

**TALLER ÍNDICES DE DIVERSIDAD 2021\_1**

CURSO ECOLOGÍA MICROBIANA Y HERRAMIENTAS DE ANÁLISIS BIOINFORMÁTICO

PROFESOR: Diego Javier Jiménez PhD.

**Carlos Andrés Diaz Rodríguez , David León, César Patiño**

**Objetivos del taller:**

* Explicar como la profundidad de muestreo (e.g. secuenciación) influye en la evaluación de la diversidad microbiana
* Describir como la evaluación de la diversidad microbiana es dependiente de las probabilidades
* Calcular índices de diversidad y realizar curvas de rarefacción de las comunidades microbianas muestreadas

**Caso:** Usted es investigador del *Smithsonian Tropical Research Institute.* Uno de los objetivos de su grupo de investigación es evaluarla diversidad microbiana asociada a bosques de manglares. En el 2020, su grupo de investigacion realizó muestreos de una especie de manglar (*Rhizophora mangle*) en el caribe Colombiano. Posteriormente, se llevo a cabo la extracción de ADN total y se amplificó el gen ribosomal 16S de la comunidad bacteriana. Estos amplicones fueron enviados a la compañía GenWiz (en USA) y usted obtuvo resultados de secuenciación de tres tipos de microbiomas: 1. Sedimentos,; 2. Agua de Mar y 3. Filosfera. Después del proceso de secuenciación, usted recibió los datos que fueron procesados en el software qiime2. Con base en las siguientes instrucciones, responda los preguntas planteadas a continuación:

**Instrucciones:**

1. Cada grupo tiene un archivo en .pwp, en el cual hay unas cajas que contiene una comunidad de circulos de colores. Cada especie/taxa u OTU (97%) esta representado por un color.
2. Cuente el numero de circulos por cada caja, y cuente el numero de colores únicos (su abundancia).

**Agua de Mar**

|  |  |
| --- | --- |
| **Muestra 1 (200 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 1 |
| OTU Rojo | 4 |

|  |  |
| --- | --- |
| **Muestra 2 (2000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 2 |
| OTU Rojo | 5 |
| OTU Amarillo | 3 |

|  |  |
| --- | --- |
| **Muestra 3 (5000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 3 |
| OTU Rojo | 7 |
| OTU Amarillo | 3 |
| OTU Verde | 2 |

|  |  |
| --- | --- |
| **Muestra 4 (7000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 2 |
| OTU Rojo | 7 |
| OTU Amarillo | 3 |
| OTU Verde | 3 |
| OTU Negro | 3 |

|  |  |
| --- | --- |
| **Muestra 5 (10000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 3 |
| OTU Rojo | 10 |
| OTU Amarillo | 4 |
| OTU Verde | 3 |
| OTU Negro | 3 |
| OTU Rosa | 7 |

|  |  |
| --- | --- |
| **Muestra 6 (12000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 7 |
| OTU Rojo | 18 |
| OTU Amarillo | 5 |
| OTU Verde | 3 |
| OTU Negro | 4 |
| OTU Rosa | 3 |

**Filosfera**

|  |  |
| --- | --- |
| **Muestra 1 (100 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 1 |
| OTU Amarillo | 1 |
| OTU Rojo | 3 |

|  |  |
| --- | --- |
| **Muestra 2 (1000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 2 |
| OTU Amarillo | 3 |
| OTU Rojo | 4 |
| OTU Verde | 1 |

|  |  |
| --- | --- |
| **Muestra 3 (2000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 3 |
| OTU Amarillo | 3 |
| OTU Rojo | 5 |
| OTU Verde | 2 |
| OTU Negro | 2 |

|  |  |
| --- | --- |
| **Muestra 4 (10000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 2 |
| OTU Amarillo | 3 |
| OTU Rojo | 7 |
| OTU Verde | 3 |
| OTU Negro | 3 |

|  |  |
| --- | --- |
| **Muestra 5 (15000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 3 |
| OTU Amarillo | 4 |
| OTU Rojo | 10 |
| OTU Verde | 3 |
| OTU Negro | 3 |
| OTU Rosa | 5 |
| OTU Verde Pastel | 2 |

|  |  |
| --- | --- |
| **Muestra 6 (20000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 7 |
| OTU Amarillo | 5 |
| OTU Rojo | 16 |
| OTU Verde | 3 |
| OTU Negro | 4 |
| OTU Rosa | 3 |
| OTU Verde Pastel | 2 |

**Sedimentos**

|  |  |
| --- | --- |
| **Muestra 1 (200 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 4 |
| OTU Rojo | 1 |

|  |  |
| --- | --- |
| **Muestra 2 (2000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 4 |
| OTU Rojo | 5 |
| OTU Amarillo | 1 |

|  |  |
| --- | --- |
| **Muestra 3 (5000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 8 |
| OTU Rojo | 3 |
| OTU Amarillo | 3 |
| OTU Verde | 1 |

|  |  |
| --- | --- |
| **Muestra 4 (7000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 8 |
| OTU Rojo | 6 |
| OTU Amarillo | 3 |
| OTU Verde | 1 |
| OTU Negro | 2 |

|  |  |
| --- | --- |
| **Muestra 5 (10000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 11 |
| OTU Rojo | 7 |
| OTU Amarillo | 4 |
| OTU Verde | 3 |
| OTU Negro | 3 |
| OTU Rosa | 2 |

|  |  |
| --- | --- |
| **Muestra 6 (15000 seq)** | |
| **OTU** | **Cantidad** |
| OTU Azul | 23 |
| OTU Rojo | 7 |
| OTU Amarillo | 4 |
| OTU Verde | 2 |
| OTU Negro | 2 |
| OTU Rosa | 2 |

1. Grafique el numero de colores únicos vs el numero de circulos muestreadas con el fin de hacer una curva de rarefacción. Para eso grafique #OTUs vs # de secuencias

**Agua de Mar**

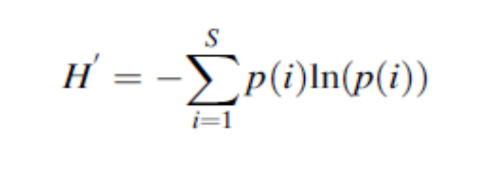
|  |  |
| --- | --- |
| **Rarefacción** | |
| **OTUs** | **Seq** |
| 5 | 200 |
| 10 | 2000 |
| 15 | 5000 |
| 18 | 7000 |
| 30 | 10000 |
| 40 | 12000 |

**Filosfera**

|  |  |
| --- | --- |
| **Rarefacción** | |
| **OTUs** | **Seq** |
| 5 | 100 |
| 10 | 1000 |
| 15 | 2000 |
| 18 | 10000 |
| 30 | 15000 |
| 40 | 20000 |

**Sedimentos**

|  |  |
| --- | --- |
| **Rarefacción** | |
| **OTUs** | **Seq** |
| 5 | 200 |
| 10 | 2000 |
| 15 | 5000 |
| 20 | 7000 |
| 30 | 10000 |
| 40 | 15000 |

1. Con los datos del mayor numero de secuencias, realice un grafico de abundancia de cada OTU, estime la riqueza y el indice de Shannon con la siguiente formula:

En donde H es el índice de Shannon

**Agua de Mar, cuyo H es 1,543076869**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Especie** | **Abundancia** | **p(i)** | **ln(p(i))** | **p(i) x ln(p(i))** |
| **OTU Azul** | 7 | 0,175 | -1,74297 | -0,305019628 |
| **OTU Rojo** | 18 | 0,45 | -0,79851 | -0,359328463 |
| **OTU Amarillo** | 5 | 0,125 | -2,07944 | -0,259930193 |
| **OTU Verde** | 3 | 0,075 | -2,59027 | -0,194270037 |
| **OTU Negro** | 4 | 0,1 | -2,30259 | -0,230258509 |
| **OTU Rosa** | 3 | 0,075 | -2,59027 | -0,194270037 |
| **N** | 40 |  |  | -1,543076869 |

**Filosfera, cuyo H es 1,700051312**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Especie** | **Abundancia** | **p(i)** | **ln(p(i))** | **p(i) x ln(p(i))** |
| **OTU Azul** | 7 | 0,175 | -1,742969305 | -0,305019628 |
| **OTU Amarillo** | 5 | 0,125 | -2,079441542 | -0,259930193 |
| **OTU Rojo** | 16 | 0,4 | -0,916290732 | -0,366516293 |
| **OTU Verde** | 3 | 0,075 | -2,590267165 | -0,194270037 |
| **OTU Negro** | 4 | 0,1 | -2,302585093 | -0,230258509 |
| **OTU Rosa** | 3 | 0,075 | -2,590267165 | -0,194270037 |
| **OTU Verde Pastel** | 2 | 0,05 | -2,995732274 | -0,149786614 |
| **N** | 40 |  |  | -1,700051312 |

**Sedimentos, cuyo H es 1,302834491**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Especie** | **Abundancia** | **p(i)** | **ln(p(i))** | **p(i) x ln(p(i))** |
| **OTU Azul** | 23 | 0,575 | -0,553385238 | -0,318196512 |
| **OTU Rojo** | 7 | 0,175 | -1,742969305 | -0,305019628 |
| **OTU Amarillo** | 4 | 0,1 | -2,302585093 | -0,230258509 |
| **OTU Verde** | 2 | 0,05 | -2,995732274 | -0,149786614 |
| **OTU Negro** | 2 | 0,05 | -2,995732274 | -0,149786614 |
| **OTU Rosa** | 2 | 0,05 | -2,995732274 | -0,149786614 |
| **N** | 40 |  |  | -1,302834491 |

**Preguntas:**

1. Con las gráficas de rarefacción obtenidas, ¿qué puede usted concluir de la profundidad de secuenciación para cada microbioma?

En cada microbioma se observa que, a medida que aumenta la cantidad de lecturas también aumenta la cantidad de OTUs que se encuentran, por lo que se observa crecimiento constante en la diversidad; sin embargo, la profundidad de secuenciación resulta insuficiente para alcanzar a describir la riqueza de las especies en términos de OTUs, esto se debe a que como se observa en la gráfica, no hay un punto en el que la cantidad de de OTUs descritas, se estabilice.

1. ¿Que microbioma tiene una mayor riqueza (e.g. # OTUs) y diversidad? ¿Hay alguna especie clave en cada microbioma? ¿Qué comunidades tienen una estructura similar?

El Microbioma que tiene mayor riqueza es el de la Filosfera debido a que tiene un OTU perteneciente al color verde pastel, con una abundancia relativa de 2; además posee el indice de Shannon más alto entre los diferentes microbiomas: 1,700051312.

En el microbioma perteneciente a agua de mar la especie clave es el OTU de color rojo, pues es la que se encuentra en mayor proporción.

En el microbioma perteneciente a la filosfera la especie clave es el OTU de color rojo, pues es la que se encuentra en mayor proporción.

En el microbioma perteneciente a los sedimentos la especie clave es el OTU de color azul, pues es la que se encuentra en mayor proporción.

Las comunidades que tienen una estructura similar son la comunidad perteneciente al agua de mar y la comunidad perteneciente a los sedimentos, pues presentan la misma diversidad en términos de OTUs, sin embargo, vale aclarar que se presentan en diferente proporción ya que la especie predominante en el primero es el OTU rojo, y en el segundo es el OTU azul.

1. ¿Como podría evaluar beta diversidad con estos valores de abundancia relativa?

Para evaluar la beta diversidad usaríamos los valores de abundancia relativa para comparar las diferentes especies que se encuentran en los distintos ambientes (agua de mar, filosfera y sedimentos). Al análizar estos datos nos damos cuenta que en el microbioma de la filosfera se encuentra una especie que no se encuentra en los otros dos microbiomas, por lo que la clasificaríamos como una especie local, a diferencia de las otras especies que se encuentran presentes en los tres microbiomas.

1. ¿Es posible comparar los valores de Shannon con los datos obtenidos? Indique una desventaja de este indice

Sí es posible, pues el valor del índice es representativo de la diversidad que se observa en los microbiomas, por ejemplo, el índice tiene un valor más alto en el microbioma que representa la filosfera y dicha diversidad la podemos observar al análizar la cantidad de OTUs distintos presentes en el microbioma. Por otra parte, en el microbioma que representa los sedimentos, el índice es el mas bajo y esto se debe a que tiene una especie menos que la filosfera y a que más de la mitad de las especies pertenecen a un OTU en partícular, cosa que no pasa en los otros microbiomas.

Una desventaja que presenta en índice de Shannon es que no tiene en cuenta la taxonomía ,este índice refleja la heterogeneidad de una comunidad teniendo encuenta dos factores: el número de especies presentes y su abundancia relativa.

Conceptualmente es una medida del grado de incertidumbre asociada a la selección aleatoria de un individuo en la comunidad, lo cual puede constituirse en esa desventaja, ante los otros métodos por la representatividad de las especies y los individuos; esto es, si una comunidad de S especies es muy homogénea, por ejemplo porque existe una especie claramente dominante y las restantes S-1 especies apenas presentes, el grado de incertidumbre será más bajo que si todas las S especies fueran igualmente abundantes.

Pla, L. (2021). Biodiversidad: Inferencia basada en el índice de Shannon y la riqueza. Retrieved 27 February 2021, from http://ve.scielo.org/scielo.php?script=sci\_arttext&pid=S0378-18442006000800008

1. Concluya y de su punto de vista acerca de estas comunidades microbianas en cada uno de los microbiomas muestreados

En general, se observa que a medida que aumenta la cantidad de secuencias, también aumenta la abundancia de los OTUs presentes, como se puede ver en las gráficas de rarefacción; sin embargo, estas secuencias no son suficientes para identificar la diversidad total de las especies pues no hay un punto en el que las gráficas se estabilicen. Por otra parte, es importante resaltar que todas las especies presentes en el microbioma de agua de mar y sedimentos están presentes en el microbioma de la filosfera.

Con respecto a la comunidad microbiana perteneciente al agua de mar, se observa un valor de 1,543076869 para su índice de Shannon, lo cual significa que tiene una diversidad media al compararla con las otras dos comunidades; esto se debe a que la especie representada por el OTU rojo se presenta en mayores cantidades, sin embargo, esta especie no alcanza a representar el 50% de la abundancia en el microbioma muestreado.

Con respecto a la comunidad microbiana perteneciente a la filosfera, se observa un valor de 1,700051312 para su índice de Shannon, lo cual significa que tiene una diversidad alta al compararla con las otras dos comunidades; esto se debe a que hay una especie que no se encuentra en ninguna de las otras comunidades; además, al análizar la abundancia de las especies presentes, se observa que ninguno de los OTUs supera el 40% de la abundancia del microbioma muestreado.

Con respecto a la comunidad microbiana perteneciente a los sedimentos, se observa un valor de: 1,302834491 para su índice de Shannon, lo cual significa que tiene una diversidad baja al compararla con las otras dos comunidades; esto se debe a que hay una especie que constituye más del 50% de la abundancia del microbioma muestreado, la especie representada por el OTU azul.

**Qiime 2**

**Nota: Desde este momento su tabla inical será la** TreeFiltered\_Table.qza **del taller pasado**

*Rarefacción*

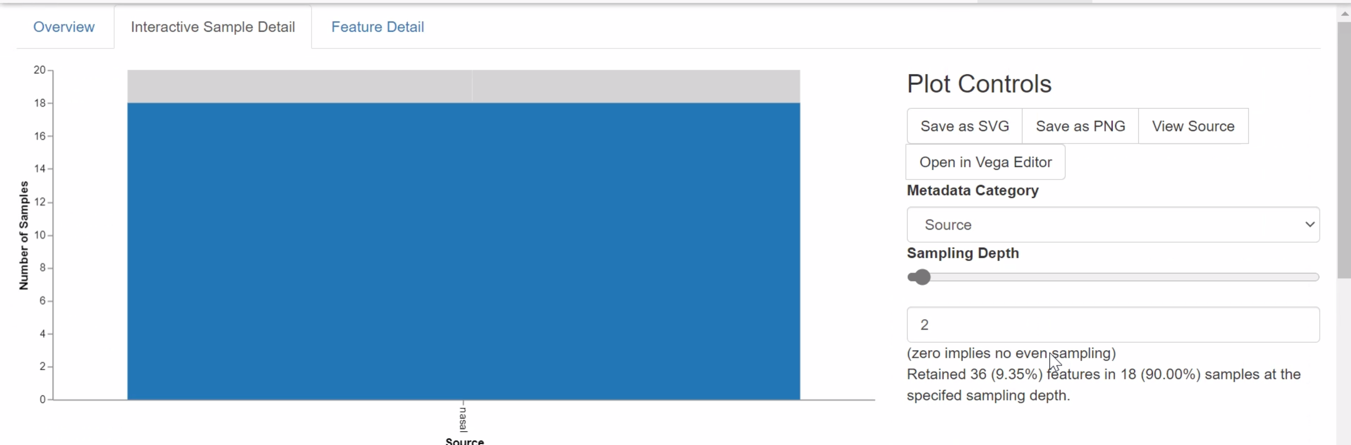
En este paso se ajustarán las diferencias en el tamaño de las librerías, esto con el fin de que no haya sobreestimaciones de diversidad debido a un mayor número de secuencias en determinadas muestras.

1. Vuelva a visualizar el archivo TreeFiltered\_Table.qzv y defina un número de reads que se usará para rarificar. **Pista:** Encuentre un balance en donde pierda el menor número de muestras y donde conserve un buen número de reads por muestra para los análisis subsecuentes.

En este caso se va usar el método “ alpha-rarefaction” del plugin “Diversity”. Busque la forma de correrlo en la documentación.

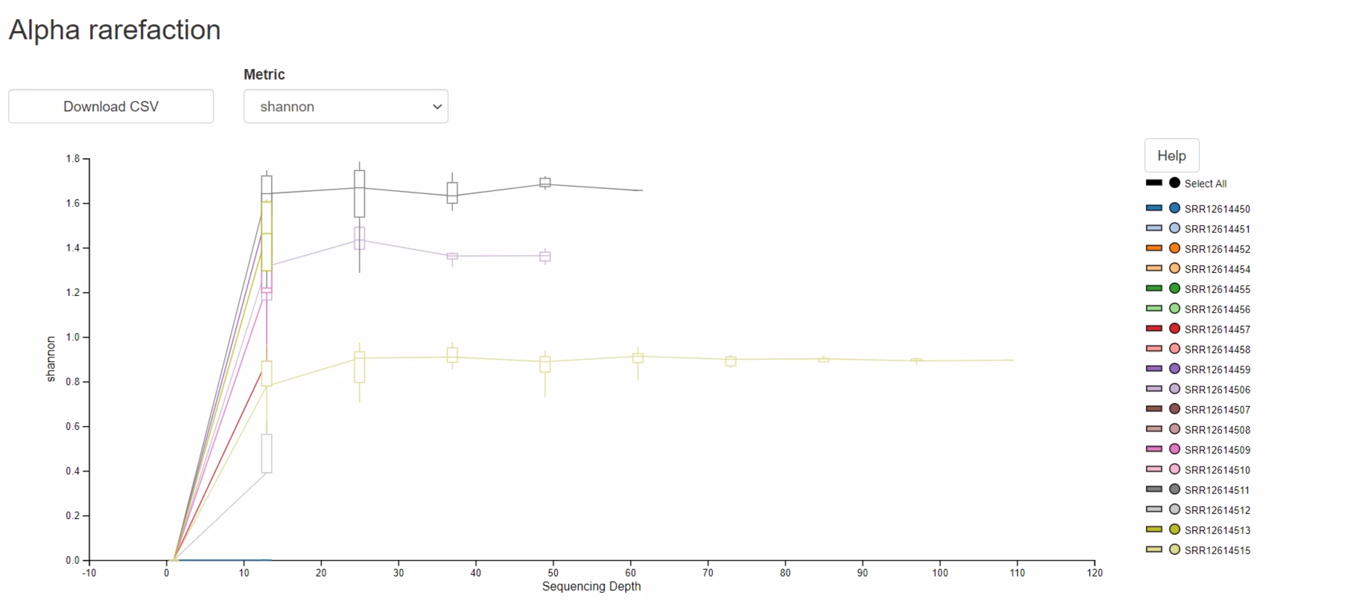
**Para entregar:**

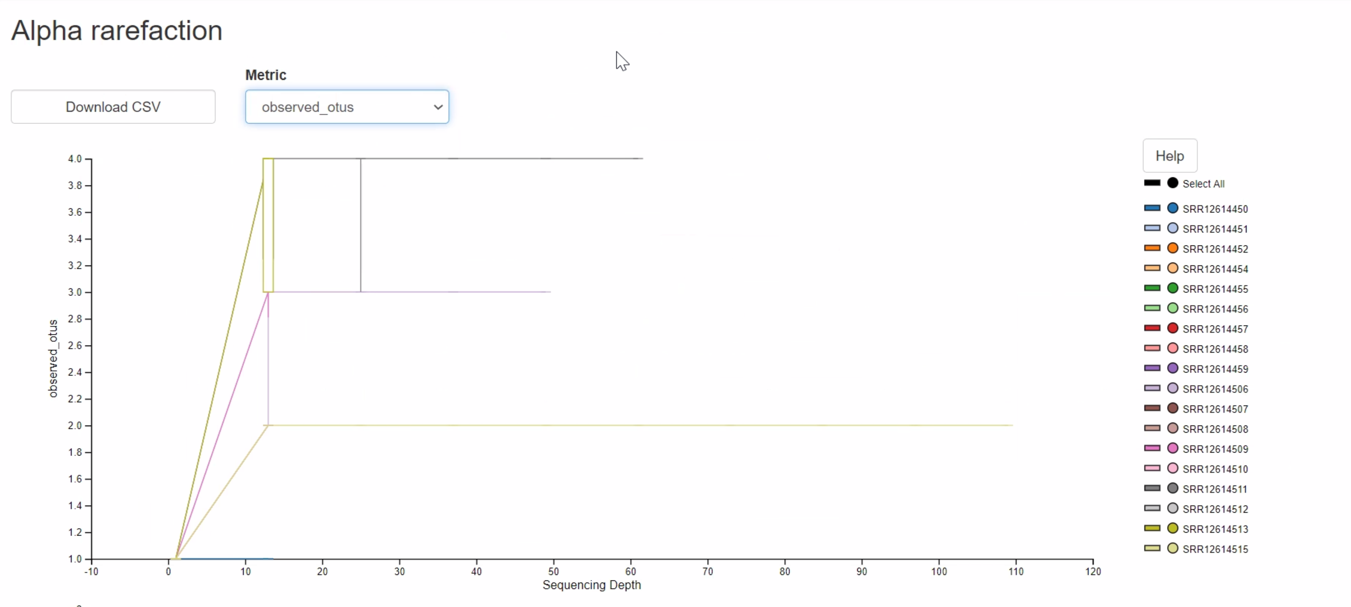
* Ponga el valor de rarefacción seleccionado y sustente el por qué lo escogió.
* Añada las imágenes de soporte que crea pertinentes.

El valor de rarefacción que escogímos es dos, porque es un valor más alto (relativamente) que las muestras que tienen menor número de lecturas y es cercano a las muestras que se ubican más arriba, las cuales se encuentran en el orden de las unidades.

Al observar la figura también es claro como al escoger esa profundidad de muestreo, se retienen aproximadamente un 9% de las muestras totales.

* Pegue la imagen/es generada por alpha-rarefaction y discuta si con ese valor de rarefacción se llegó a una saturación en la **diversidad y la riqueza.**

****

****

Al analizar los datos nos damos cuenta que la pendiente se aproxima a cero, esto sugiere que obtener secuencias adicionales más allá de las que se poseen no proveería información nueva pues en el momento en el que la pendiente se estabiliza, podemos afirmar que se llegó a una saturación en la diversidad y riqueza.

* ¿Qué métricas utilizó y que ventajas considera que tienen?

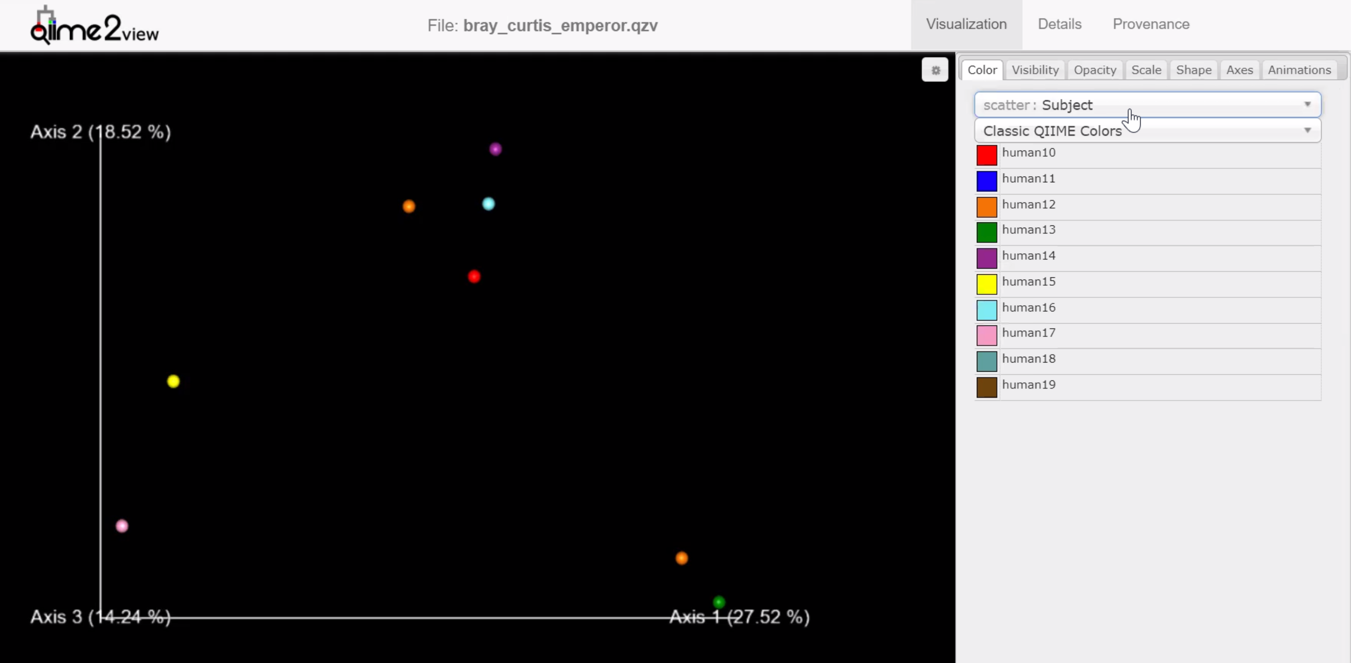
Las métricas que utilizamos fueron el índice Shannon y la cantidad observada de OTUs.

1. Ahora generaremos los valores de alfa y beta diversidad. En este caso se usará un workflow que produce todos los artefactos necesarios. (alpha-qiime diversity core-metrics-phylogenetic). **Tip**: Aunque la documentación pide una gran cantidad de outputs, no es necesario ponerlos todos. Pueden usar el flag “--output-dir” para generar un directorio con todos los archivos de salida.

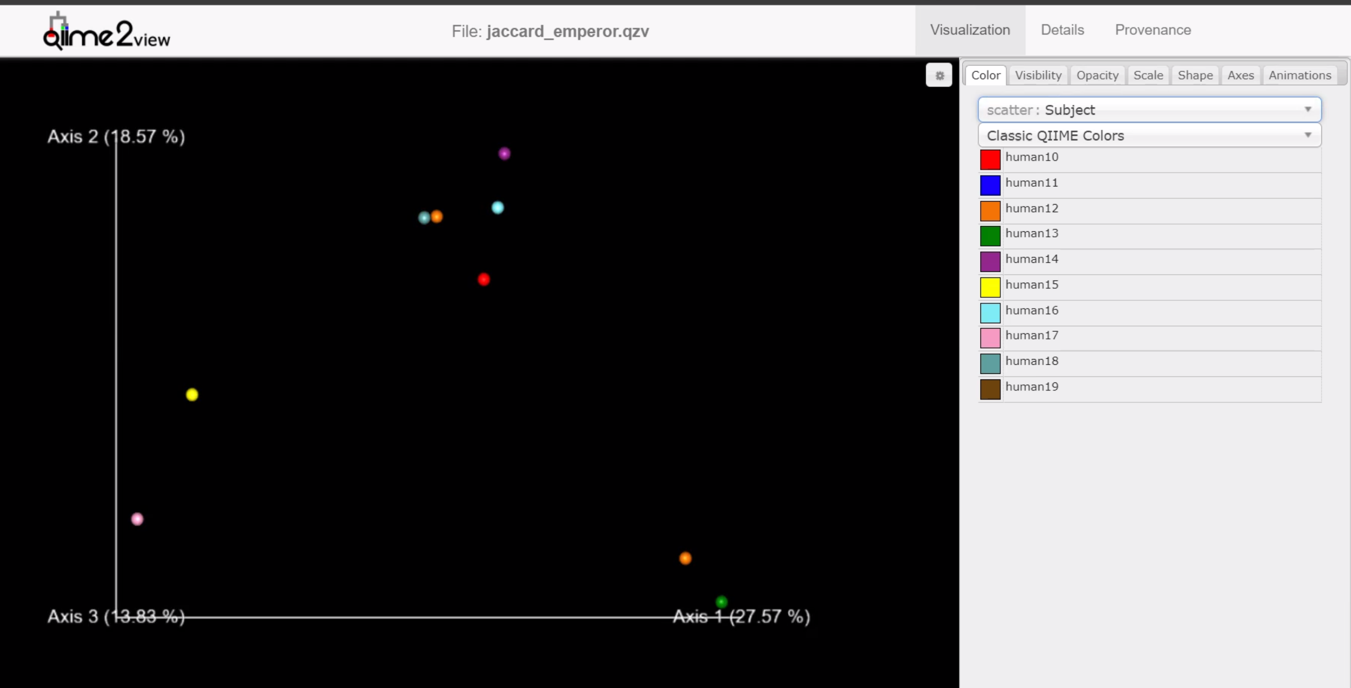
**Para entregar:**

* Visualice y pegue las imágenes para dos graficas de métricas de beta diversidad.

Bray\_curtis analysis



Jaccard\_analysis



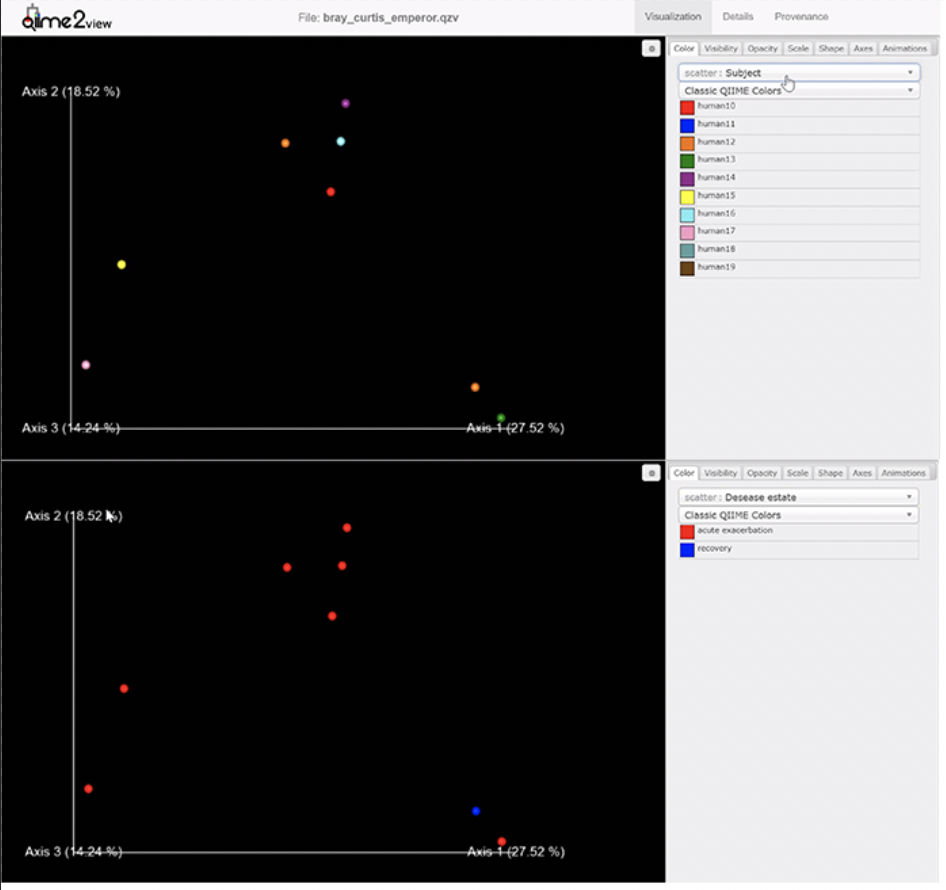
* Reporte si hay diferencias en el ordenamiento espacial de las muestras y valor de los ejes entre las dos y discuta a que se puede deber esto.

Sí hay diferencias en el ordenamiento espacial, partícularmente en el humano 18, el cual, en el caso de Jaccard\_emperor.qzv se presenta a la izquierda del humano 12 pero en el caso de Bray\_curtis\_emperor.qzv no se ve.

Por otra parte, hay diferencias en el valor de los ejes de Jaccard\_analysis con respecto a Bray\_curtis analysis, debido a que el análisis de Jaccard es una medida cualitativa que representa las diferencias de la comunidad; mientras que el análisis de Bray\_curtis es una medida cuantitativa que representa las diferencias de la comunidad.

Al observar atentamente las gráficas, resulta evidente que la localización cada una de las muestras es muy similar, variando únicamente en unas pocas muestras

* Ahora escoja una sola gráfica y analice si hay clusterización de muestras por alguna categoría de metadatos.

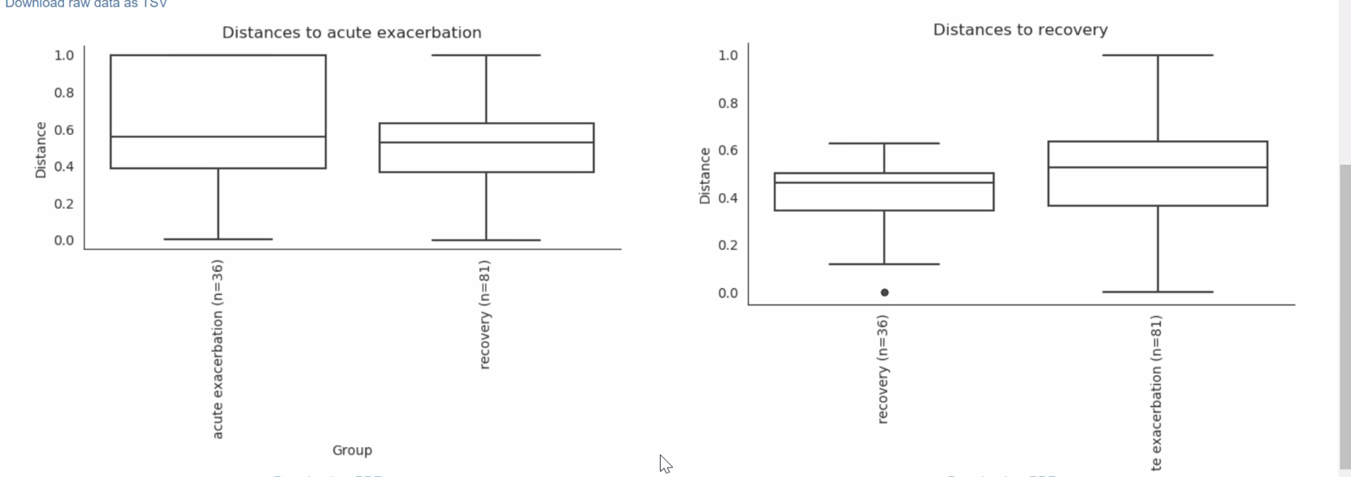
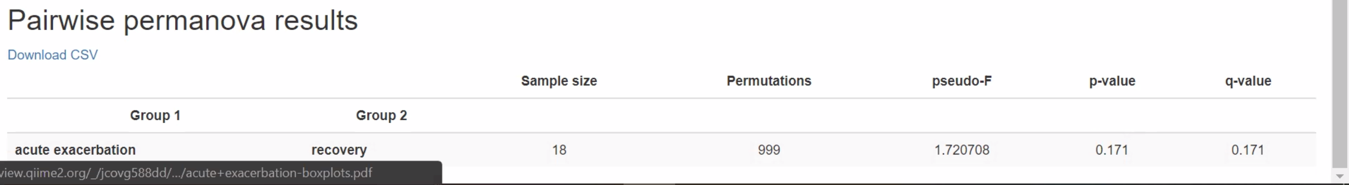


Al cambiar sucesivamente cada una de las categorías de los metadatos ( #SampleID, Source, Subject , Desease estate ,Location, sample\_title) resulta evidente que la categoría que presenta una agrupación mayor de los datos es la categoría de Desease estate mostrada en el lado inferior, en comparación con la muestra categorizada por Subject, mostrada en el lado superior.

1. Finalmente se analizará los valores de alfa diversidad y su significancia al agrupar por categorías. Para esto puede usar el método. ”beta-group-significance”.

**Para entregar:**

* Visualice el archivo de salida y analice si hay diferencias significativas en los valores de alfa diversidad entre pacientes en fase aguda y pacientes en etapa de tratamiento.

Al observar el valor p: 0.176, del permanova resulta claro que no existe una diferencia estadisticamente significativa entre los pacientes que se encuentran en fase aguda y los que se encuentran en fase de tratamiento; además, se observa cómo los diagramas de cajas y bigotes superponen sus rangos intercuartílicos por lo que no es posible distinguir entre ambas poblaciones visualmente. 

**Tomado de :**

<https://www.researchgate.net/post/Does-it-make-any-sense-to-use-both-Jaccard-> index-and-Bray-Curtis-coefficient-to-compare-similarity-matrices-of-two-datasets