were collected, we find a statistically significant effect of time of year over diversity ($t=2.5429$, $df=37$, $P=0.01534$). From the observation of the data (figure 1B) we can conclude that there is a higher microbial diversity in August ($M=0.5589$, $SD=1.5519$) than in January ($M=-0.8236$, $SD=1.8717$).

These are not in accord with previously described trends, where oceanic microbial diversity was found to increase in the winter (Ladau et al., 2013), or where no effect of seasonality was found (Suh et al., 2015). We tested whether taking into account the interaction of latitude and season could better explain our results, but found no statistically significant evidence for this ($F=2.3069$, $df=1$, $P=0.13753$).

References

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol*, 215, 403-10.

LADAU, J., SHARPTON, T. J., FINUCANE, M. M., JOSPIN, G., KEMBEL, S. W., O'DWYER, J., KOEPEL, A. F., GREEN, J. L. & POLLARD, K. S. 2013. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J*, 7, 1669-77.

LOZUPONE, C. & KNIGHT, R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*, 71, 8228-35.

SUH, S. S., PARK, M., HWANG, J., KIL, E. J., JUNG, S. W., LEE, S. & LEE, T. K. 2015. Seasonal Dynamics of Marine Microbial Community in the South Sea of Korea. *PLoS One*, 10, e0131633.

Dataset 2 Results Report

30 genes with homology to luciferase were found amongst the 900 coding loci from 208 species of the *Brassicaceae* family sequenced in the context of this RNA seq study. The expression levels of these luciferase orthologues was determined for each species and the pairwise genetic distance between each luciferase homologue and that found in *Arabidopsis thaliana* was calculated. No statistically significant correlation was found between expression levels of the putative luciferase homologues and the genetic distance to the *Arabidopsis thaliana* homologue sequence ($R^2=0.1037$, $F(1, 28)= 3.241$, $P=0.08261$). This may suggest that there is no clear selective pressure over these putative genes.

Dataset 3 Results Report

CNS samples (spinal cord and brain) were collected from an HIV-positive patient over the course of 40 weeks in order to address the impact of CD4-positive T-cell permeability, tissue, viral diversity and evolutionary distance on the viral load (figure 2). We found that the distance from the reference sample (distance) and the viral diversity (shannon) scores' impact on viral load depend on both the tissue and CD4 level ($AIC=33.359$) (equation 1 and table 1).

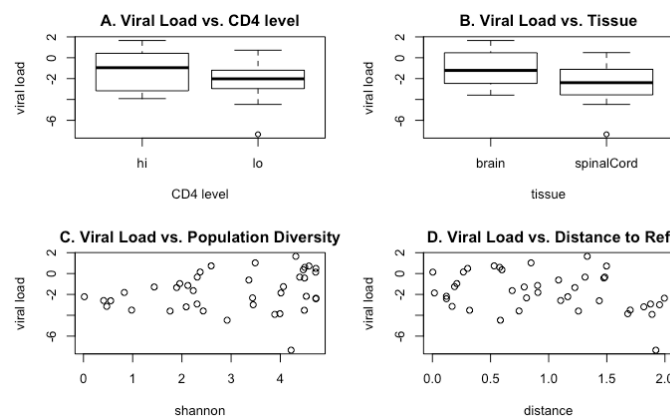


Figure 2. The effect of CD level (A), tissue (B), viral population diversity (C) and evolutionary distance to reference strain (D) on viral load.

$$viral\ load = a + b * distance + c * diversity + errors \quad \text{(equation 1)}$$

Table 1. Coefficients for model (equation 1) expressing viral load in terms of evolutionary distance to the reference strain and viral diversity (Shannon score) depending on tissue and CD4 level. p-values: ** 0.001-0.01, * 0.01-0.05, . 0.05-0.1.

CD4 level	a	b	c
	Spinal cord		
low	0.7471	-1.6148	-0.6946 *
high	-1.2977 .		0.3061 **
	Brain		
low	-1.2579 *	-0.6635	0.0302 **
high	-3.3027		1.0309 **

We find that the viral load is negatively correlated with the distance, perhaps suggesting a fitness loss associated with the evolutionary pressure of replicating within the host. This effect is larger in the spinal cord than in the brain. In contrast, viral load is, in most contexts, positively correlated with the viral diversity observed. This could be a consequence of highly diverse viral populations being more likely to adapt to the host and evade the immune system. This effect is accentuated when CD4 levels are high, possibly due to a higher evolutionary pressure. The only circumstance in which a negative correlation is observed between diversity and viral load is in spinal cord samples with low levels of CD4-positive T-cells (table 1). A correlation between HIV viral load and CD4 levels has previously been described (Srinivasula et al., 2011).

References

SRINIVASULA, S., LEMPICKI, R., ADELSBERGER, J., HUANG, C., ROARK, J., LEE, P., RUPERT, A., STEVENS, R., SERETI, I., LANE, H., DI MASCIIO, M. AND KOVACS, J. (2011). Differential effects of HIV viral load and CD4 count on proliferation of naive and memory CD4 and CD8 T lymphocytes. *Blood*, 118(2), pp.262-270.