

Archivo de Consultas de Filtrado sobre Análisis Metagenómicos¹

Alejandro Navas González

2023-06-16

¹Andera Projekt

1. ¿Qué significa el filtrado por abundancia total de las OTUs?

La abundancia total de una OTU (Unidad Taxonómica Operativa) en todas las muestras se refiere a la suma de las cuentas (o lecturas) de esa OTU en todas las muestras. Por ejemplo, si se dispone de tres muestras y una OTU tiene 2 cuentas en la primera muestra, 1 cuenta en la segunda muestra y 2 cuentas en la tercera muestra, la abundancia total de esa OTU en todas las muestras sería $2 + 1 + 2 = 5$. Esto es diferente del umbral de presencia en un número de muestras determinado. Es decir, una OTUs puede tener una abundancia total de 10 y estar solamente presente en una muestra. En este contexto del análisis de microbiomas, una OTU con una abundancia total baja podría ser una especie rara en la comunidad, o podría ser un artefacto de secuenciación.

2. ¿Cuáles son los métodos de filtrado más comunes y en qué radica su importancia?

- **Filtrado por abundancia:** Este tipo de filtrado implica eliminar las OTUs que tienen una abundancia total (es decir, la suma de las cuentas en todas las muestras) por debajo de un cierto umbral. Este tipo de filtrado puede ser relevante para eliminar las OTUs raras o los posibles artefactos de secuenciación, que podrían añadir ruido a los análisis.
- **Filtrado por presencia en muestras:** Este tipo de filtrado implica eliminar las OTUs que están presentes en menos de un cierto número de muestras. Este tipo de filtrado puede ser relevante para eliminar las OTUs que son raras o que podrían ser artefactos de secuenciación.
- **Filtrado por taxonomía:** Este tipo de filtrado implica eliminar las OTUs que pertenecen a ciertos grupos taxonómicos. Este tipo de filtrado puede ser relevante, por ejemplo, para eliminar las OTUs de la clase de los cloroplastos o la familia de las mitocondrias, que son organismos que no son bacterias y que pueden estar presentes en los datos debido a la contaminación durante el proceso de secuenciación. También puede ser relevante para eliminar una especie

testigo que se haya añadido al experimento para controlar la calidad del proceso de secuenciación.

- **Filtrado por variabilidad:** Este tipo de filtrado implica eliminar las OTUs que tienen una variabilidad baja o alta en su abundancia entre las muestras. Este tipo de filtrado puede ser relevante para eliminar las OTUs que son constantes o muy variables entre las muestras, ya que podrían no ser informativas para los análisis posteriores.

3. ¿Qué sucede si en el filtrado por taxonomía se hallan especies del filo *Cyanobacteria* con la clase *Chloroplast*?

En los estudios de microbiomas, la presencia de secuencias asignadas al filo *Cyanobacteria* y a la clase *Chloroplast* puede ser indicativa de contaminación por ADN de origen no bacteriano. Los cloroplastos, que son los orgánulos responsables de la fotosíntesis en las plantas y las algas, son en realidad descendientes de cianobacterias que fueron incorporadas por una célula eucariota por un proceso llamado endosimbiosis. Como resultado, los cloroplastos comparten muchas similitudes genéticas con las cianobacterias, y las secuencias de ADN de los cloroplastos pueden ser erróneamente asignadas a las cianobacterias en los análisis de microbiomas. Por ende, si se observa una gran cantidad de secuencias asignadas a cloroplasto en los datos de la secuenciación, hay varias posibilidades:

- **Contaminación durante la recolección de muestras:** Si las muestras fueron recolectadas de un ambiente donde las plantas o las algas son abundantes, es posible que el ADN de los cloroplastos haya contaminado las muestras.
- **Contaminación durante la extracción de ADN o la secuenciación:** Si las muestras fueron procesadas en un laboratorio donde también se estaban manejando plantas o algas, es posible que el ADN de los cloroplastos haya contaminado las muestras o los reactivos.
- **Contaminación en la base de datos de referencia:** Si la base de datos de referencia que estás utilizando para asignar las secuencias a los

taxones contiene secuencias de cloroplastos que están mal etiquetadas como cianobacterias, esto podría llevar a una asignación errónea de las secuencias.

Ahora bien, se ofrece una opción alternativa y es que las secuencias asignadas a la clase *Chloroplast* sean de hecho procedentes de cianobacterias. Las cianobacterias son un grupo diverso de bacterias fotosintéticas que se encuentran en una variedad de ambientes, incluyendo el agua dulce, el agua de mar, el suelo y algunas condiciones extremas como los manantiales termales y las zonas áridas. Si las muestras provienen de un ambiente de este calado, donde las cianobacterias son comunes, es posible que las secuencias que se observan sean realmente de cianobacterias y no de cloroplastos. En este caso, no interesa filtrarlas. Para determinar si las secuencias son realmente de cianobacterias, se pueden considerar varias estrategias:

- **Revisar la asignación taxonómica a niveles más bajos:** Si las secuencias están asignadas a géneros o especies que son conocidos por ser cianobacterias, esto podría indicar que son realmente de cianobacterias.
- **Revisar la literatura y los datos existentes:** Si otros estudios han encontrado cianobacterias en el mismo tipo de muestras o en el mismo ambiente, esto podría apoyar la idea de que las secuencias son realmente de cianobacterias.
- **Realizar análisis filogenéticos:** Se podría construir un árbol filogenético con las secuencias y compararlo con árboles de referencia para ver si las secuencias se agrupan con las cianobacterias conocidas.

En última instancia, la interpretación de estos datos requerirá un juicio basado en el conocimiento del ambiente de muestreo, el proceso de secuenciación y análisis, y la biología de las cianobacterias y los cloroplastos.

4. ¿Qué significa el uso de *Aliivibrio fischeri* como testigo en el filtrado por taxonomía?

Aliivibrio fischeri es una especie de bacteria gram-

negativa que se encuentra comúnmente en ambientes marinos. *A. fischeri* es conocida por su capacidad para producir luz, un fenómeno conocido como bioluminiscencia. Esta bacteria forma una relación simbiótica con varios animales marinos, como el calamar hawaiano *Euprymna scolopes*, donde la bacteria coloniza un órgano especializado en el calamar y produce luz que el calamar utiliza para camuflarse. *A. fischeri* se utiliza a menudo como organismo modelo en la investigación de la bioluminiscencia y la simbiosis.

En el contexto de los análisis de microbiomas, un **organismo testigo**, también conocido como **control positivo**, es un organismo que se añade intencionalmente a las muestras o al proceso de secuenciación para controlar la calidad del experimento. Por ejemplo, se puede añadir una cantidad conocida de *A. fischeri* a tus muestras antes de la extracción de ADN. Luego, después de la secuenciación, se es capaz de detectar *A. fischeri* en los datos de secuenciación. Si no se pudiera detectar *A. fischeri*, o si la abundancia de *A. fischeri* fuera muy diferente de la cantidad añadida, esto podría indicar un problema con la extracción de ADN o la secuenciación. Además, al conocer la cantidad exacta de *A. fischeri* que se ha añadido, es posible utilizar este organismo testigo para calibrar los datos de abundancia. Esto puede ser útil para convertir las cuentas de lecturas de secuenciación, que pueden ser afectadas por factores como la eficiencia de la extracción de ADN y la profundidad de secuenciación, en estimaciones más precisas de la abundancia de los organismos en tus muestras.

Por lo tanto, el uso de un organismo testigo como *A. fischeri* puede ser una herramienta valiosa para controlar la calidad de los experimentos y mejorar la precisión de los análisis de microbiomas.