



Proyecto: Entrega 2

AUTORES:

Mariana Aguirre Llano

Sofia Ospina Aristizábal

Mariana Salazar Díaz Granados

Juanita Vélez Uribe

**Profesor
Simón Villegas
Bioinformática**

**Universidad EIA
ENVIGADO
2025 – 02**

1. Descripción de clasificación taxonómica.

La clasificación taxonómica se realizó a partir de los reads de ADN que no fueron mapeados, es decir, aquellas secuencias que no lograron alinearse al genoma ancestral durante el mapeo con BWA. Por lo que, estas lecturas pueden corresponder a contaminantes biológicos, bacterias externas, virus asociados o elementos genéticos móviles que no forman parte del genoma de referencia. Así mismo, analizar estos reads permite detectar posibles organismos presentes en las muestras evolucionadas o cambios genéticos adquiridos durante el proceso experimental, lo cual resulta esencial para comprender la composición microbiana y el contexto evolutivo de las muestras (Wood et al., 2019).

Ahora bien, para llevar a cabo el análisis hicimos uso de dos programas que realizan la clasificación taxonómica, los cuales son Kraken2 y Krona. En el programa Kraken2 se desarrolla la clasificación taxonómica mediante la asignación de cada lectura a un taxón, utilizando un algoritmo basado en k-mers (fragmentos cortos de secuencia) que compara las lecturas con una base de datos de referencia (Wood et al., 2019). Este método ofrece una clasificación ultrarrápida y de alta precisión, generando reportes con el número y porcentaje de lecturas asignadas a cada nivel taxonómico, desde el dominio hasta la especie.

Por otro lado, Krona no clasifica directamente las secuencias, sino que visualiza los resultados generados por Kraken, permitiendo explorar de forma interactiva la distribución taxonómica mediante gráficos circulares jerárquicos. En estos diagramas, los niveles más internos representan categorías amplias (como Bacteria o Viruses), mientras que los anillos externos muestran clasificaciones más específicas, facilitando la interpretación de la diversidad taxonómica de la muestra (Ondov et al., 2011).

Resultados generales

En los resultados obtenidos para Evo1, la mayoría de las lecturas no mapeadas se clasificaron dentro del dominio Bacteria, principalmente del género *Escherichia* y especies relacionadas como *Escherichia coli* y *Escherichia virus T1*. Además, se detectaron lecturas correspondientes a virus bacteriófagos, lo que sugiere la posible presencia de elementos virales asociados al hospedador bacteriano.

En el caso de Evo2, se observó un patrón similar, aunque con un ligero aumento en la proporción de lecturas virales, incluyendo secuencias relacionadas con Enterobacteria phage lambda. Este incremento podría estar asociado a una mayor actividad de fagos o transferencia horizontal de genes, lo que es común en bacterias sometidas a presión selectiva o condiciones de evolución experimental (Touchon et al., 2017).

Ambas muestras mantienen una composición dominada por *Escherichia coli*, lo que indica la ausencia de contaminación biológica significativa y la persistencia de la especie ancestral como componente principal.

Teniendo en cuenta lo anterior, es válido mencionar que el hallazgo de secuencias virales dentro de los reads no mapeados es coherente con la naturaleza de las cepas bacterianas en evolución, ya que los bacteriófagos pueden integrarse o coexistir con sus hospedadores bacterianos, influyendo en su evolución genómica (Bobay et al., 2014). Por tanto, el análisis taxonómico sugiere que las líneas evolucionadas mantienen su identidad bacteriana, pero muestran diferencias en la presencia y abundancia de secuencias virales, lo cual podría representar procesos de interacción o coevolución entre fagos y bacterias a lo largo del experimento.

2. Tabla con variantes seleccionadas.

ID del Contig	Posición	Cambio (REF → ALT)	Tipo de variante	Calidad (QUAL)	Muestra	Interpretación Biológica
NODE_1_length_167691_cov_6.816352	95144	T → A	SNP	225.41	Evo1	Presenta un cambio puntual que podría modificar un codón en una región funcional, afectando la proteína.
NODE_1_length_167691_cov_6.816352	106180	A → T	SNP	206.32	Evo1 y Evo2	Es una variante compartida por ambas líneas, la cual, sugiere una mutación estable o adaptación común.
NODE_1_length_167691_cov_6.816352	106185	A → C	SNP	225.41	Evo1 y Evo2	Es una mutación puntual recurrente que puede presentar una posible alteración en la estructura proteica local.
NODE_1_length_167691_cov_6.816352	106186	A → T	SNP	180.17	Evo1 y Evo2	Son mutaciones consecutivas que indican una posible presión selectiva en esa región.

NODE_4_length_101311_cov_7.293844	73	C → A	SNP	221.06	Evo1	Es una variante heterocigota que podría influir en la regulación génica o expresión.
NODE_4_length_101311_cov_7.293844	6352	T → C	SNP	228.40	Evo1 y Evo2	Es una variante compartida, posiblemente neutral o conservada tras la evolución.
NODE_4_length_101311_cov_7.293844	94046	INDEL (AGTAATACNN NNNNNNNGT AATAC → AGTAATAC)	INDEL	228.15	Evo2	Ocurre una delección de bases ambiguas, probablemente mejora la calidad del ensamblaje o la estabilidad genómica.
NODE_4_length_101311_cov_7.293844	98871	C → T	SNP	71.91	Evo2	Corresponde a una mutación heterocigota, la cual, podría indicar variabilidad intracelular o mutación reciente.
NODE_4_length_101311_cov_7.293844	98875	T → C	SNP	88.08	Evo2	Cambio complementario cercano al anterior, sugiere una microvariación local.

3. Discusión breve sobre variantes de interés.

Las variantes identificadas en las muestras Evo1 y Evo2 muestran patrones consistentes con procesos de adaptación evolutiva bacteriana. Un ejemplo de esto es la presencia de mutaciones puntuales compartidas entre ambas líneas, como en las posiciones 106180, 106185 y 106186 del *contig* principal, sugiere la presencia de cambios beneficiosos que pudieron dar ventajas selectivas bajo las condiciones específicas del experimento. Al contrastar estos resultados con estudios previos en *E. coli*, se encuentra una concordancia entre estos, a partir de la cual se puede

decir que, la acumulación de mutaciones en genes relacionados con metabolismo, transporte y regulación confiere adaptaciones beneficiosas graduales y acumulativas a largo plazo (Tenaillon et al., 2016; Blount et al., 2012).

Como ya fue mencionado, se encontraron varias mutaciones consecutivas en regiones cercanas, lo que podría indicar zonas de inestabilidad local o “*hotspots*” de mutación, que es un fenómeno común en bacterias sometidas a presión ambiental o estrés oxidativo (Jensen & Galán, 2018). Estos cambios, aunque sean pequeños, pueden influir en la actividad enzimática e incluso en la expresión génica, lo que contribuye a la variación en los fenotipos e influye y la capacidad adaptativa de las bacterias.

En contraste, la delección observada en Evo2 en la posición 94046 puede asociarse con la pérdida de información genética como resultado de procesos de adaptación. La eliminación de secuencias no codificantes es un proceso que ya ha sido observado en estudios previos y que permite disminuir la demanda de ATP de ciertos procesos metabólicos ambientes controlados (Koskiniemi et al., 2012; Ochman & Moran, 2001). Los fagos, cuya presencia fue detectada en los análisis taxonómicos, también pueden llegar a tener un rol significativo en este tipo de mutaciones ya que, por ejemplo, los bacteriófagos pueden promover la recombinación y reorganización genómica y, como consecuencia, derivar en una aceleración de la evolución bacteriana (Bobay et al., 2014; Touchon et al., 2009).

En conjunto, los resultados muestran que las dos líneas evolucionadas conservan la identidad de *E. coli*, pero presentan cambios genéticos que apuntan a un proceso activo de adaptación. Las mutaciones compartidas, las delecciones y las posibles interacciones con fagos revelan cómo incluso en períodos relativamente cortos pueden darse modificaciones genómicas que reflejan la dinámica y el ritmo de la evolución bacteriana (Tenaillon et al., 2016).

4. Bibliografía

- Bobay, L. M., Touchon, M., & Rocha, E. P. C. (2014). *Pervasive domestication of defective prophages by bacteria. Proceedings of the National Academy of Sciences*, 111(33), 12127–12132. <https://doi.org/10.1073/pnas.1405336111>
- Blount, Z. D., Barrick, J. E., Davidson, C. J., & Lenski, R. E. (2012). *Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature*, 489(7417), 513–518. <https://doi.org/10.1038/nature11514>

- Jensen, P. A., & Galán, J. E. (2018). *Sensing and adapting to environmental stress: Lessons from enteric bacteria*. *Nature Reviews Microbiology*, 16(8), 465–476. <https://doi.org/10.1038/s41579-018-0032-9>
- Koskiniemi, S., Sun, S., Berg, O. G., & Andersson, D. I. (2012). *Selection-driven gene loss in bacteria*. *PLoS Genetics*, 8(6), e1002787. <https://doi.org/10.1371/journal.pgen.1002787>
- Ochman, H., & Moran, N. A. (2001). *Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis*. *Science*, 292(5519), 1096–1099. <https://doi.org/10.1126/science.1058543>
- Tenaillon, O., Barrick, J. E., Ribeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., ... & Lenski, R. E. (2016). *Tempo and mode of genome evolution in a 50,000-generation experiment*. *Nature*, 536(7615), 165–170. <https://doi.org/10.1038/nature18959>
- Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., ... & Rocha, E. P. C. (2009). *Organized genome dynamics in the Escherichia coli species results in highly diverse adaptive paths*. *PLoS Genetics*, 5(1), e1000344. <https://doi.org/10.1371/journal.pgen.1000344>