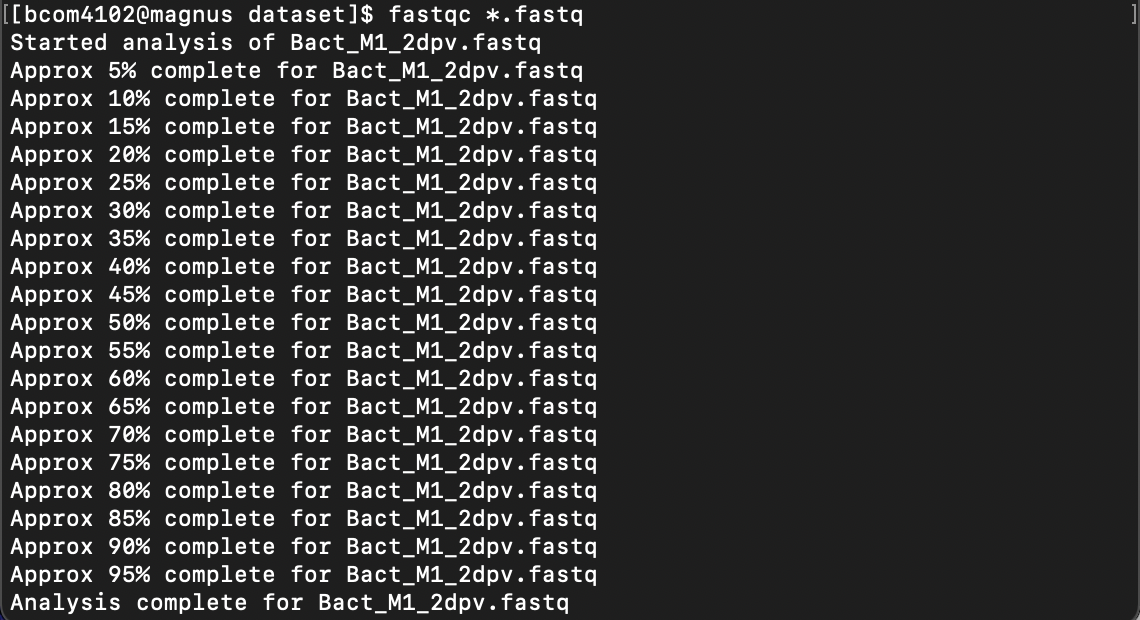
**MetaPhLAn**

MetaPhlAn (Metagenomic Phylogenetic Analysis) es una herramienta computacional para caracterizar la composición de comunidades microbianas a partir de datos de secuenciación de metagenómica tipo *shotgun*. MetaPhlAn se basa en genes marcadores específicos de clado únicos identificados a partir de genomas de referencia, lo que permite más eficiencia de uso computacional y asignaciones taxonómicas inequívocas.

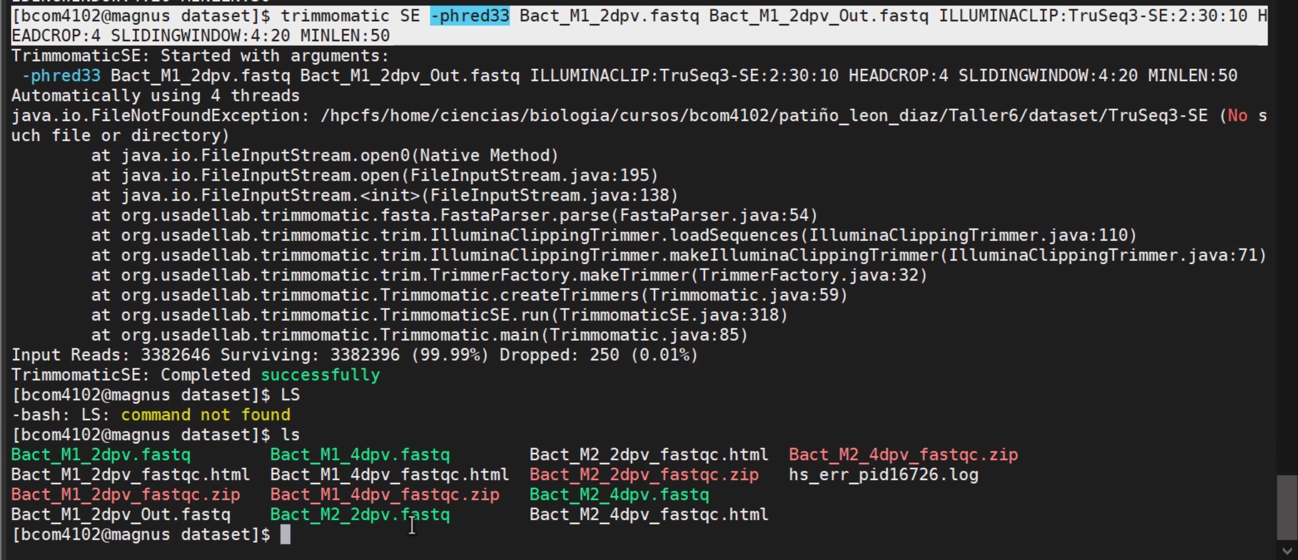
Para este taller usaremos secuencias Single End basadas en la tecnología Illumina derivada de ratones Gnotobióticos que contiene una comunidad de 15 bacterias distintas. Para este ejemplo particular usaremos datos de solo 2 ratones (M1, M2) de 2 y 4 días después de agregar partículas virales (2dpv, 4dpv) y con un submuestreo de solo 50,000 lecturas por muestra. Los datos se encuentran en **~/Datasets/Metagenomes/** los archivos de interés son todos los que comienzan con **Bact\_\***

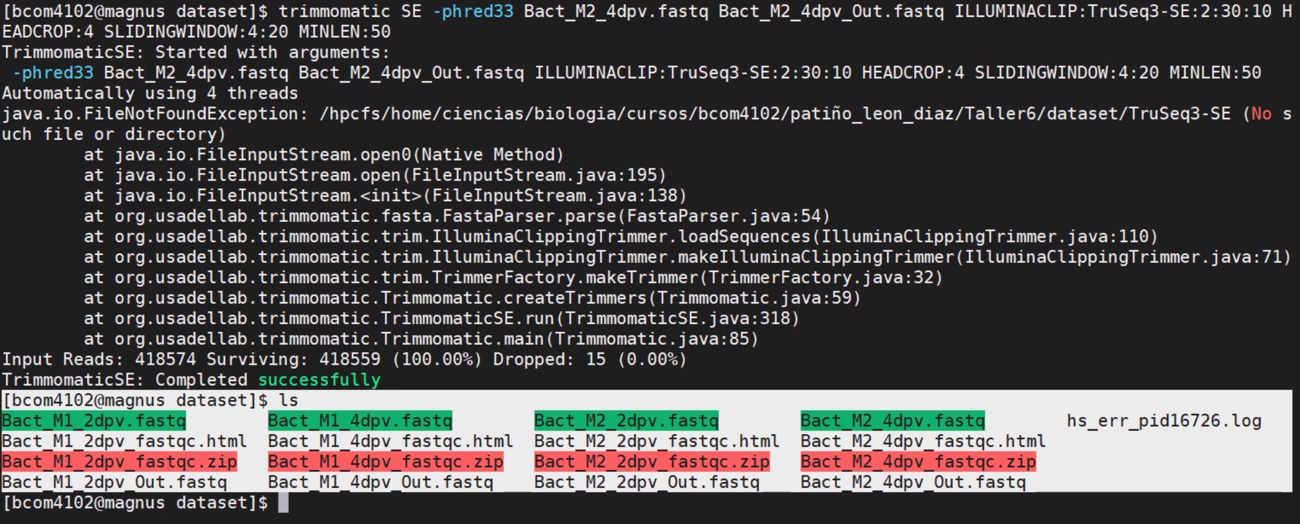
1. Primero es necesario limpiar las lecturas, lo haremos usando primero FastQC y el Trimmomatic como ya lo han usado anteriormente. Recuerden los módulos para estos dos programas: **fastqc / 0.11.2** y **trimmomatic/0.3 9**. Para el Trimmomatic use unos parámetros para el sliding window de 4:20, un Headcrop de 4 y un tamáño mínimo de secuencia de 50.
   1. En ambos casos adjunte los comandos usados y los pantallazos de los cambios en la calidad de las secuencias antes y después de la limpieza. También indique el porcentaje de secuencias que NO pasaron los filtros de la limpieza.

Comando usado: fastqc \*.fastq



Comando usado: trimmomatic SE -phred33 Bact\_M1\_2dpv.fastq Bact\_M1\_2dpv\_Out.fastq ILLUMINACLIP:TruSeq3-SE:2:30:10 HEADCROP:4 SLIDINGWINDOW:4:20 MINLEN:50





Repita el proceso para cada conjunto de archivos.

Aunque MetaPhlAn puede usar BlastN y BowTie2 en el paso de mapeo de lecturas a los marcadores, vamos a usar BowTie2 por razones computacionales.

Para correr MetaPhlAn es necesario cargar los siguientes módulos:

**bowtie/2-2.2.4**

**anaconda/python2.7**

**metaphlan2/2.0**

Familiarícese con la ejecución de MetaPhlAn con el siguiente commando:

**metaphlan2.py --help**

Si desea más información sobre el software lo puede consultar acá:

<https://github.com/biobakery/MetaPhlAn>

Ejecute MetaPhlAn sobre cada uno de los archivos FASTA limpios asegurándose de usar los siguientes parámetros:

bowtie2db: **/lustre/apps/metaphlan/2.0/db\_v20/mpa\_v20\_m200 ya**

bt2\_ps: **very-sensitive ya**

tax\_lev **'s' ya**

t: **rel\_ab**

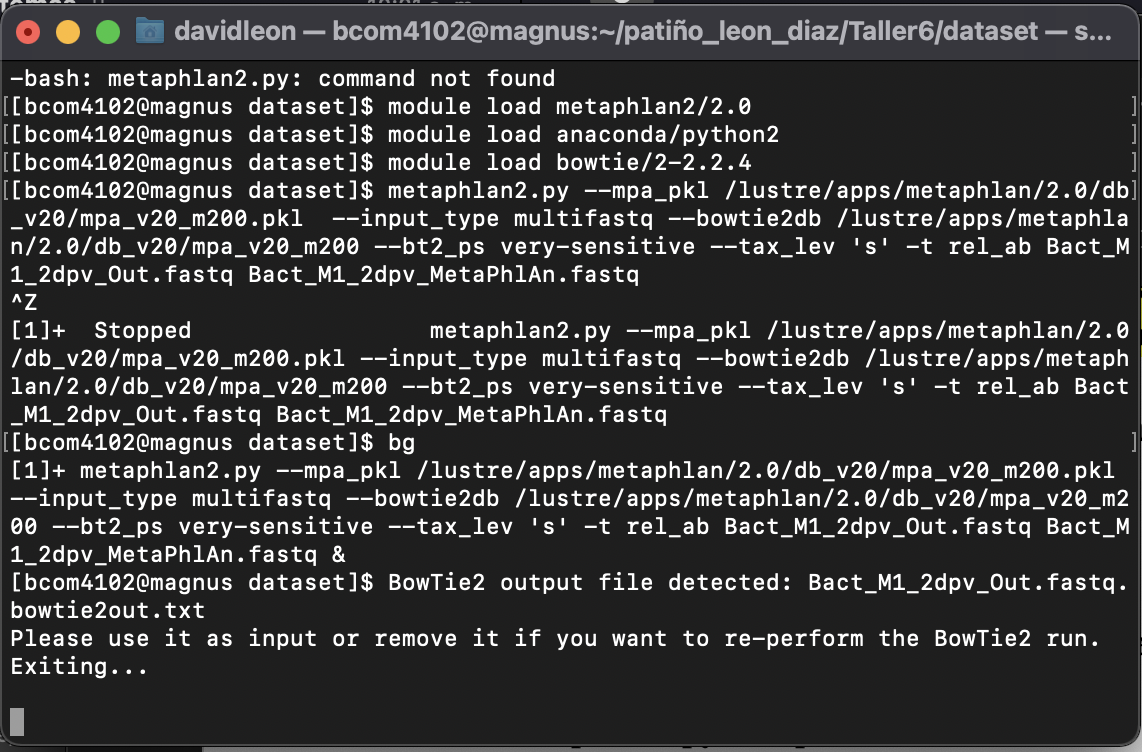
input\_type **multifastq ya**

mpa\_pkl **/lustre/apps/metaphlan/2.0/db\_v20/mpa\_v20\_m200.pkl ya**

metaphlan2.py --mpa\_pkl /lustre/apps/metaphlan/2.0/db\_v20/mpa\_v20\_m200.pkl  **--input\_type multifastq** --bowtie2db /lustre/apps/metaphlan/2.0/db\_v20/mpa\_v20\_m200 --bt2\_ps **very-sensitive** --tax\_lev **'s'** -t **rel\_ab** Bact\_M1\_2dpv\_Out.fastq Bact\_M1\_2dpv\_MetaPhlAn.fastq

1. Pruebe con un archivo primero y asegúrese que corre bien y envíe el resto como un trabajo de clúster.
   1. Cuál fue el comando que ejecutó?

metaphlan2.py --mpa\_pkl /lustre/apps/metaphlan/2.0/db\_v20/mpa\_v20\_m200.pkl  **--input\_type multifastq** --bowtie2db /lustre/apps/metaphlan/2.0/db\_v20/mpa\_v20\_m200 --bt2\_ps **very-sensitive** --tax\_lev **'s'** -t **rel\_ab** Bact\_M1\_2dpv\_Out.fastq Bact\_M1\_2dpv\_MetaPhlAn.fastq



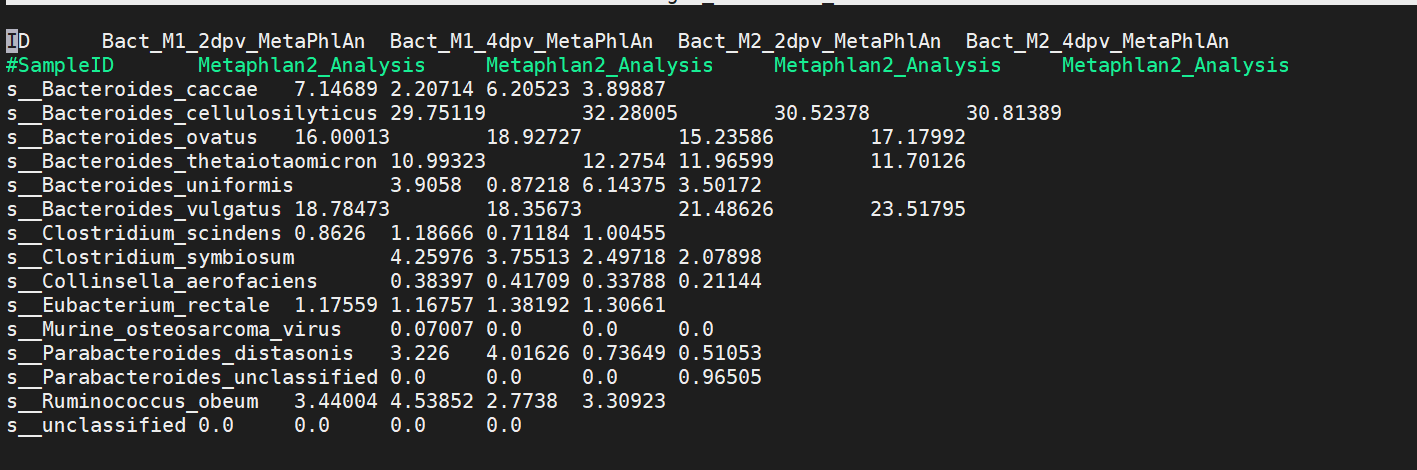
En el archivo de salida, verá el porcentaje de lecturas correspondientes a cada clasificación taxonómica, solo a nivel de especie. Es importante señalar que MetaPhlAn compara las lecturas con una base de datos de arqueas y bacterias.

Para visualizar los resultados , primero necesitamos fusionar las diferentes salidas en un solo archivo, para esto use el comando

**merge\_metaphlan\_tables.py**

**(es necesario nombrar por extensión )**

**merge\_metaphlan\_tables.py Bact\_M1\_2dpv\_MetaPhlAn.fastq Bact\_M1\_4dpv\_MetaPhlAn.fastq Bact\_M2\_2dpv\_MetaPhlAn.fastq Bact\_M2\_4dpv\_MetaPhlAn.fastq > merged\_taxonomic\_table**



Luego podemos generar un mapa de calor en PDF de los resultados metaphlan2/2.96.1

**metaphlan\_hclust\_heatmap.py**

Al implementar la siguiente línea de comando en la terminal se obtiene el mapa de calor descrito en la figura1;

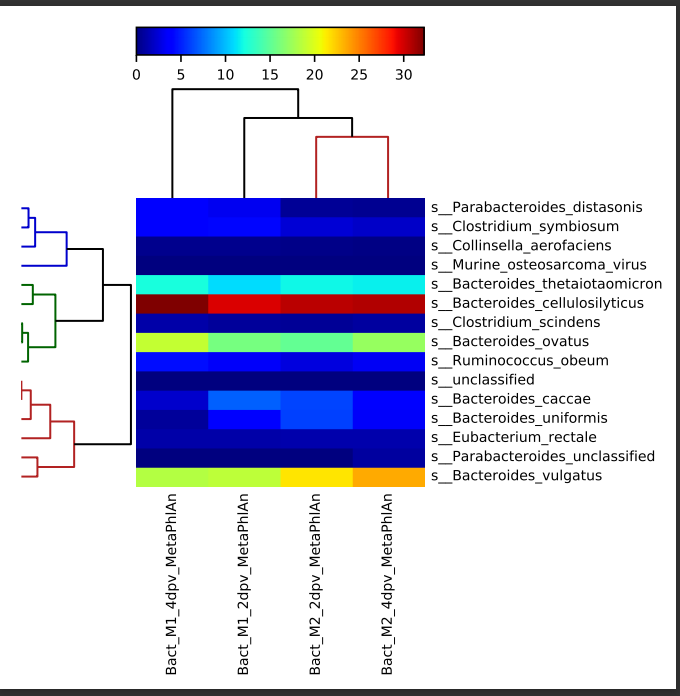
metaphlan\_hclust\_heatmap.py --in merged\_taxonomic\_table –out taxonomic\_heatmap.pdf

Descargue los resultados a su computadora y examínelos.

* 1. ¿Cuántos y cuales genomas diferentes se encontraron en las muestras?

De acuerdo con el mapa de calor resulta evidente que luego del análisis se logran identificar 14 genomas con su respectiva clasificación taxonómica y un genoma que no pudo ser clasificado correctamente. Dentro de los genomas identificado se observa que el género Bacterioides posee la mayor frecuencia en cuanto a las especies identificadas (6) seguido del género Clostridium (2) y por último géneros con una frecuencia de especies única (Parabacteroides,Murine, Ruminococcus,Eubacterium, entre otros)

* 1. Adjunte la imagen del PDF.



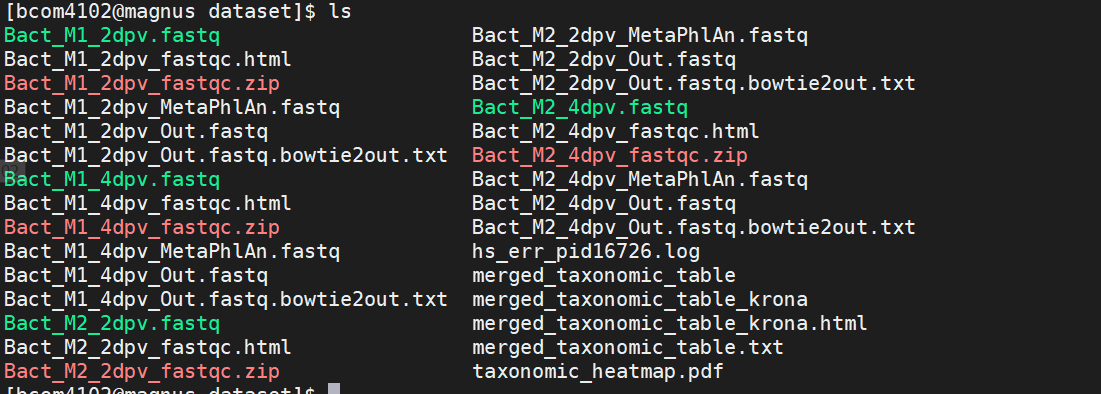
**Figura1.** Mapa de calor para las diferentes muestras

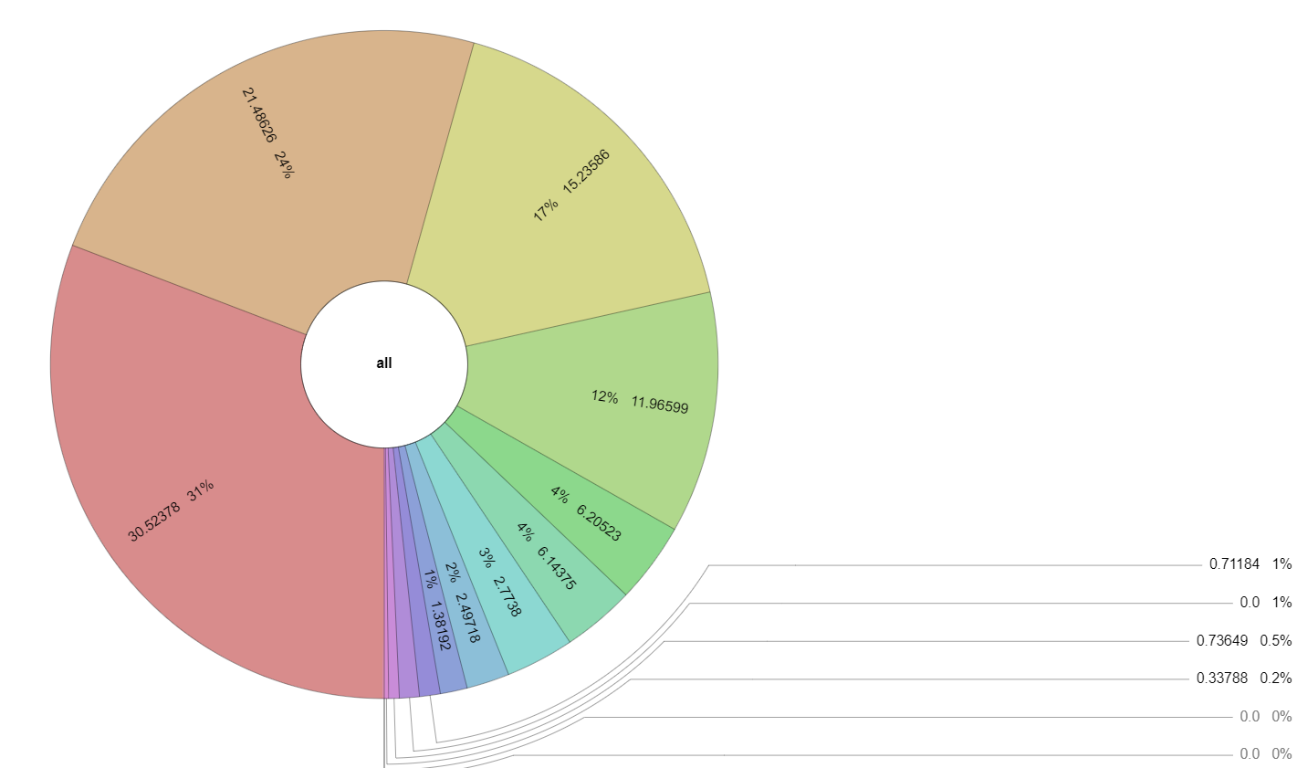
Es posible crear otras visualizaciones usando *pie charts*, puede hacerlo usando Krona, para esto primero necesita el módulo **krona/2.7** y el script **metaphlan2krona.py**. Necesitará este archivo para trazar gráficos circulares de Krona.

metaphlan2krona.py -p merged\_taxonomic\_table -k merged\_taxonomic\_table.krona

Finalmente, para construir el gráfico circular de Krona, use las herramientas de Krona, en particular **ktImportText**.

1. Descargue y explore los resultados con un navegador.
   1. Que comando usó para generar el gráfico?

ktImportText merged\_taxonomic\_table\_krona -o merged\_taxonomic\_table\_krona.htm

* 1. Adjunte el pantallazo de la imagen generada en el html.
  2. Que ventaja observa de usar este tipo de imágenes?

Este tipo de imágenes permite observar rápidamente la distribución de las especies

Si desea ver un gráfico de Krona con todos los niveles taxonómicos, puede volver a ejecutar el comando MetaPhlAn sin necesidad de ejecutar bowtie nuevamente para esto es importante modificar los siguientes parámteros:

tax\_lev: **'a'**

input\_type: **bowtie2out**

No es necesario darle los parámetros específicos de bowtie y el archivo de entrada es el que guardaron como archivo de salida de bowtie en la primera vez que lo corrieron.

Ejecute con la nueva salida los comandos **metaphlan2krona.py** y **ktImportText**

1. Descargue y explore los resultados con un navegador.
   1. Que comandos usó para generar este gráfico?

metaphlan2.py --mpa\_pkl /lustre/apps/metaphlan/2.0/db\_v20/mpa\_v20\_m200.pkl --input\_type bowtie2out --tax\_lev 'a' Bact\_M1\_2dpv\_MetaPhlAn.fastq Bact\_M1\_2dpv\_OutALL.fastq

* 1. Adjunte el pantallazo de la imagen generada en el html
  2. Que ventaja observa al usar todos los niveles taxonómicos?

**Qiime 2**

Ahora volveremos a los datos de 16S utilizados en las dos prácticas pasadas. En este caso utilizaremos un clasificador bayesiano para intentar asignar cada ASV a un grupo taxonómico.

*Entrenar el clasificador*

Debido a que este es un método basado en machine learning, es necesario utilizar un set de datos de entrenamiento del cual el clasificador pueda aprender. En vista de que para 16S hay bases de datos con una gran cantidad de secuencias y su respectiva clasificación taxonómica (GreenGenes, Silva RDP), normalmente estas son las que se utilizan como set de entrenamiento. En esta práctica usaremos la base de datos de GreenGenes.

Los archivos de entrada para entrenar el clasificador los encuentran en: /hpcfs/home/ciencias/biologia/cursos/bcom4102/Datasets/Talle6\_Qiime2

# Para hacer esto use el método “fit-classifier-naive-bayes” y solo ponga los parámetros requeridos. Nota: Se puede demorar un rato.

qiime feature-classifier fit-classifier-naive-bayes

--i-reference-reads ARTIFACT FeatureData[RepSeqs\_99\_otus.qza]

--i-reference-taxonomy ARTIFACT FeatureData[Taxonomy\_99\_otus.qza]

# *Crear la tabla de taxonomía*

# Usando como base el clasificador entrenado y las secuencias representativas, resultado del proceso de DADA2, use el método “classify-sklearn” para generar la tabla de taxonomía.

# qiime feature-classifier classify-sklearn --i-reads RepSeqs\_99\_otus.qza --i-classifier TestCla.qza --o-classification Taxonomy\_99\_otus.qza

# *Visualizar la taxonomía asociada a a cada ASV*

# En este caso use el método “barplot”, que resulta en un archivo que se puede visualizar en qiime view y que permite ver los resultados a diferentes niveles taxonómicos, así como por categorías de metadatos.

qiime taxa barplot \

--i-table feature\_table\_samples.qza \ <la tabla generada arriba>

--i-taxonomy Taxonomy\_99\_otus.qza\

--m-metadata-file ../metadata\_for\_qiime2.txt \ <no creo que vaya xq hasta ahora no tenemos metadata xd>

--o-visualization resultados\_barplot.qzv

# Para entregar

# Genere los barplots a nivel de phylum y género e inclúyalos en el taller.

* ¿Que tan común es encontrar “unassigned” en cada muestra? A nivel de género, ¿tan común es encontrar “k\_\_Bacteria;\_\_;\_\_;\_\_;\_\_;\_\_” en cada muestra?¿ Que le dice esto sobre la base de datos utilizada?

# Discuta si hay diferencias claras en la composición taxonómica entre pacientes en fase aguda y pacientes en etapa de tratamiento. Use el nivel taxonómico que prefiera.

# qiime feature-classifier fit-classifier-naive-bayes --i-reference-reads RepSeqs\_99\_otus.qza --i-reference-taxonomy Taxonomy\_99\_otus.qza --output-dir qiimefeautreclassifier

metaphlan2krona.py -p merged\_taxonomic\_table.txt -k

merged\_taxonomic\_table.html

metaphlan2krona.py -p merged\_taxonomic\_table.txt -k merged\_taxonomic\_table.krona.txt

/hpcfs/home/ciencias/biologia/cursos/bcom4102/patiño\_leon\_diaz/Taller6/Taller6Qiime