Análisis metagenómico de la diversidad de plancton en la zona epipelágica del golfo de México

Grupo Bioinformatica

Informe 2018-08-09

## Generalidad del análisis

Los datos provienen de muestras obtenidas en *42* estaciones que cubrió el crucero exploratorio *XIXIMI-06* Para cada muestra se generaron amplicones de la región v9 del gen ribosomal 18S. Los amplicones fueron secuenciados desde ambos extremos generando secuencias con una longitud de 150 pares de bases (pb) cada uno, en un secuenciador MiSeq de Illumina (del cicese-cigom). Los siguientes resultados corresponden al análisis bioinformático de las secuencias generadas.

El resumen del análisis se divide entre en tres etapas cruciales: *Pre-procesamiento y alineamiento de lecturas.* Identificación, abundancia y clasificación taxonómica de OTUs. \*Estimación de los índices de diversidad y presentación gráfica resultados.

### Pre-procesamiento y alineamiento de lecturas

Durante el pre-procesamiento se demultiplexa y eliminan secuencias ambiguas y de puntaje de calidad phred menor a 20. La siguiente imagen resume las métricas de las bibliotecas illumina obtenidas para el crucero exploratorio (Figura 1 y 2).

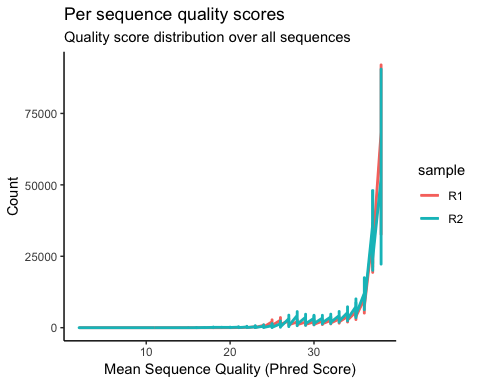


Figura 1. Promedio del puntaje de calidad Phred a lo largo de las lecturas de las bibliotecas. Bibliotecas Forward (R1) y Reverse (R2).

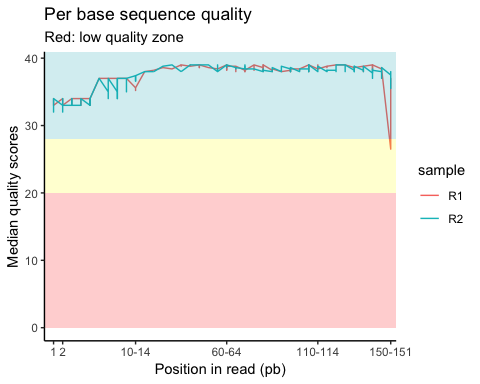


Figura 2. Promedio del puntaje de calidad Phred por posición de nucleótido a lo largo de las lectura. Bibliotecas Forward (R1) y Reverse (R2).

El análisis metagenómico se llevó a cabo en el cluster de cómputo OMICA-cicese mediante el programa mothur version 1.39.5. La tubería del análisis fue adaptada del sugerido por Schloss et al. 2013 para amplicones de la región v9 del gen 18S rRNA secuenciados por Illumina. Después del pre-procesamiento, las lecturas se transformaron a un set de secuencias unicas para implementar el alineamiento con la región v9 de la base de secuencias de referencia Silva v128 (Oliver-Glockner et. al. 2014). Entre aquellas que presentaron homología con la base de referencia mayor al 50% en identidad se implementa un paso de filtrado para eliminar falsos homopolímeros (max=10) y se volvieron a identificar secuencia únicas, posteriormente se quitaron aquellas secuencias que podrían tener al menos 2 cambios en nucleótidos, en base a la matriz generada en el alineamiento (Figura 3).

## Identificación, abundancia y clasificación taxonómica de OTUs

Se removieron quimeras implementando la herramienta VSEARCH (Mahé F. et. al., 2016); las secuencias resultantes fueron agrupadas mediante el algoritmo Opticlust (Schloss et. al., 2017) tomando un parámetro de 97% de similitud entre secuencias para ser consideradas pertenecientes una unidad taxonómica operacional (OTU). Se realizó la asignación taxonómica de las secuencias procesadas usando un algoritmo de clasificación de Bayes ingenuo (Wang, 2007) y se obtuvo la taxonomía consenso para cada OTU. La base de datos para asignación taxonómica está compuesta de las secuencias de la base de datos W2-PR2 (de Vargas, 2015) que fue generada en un estudio metagenómico de amplio muestreo (descargada de su sitio web el 2 de Mayo de 2017) en la que se implementó la nomenclatura utilizada en la base de datos del registro mundial de especies marinas: WORMS (WoRMS Editorial Board, 2017). La figura 3 resume la abundancia de secuencias obtenidas a través del análisis metagenómico :

## Estimación de los índices de diversidad y presentación gráfica resultados

La estimación de los índices de diversidad y la generación de representaciones gráficas de los resultados del análisis metagenómico, se realizó en R v3.4 (R Core Team, 2017) utilizando diferentes paquetes especializados en análisis de datos ecológicos y gráficos.

Para obtener una perspectiva biológica de la las comunidades planctónicas se calculó la abundancia relativa a nivel taxón los OTUs obtenidos por estación. El siguiente grupo de figuras (Figura 4,5,6,7 y 8) muestran en mapas de calor la composición de las comunidades a los diferentes rangos taxonómicos indicados (Filo, Clase, Orden, Familia, Género). Se muestran únicamente aquellos taxones cuya abundancia relativa supera el porcentaje señalado. Para una mejor visualización de los grupos de organismos en la figura correspondiente al rango Género, los organismos se agrupan e indican sus rangos Filos y Clase correspondientes, en las columnas del panel izquierdo (Figura 8).

Aquellos taxones denominados como ‘unclassified’ denotan que no es posible asignar un nombre al rango taxonómico actual y en cambio indican el nombre del último rango asignado posible. Esta tendencia se acentúa conforme se baja de rango taxonómico (sentido Reino a Especie) y depende del grupo de organismos evaluado.

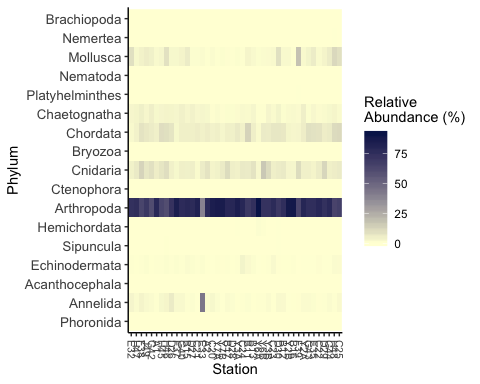


Figura 4. Composición de las comunidades a nivel Filo a lo largo de las estaciones. Abundancia relativa de los taxones que alcanzan más del 1% en la comunidad.

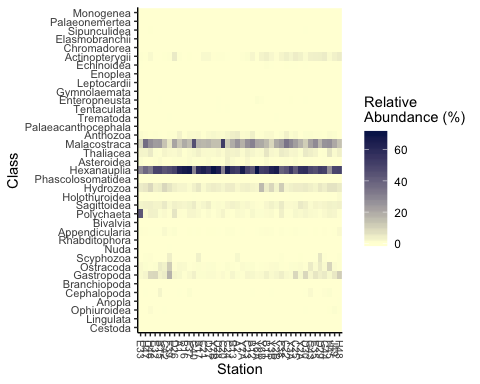


Figura 5. Composición de las comunidades al nivel Clase a lo largo de las estaciones. Abundancia relativa de los taxones que alcanzan más del 2 % en la comunidad.

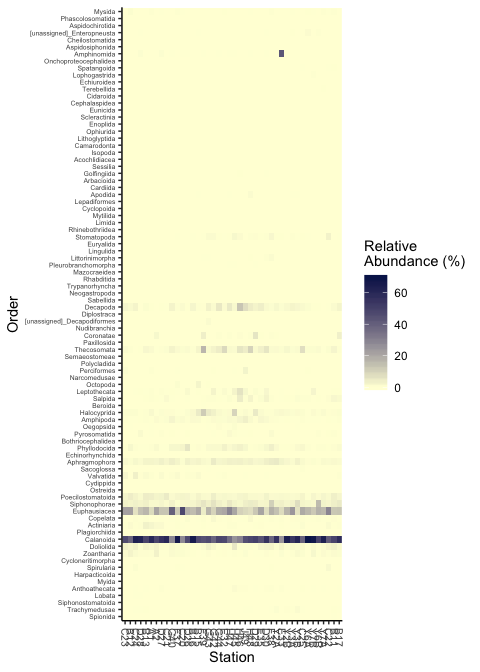


Figura 6. Composición de las comunidades a nivel Orden a lo largo de las estaciones. Abundancia relativa de los taxones que alcanzan más del 5 % en la comunidad.

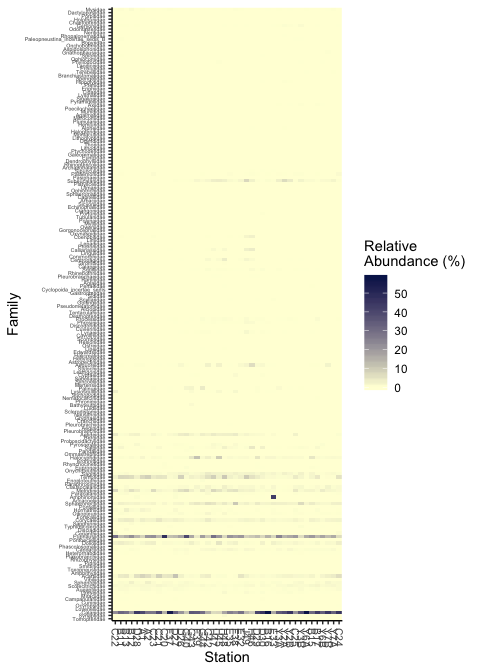


Figura 7. Composición de las comunidades a nivel Familia a lo largo de las estaciones. Abundancia relativa de los taxones que alcanzan más del 5 % en la comunidad

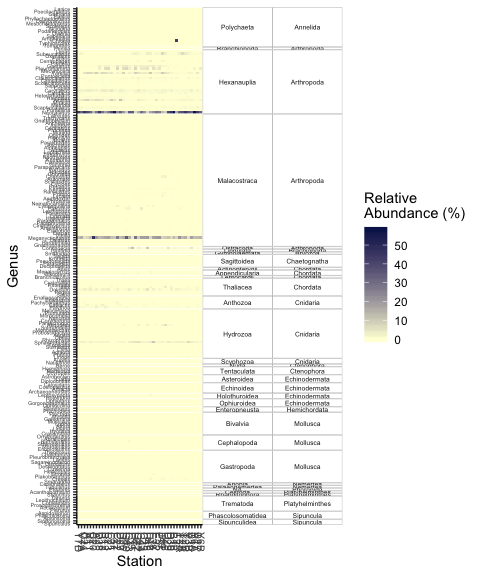


Figura 8. Composición de las comunidades a nivel Género a lo largo de las estaciones. Abundancia relativa de los taxones que alcanzan más del 5% en la comunidad. Respectivamente, se indican sus rangos Filos y Clase correspondientes en las columnas del panel izquierdo

## rarefying sample 10\_X06\_A1\_18S\_AMB  
## rarefying sample 10\_X06\_A2\_18S\_AMB  
## rarefying sample 10\_X06\_B11\_18S\_AMB  
## rarefying sample 10\_X06\_B12\_18S\_AMB  
## rarefying sample 10\_X06\_B13\_18S\_AMB  
## rarefying sample 10\_X06\_B15\_18S\_AMB  
## rarefying sample 10\_X06\_B16\_18S\_AMB  
## rarefying sample 10\_X06\_B17\_18S\_AMB  
## rarefying sample 10\_X06\_C20\_18S\_AMB  
## rarefying sample 10\_X06\_C21\_18S\_AMB  
## rarefying sample 10\_X06\_C22\_18S\_AMB  
## rarefying sample 10\_X06\_C23\_18S\_AMB  
## rarefying sample 10\_X06\_C24\_18S\_AMB  
## rarefying sample 10\_X06\_C25\_18S\_AMB  
## rarefying sample 10\_X06\_D26\_18S\_AMB  
## rarefying sample 10\_X06\_D27\_18S\_AMB  
## rarefying sample 10\_X06\_D28\_18S\_AMB  
## rarefying sample 10\_X06\_D29\_18S\_AMB  
## rarefying sample 10\_X06\_D30\_18S\_AMB  
## rarefying sample 10\_X06\_E32\_18S\_AMB  
## rarefying sample 10\_X06\_E33\_18S\_AMB  
## rarefying sample 10\_X06\_E35\_18S\_AMB  
## rarefying sample 10\_X06\_F37\_18S\_AMB  
## rarefying sample 10\_X06\_F38\_18S\_AMB  
## rarefying sample 10\_X06\_F39\_18S\_AMB  
## rarefying sample 10\_X06\_G40\_18S\_AMB  
## rarefying sample 10\_X06\_G42\_18S\_AMB  
## rarefying sample 10\_X06\_G43\_18S\_AMB  
## rarefying sample 10\_X06\_G44\_18S\_AMB  
## rarefying sample 10\_X06\_H45\_18S\_AMB  
## rarefying sample 10\_X06\_H46\_18S\_AMB  
## rarefying sample 10\_X06\_H47\_18S\_AMB  
## rarefying sample 10\_X06\_H48\_18S\_AMB  
## rarefying sample 10\_X06\_J49\_18S\_AMB  
## rarefying sample 10\_X06\_Y2A\_18S\_AMB  
## rarefying sample 10\_X06\_Y2B\_18S\_AMB  
## rarefying sample 10\_X06\_Y3A\_18S\_AMB  
## rarefying sample 10\_X06\_Y3B\_18S\_AMB  
## rarefying sample 10\_X06\_Y6A\_18S\_AMB  
## rarefying sample 10\_X06\_Y6B\_18S\_AMB  
## rarefying sample 10\_X06\_Y7A\_18S\_AMB  
## rarefying sample 10\_X06\_Y7B\_18S\_AMB

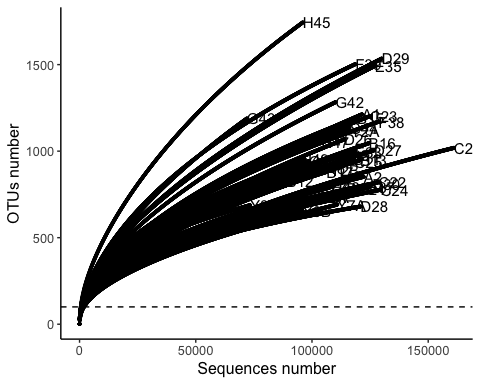
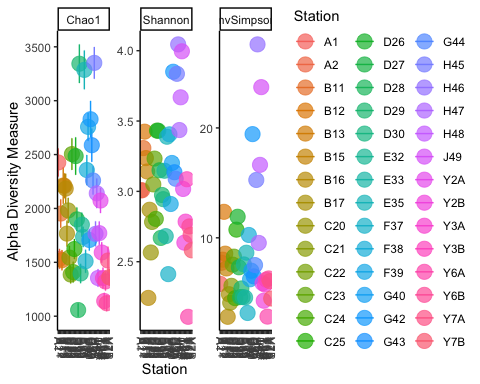


Figura 9. Curva de rarefacción que presenta la riqueza de OTUs a lo largo de las estaciones en relación al número de secuencias

En el siguiente enlace externo se encuentran la tabla de los índices de diversidad alfa: PATHTOTHEFILE

## Warning: Removed 84 rows containing missing values (geom\_errorbar).



Diversidad alpha

La diferencia en composición de la comunidad de zooplancton entre estaciones se evaluó mediante un análisis de coordenadas principales usando una matriz de distancias Jaccard. En la figura 10 se observa que la mayoría de las estaciones forman agrupaciones indicando comunidades similares mientras que son pocas estaciones que se separan, indicando comunidades diferentes al resto del grupo.

## Warning: Ignoring unknown aesthetics: label.size

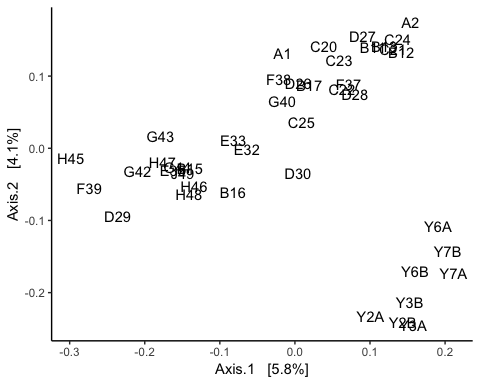
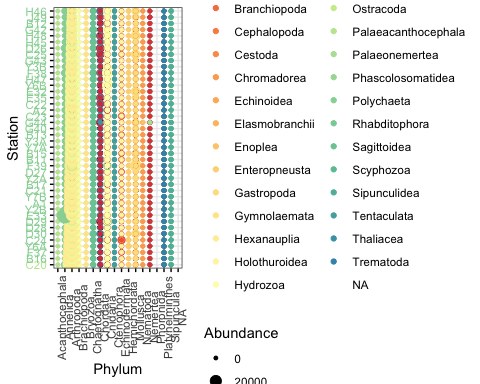


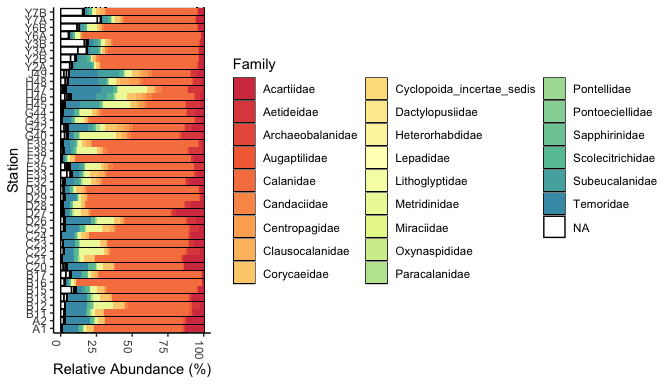
Figura 10. PCoA. Componentes principales (Distancias Jaccard).

## Warning: Removed 230454 rows containing missing values (geom\_point).



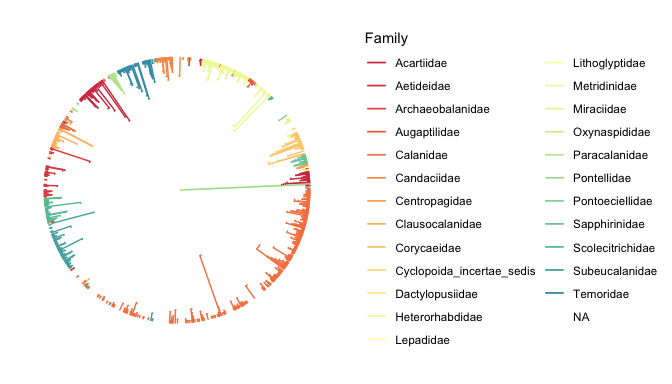
Additional plots - otus (abundantes) en una/ dos estaciones unicamente

Visualizacion de hexanauplia



Hexanauplia subset through the samples (Family label in the xlab)

## Warning: package 'dplyr' was built under R version 3.5.1



Hexanauplia tree (Family labeled)

## Citas

1 Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. (2013): Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology. 79(17):5112-20

2 Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. (2014): The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucl. Acids Res. 42:D643-D648

3 Rognes T, Flouri T, Nichols B, Quince C, Mahé F. (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584

4 Westcott SL, Schloss PD. (2017). OptiClust, an improved method for assigning amplicon-based sequence data to operational taxonomic units. mSphere 2:e00073-17

5 De Vargas, C., et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348.6237 1261605