

## Haemophilus influenzae:

primer genoma de bacteria secuenciado

Laura Brenes-Guillén



Rev. Biol. Trop. Blog Serie 4

Actualmente existen más de 25 000 genomas bacterianos disponibles en diferentes bases de datos y numerosos recursos bioinformáticos que permiten estudiar genomas de **organismos procariotas**. Estos genomas sirven de base para hacer múltiples comparaciones y fortalecer el estudio de los microorganismos. En este blog destaco la historia del primer genoma bacteriano secuenciado.

Varios genomas virales y organelas habían sido secuenciados antes de 1995 — por ej., el del bacteriófago Phi-X174 por Fred Sanger y colaboradores en 1977, el genoma de *Vaccinia* (virus vacuna), genomas de cloroplastos de *Marchantia* y el genoma del virus que causa **viruela**—, año en el que Venter y colaboradores publican el primer genoma bacteriano: ***Haemophilus influenzae***.

La primera revolución en el estudio de la secuencia del genoma bacteriano inició en 1990, con proyectos relacionados a la secuenciación del genoma de bacterias modelo, como *Escherichia coli* y *Bacillus subtilis*<sup>1</sup>. No obstante, fue hasta 1995 cuando secuencian por primera vez el genoma de una cepa no patogénica de *H. influenzae*, utilizando la tecnología de **secuenciación shotgun**<sup>2</sup>.

*El tamaño del genoma de H. influenzae era diez veces mayor que cualquier virus secuenciado hasta ese momento*

Craig Venter fundó el Instituto para la Investigación del Genoma (**TIGR**). En una colaboración con Hamilton Smith, de la Escuela de Medicina de la Universidad Johns Hopkins —quien descubrió enzimas de restricción específicas en 1970 y recibió un Premio Nobel en 1978—, decidieron secuenciar un organismo mucho más complejo, utilizando lo que llamaron “secuenciación aleatoria de todo el genoma”. Con 1.8 millones de **pares de bases (bp)**, el tamaño del genoma de *H. influenzae* era diez veces mayor que cualquier virus secuenciado hasta ese momento. Para dicha secuenciación, se generaron bibliotecas de fragmentos de ADN desde 1 500 bp hasta 20 000 bp, y para su ensamblaje se utilizó el software TIGR, el cual permitía comparar, agrupar y ensamblar los diferentes fragmentos. Posteriormente, se realizó la anotación y el mapa genético de la bacteria.

Estos pasos básicos son la base de los análisis que se realizan actualmente, no obstante, los algoritmos han mejorado y se han implementado más herramientas y programas, así como mejoras en las técnicas de secuenciación y preparación de bibliotecas. La *secuenciación aleatoria de todo el genoma* tuvo como objetivo ensamblar la totalidad de un genoma a partir de fragmentos de ADN parcialmente secuenciados con la ayuda de un modelo computacional. Este enfoque no necesitaba de un mapa físico preliminar del genoma.

Años después, a mediados de los años 2000, se inicia con la secuenciación de “nueva generación” o “alto rendimiento” (“*high-throughput sequencing*”), acompañada de nuevos enfoques bioinformáticos. Recientemente, nuevos avances han permitido generar fragmentos de mayor longitud y en conjunto con la información generada por la secuenciación de nueva generación se han logrado ensamblar genomas de mejor calidad.

Laura Brenes-Guillén

Centro de Investigación en Biología Celular y Molecular (CIBCM), Universidad de Costa Rica  
San José, Costa Rica

### Imágenes

Ilustración del ADN. Fuente: Darryl Leja, National Human Genome Research Institute (CC BY 2.0)

### Referencias

<sup>1</sup>Loman, N. J., & Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nature Reviews Microbiology*, 13(12), 787-794.

<sup>2</sup>Fleischmann, R. D., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223), 496-512.

Publicado: 8 de julio, 2019. Serie 4.