# Introducción a la Metagenómica $^1$

Alejandro Navas González

2023-10-23

 $<sup>^1</sup>$ Andera Projekt

### Introducción a la Metagenómica

La **metagenómica** es el estudio de los genomas microbianos colectivos que se encuentran en su entorno natural compartido. La metagenómica describe tanto un conjunto de técnicas de investigación como un campo de investigación (Ursell et al. 2012).

La caracterización genética de organismos demasiado pequeños para ser identificados visualmente requiere métodos computacionales para determinar cómo y dónde la expresión génica diferencial dentro de un microbioma contribuye a un ecosistema cooperativo y permisivo. Esta información proporciona un marco para estudiar la homología de secuencias, realizar la clasificación taxonómica y unir eficazmente conceptos de microbiología clásica con la genómica a nivel de organismos. Así, desarrollar un enfoque analítico de sistemas para comprender la genética de un microbioma proporciona un medio para comprender las relaciones entre diversas poblaciones microbianas que de otro modo nunca se estudiarían (Team 2019).

Un microbioma se define por las contribuciones colectivas de los organismos que lo componen y que generan un entorno metabólica y energéticamente favorable, un bioma. Con una caracterización genética más detallada de estas comunidades, se convierte en posible comprender los mecanismos evolutivos y describir las funciones de genes y proteínas (Committee on Metagenomics: Challenges and Applications 2007).

Además de proporcionar información sobre el microentorno en el que prosperan miles de organismos, la metagenómica nos proporciona una comprensión de los roles que desempeñan los organismos individuales en la salud y el bienestar humanos. La diversidad microbiana es vasta y las funciones que realizan incluyen la síntesis de nutrientes y vitaminas, la desintoxicación de compuestos ingeridos, la ayuda en la producción de energía y la protección contra patógenos. En este contexto la importancia de la metagenómica es tal que algunos de los descubrimientos más importantes de este campo provienen del estudio de las firmas genéticas de un microbioma. Esta información puede identificar patrones de ex-

presión génica novedosos o vías biosintéticas necesarias para mantener una comunidad ambiental. Al utilizar deliberadamente nuestra capacidad para extraer información significativa de un microbioma mediante técnicas de metagenómica, tendremos un impacto significativo en la salud global, la regulación de las vías metabólicas y las metodologías clínicas utilizadas para tratar enfermedades (Center 2014).

# Tipos de Secuenciación en la Metagenómica

La gran cantidad y diversidad de organismos disponibles para el estudio en microbiología y las condiciones sideralmente diferentes en las que viven hacen que la recopilación e interpretación de datos de secuencias metagenómicas sea algo necesario para comprender las relaciones entre y dentro de estas poblaciones. Estas herramientas para el análisis metagenómico se utilizan para extraer información contenida en diferentes secuencias genómicas de miles de organismos. Cada técnica varía en la cantidad de información que puede proporcionar, pero todas son útiles para comprender la conservación evolutiva entre organismos. Los métodos metagenómicos también son importantes para comprender las interacciones microbianas, realizar comparaciones de secuencias genómicas y realizar clasificaciones taxonómicas (Land et al. 2015).

Uno de los objetivos bien establecidos utilizados en la secuenciación es el ARN ribosómico (ARNr), que proporciona un molde de secuencia corta, conservada y fácil de leer. Los ARNr 16S y 18S son abundantes en prácticamente todos los miembros de un microbioma y contienen dominios tanto conservados como variables. Las diferencias en las secuencias de los dominios variables proporcionan estimaciones de la tasa y el alcance de la divergencia evolutiva entre los microorganismos. La homología entre las secuencias de ARNr 16S puede ser tan alta como el 99% dentro de un bioma. Como resultado, la secuenciación de dominios variables a menudo carece de la sensibilidad para identificar relaciones filogenéticas, lo que dificulta las comparaciones genéticas dentro de un

#### Secuenciación de ADN/ARN Shotgun

El análisis metagenómico también se puede realizar utilizando la secuenciación de shotgun de ADN o ARN. La secuenciación de shotgun identifica todas las secuencias de ADN o ARN presentes en poblaciones bacterianas, arqueas o virales. La información de la secuenciación de shotgun se obtiene a partir de múltiples lecturas de un solo genoma huésped, lo que reduce el sesgo de selección entre las muestras. La confianza en las lecturas correctas es alta y puede proporcionar suficiente información para inferir filogenias entre las muestras. Identificar cambios en las secuencias con la secuenciación de shotgun requiere un genoma de referencia para la comparación y la gran cantidad de datos proporcionará una imagen detallada del genoma microbiano y una comprensión de la función génica microbiana, la conservación de secuencias y la filogenia.

#### **ARNseq**

ARNseq es una estrategia de secuenciación metagenómica que identifica genes transcritos utilizando el ARN mensajero como plantilla. La secuenciación mediante ARNseq no se limita a los ARNm y también puede proporcionar información sobre todos los ARN codificados en una célula. Esta información ayudará a determinar qué combinaciones de genes se expresan simultáneamente en una muestra y cómo las redes de regulación génica son permisivas para la vida microbiana (Wang, Gerstein, and Snyder 2009).

Este instantánea de la transcripción activa puede describir la bioenergética de una muestra y definir redes biosintéticas clave dentro de comunidades bacterianas complejas. Así, ARNseq tiene varias ventajas sobre las tecnologías de microarrays, incluyendo un menor costo, un análisis indiscriminado de fusiones génicas, la detección de variantes de un solo nucleótido e inserciones/deleciones. La sensibilidad y especificidad de ARNseq lo hacen ideal para documentar la genética de un gran número de procariontes en nuestro entorno (Bashiardes, Zilberman-Schapira,

# Análisis de datos de metagenómica de amplicones de 16S

Como se ha explicado, el ARN ribosómico 16S (ARNr 16S) se puede amplificar y tras la amplificación comparar para identificar la información génica microbiana y su complejidad. Este ARNr 16S es el componente de la subunidad pequeña de proteínas 30S de un ribosoma procariota que se une a la secuencia de Shine-Dalgarno en un transcrito para facilitar el proceso de traducción.

La clasificación y el análisis de lecturas de ARNr se basan en la presencia conservada y prácticamente universal de secuencias de ARNr en los procariotas. Debido a las tasas lentas de evolución del ARNr, las secuencias se han vuelto útiles para reconstruir relaciones filogenéticas y taxonómicas de una comunidad microbiana.

Se cree que las diferencias entre el ARNr 16S resultan de cambios en las regiones variables menos conservadas del ARNr. Esto tiene su base en los estudios de Carl Woese y George E. Fox, pioneros en el uso del ARNr 16S en filogenia en 1990, en los cuales utilizaron secuencias de ARNr 16S para identificar linajes evolutivos distintos entre las arqueas. En sus estudios hallaron diferencias en el tamaño y el número de copias genómicas, así como la variabilidad inherente en las secuencias de ARNr 16S (Koonin 2014). Hacer estas comparaciones en una muestra ambiental puede proporcionar información sobre la abundancia de una bacteria y su relación con otros miembros de la comunidad.

La secuencia de ARNr procariota es pequeña, de alrededor de 1500 pares de bases, y las regiones del ARNr amplificadas en ARNseq varían. Los dominios constantes dentro de las secuencias de 16S facilitan el diseño de cebadores y la amplificación. La estructura del ARNr 16S también contiene nueve regiones variables además de estos dominios conservados. Las secuencias variables V3 y V4 dentro del ARNr suelen ser el objetivo común de secuenciación. La com-

paración de las secuencias hipervariables puede proporcionar estimaciones de la divergencia genética y la clasificación filogenética. Por lo tanto, los amplicones de 16S son esenciales para comprender el momento y el alcance del cambio evolutivo. Al analizar el número de copias de ARNr y los tamaños de los genomas en especímenes secuenciados, es posible calcular la abundancia y la diversidad microbianas.

Entonces, se comparan las secuencias de los amplicones de rRNA con longitudes de lectura variables con las secuencias de rRNA de longitud completa. Se diseñan cebadores para unirse a regiones altamente conservadas llamadas amplicones. La comparación entre los amplicones permite la clasificación taxonómica de las bacterias. El preprocesamiento asegura que las secuencias de entrada sean adecuadas para el cribado de alto rendimiento. El análisis diferencial y la normalización de las secuencias de rRNA 16S proporcionan datos de alta calidad y fácil interpretación que se pueden comparar con bases de datos preexistentes como SILVA, que contienen información de secuencias actualizada y completa (Li et al. 2017; "High Quality Ribosomal RNA Database" n.d.).

## Bibliografía

- Bashiardes, S., G. Zilberman-Schapira, and E. Elinav. 2016. "Use of Metatranscriptomics in Microbiome Research." *Bioinformatics and Biology Insights* 10: 19–25. https://doi.org/10.4137/BBI.S34610.
- Center, Genetic Science Learning. 2014. "How We Study the Microbiome." 2014. https://learn.genetics.utah.edu/.
- Committee on Metagenomics: Challenges, National Research Council (US), and Functional Applications. 2007. The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. Washington (DC): National Academies Press (US). https://www.ncbi.nlm.
- "High Quality Ribosomal RNA Database." n.d. n.d. https://www.arb-silva.de/documentation.
- Koonin, E. V. 2014. "Carl Woese's Vision of Cellular Evolution and the Domains of Life." *RNA Biology* 11 (3): 197–204. https://doi.org/10.4161/rna. 27673.

- Land, M., L. Hauser, S. R. Jun, I. Nookaew, M. R. Leuze, T. H. Ahn, and D. W. Ussery. 2015. "Insights from 20 Years of Bacterial Genome Sequencing." Functional & Integrative Genomics 15 (2): 141–61. https://doi.org/10.1007/s10142-015-0433-4.
- Li, X., G. N. Brock, E. C. Rouchka, N. Cooper, D. Wu, T. E. O'Toole, and S. N. Rai. 2017. "A Comparison of Per Sample Global Scaling and Per Gene Normalization Methods for Differential Expression Analysis of RNA-Seq Data." PloS One 12 (5): e0176185. https://doi.org/10.1371/journal.pone.0176185.
- Team, NIH Human Microbiome Portfolio Analysis. 2019. "A Review of 10 Years of Human Microbiome Research Activities at the US National Institutes of Health, Fiscal Years 2007-2016." Microbiome 7 (1): 31. https://doi.org/10.1186/s40168-019-0620-y.
- Ursell, L. K., J. L. Metcalf, L. W. Parfrey, and R. Knight. 2012. "Defining the Human Microbiome." Nutrition Reviews 70 (Suppl 1): S38–44. https://www.ncbi.nlm.nih.gov/pmc/articles.
- Wang, Z., M. Gerstein, and M. Snyder. 2009. "RNA-Seq: A Revolutionary Tool for Transcriptomics." Nature Reviews. Genetics 10 (1): 57–63. https://doi.org/10.1038/nrg2484.