



# Genómica y Bioinformática, mucho más que secuencias...

**Por Dr. Carlos Ramirez y Lcdo. Rafael Puche**  
 cdramirez@yandex.com, rafaelpucheq@yandex.com.

En estas dos primeras décadas del siglo 21 no es para nada desconocido que en cualquier reunión o espacio académico se hable del ADN (ácido desoxirribonucleico), la molécula universal de la herencia que constituye la base genética de la vida, por lo que su conocimiento molecular puede darnos la clave de muchos fenómenos vitales que ocurren en los organismos vivos. Esta molécula tiene un significado muy particular para aquellos que trabajan a diario con sus secuencias y combinaciones, tanto que ha llegado a ocupar un sitio casi privilegiado entre las llamadas Ciencias Biológicas.

El estudio de las secuencias que codifican los genes y los productos que se van generando (sintetizando) en la transmisión de la información que se encuentra codificada en el ADN no hubiera sido posible conocerla sin el desarrollo de técnicas y estrategias capaces, de forma rápida y barata, abordar su estudio detallado, ya que era necesario analizar cómo se encontraban ordenadas las secuencias y el significado biológico que las mismas representaban. Nace así la era de la Genómica. La genómica comprende el estudio del contenido, organización, función y evolución de la información genética en un genoma completo (el conjunto de genes presentes en cualquier organismo vivo). El término genómica es relativamente reciente. Se considera que fue acuñado por Thomas Roderick, en 1986, para referirse a la subdisciplina de la genética dedicada al estudio de la cartografía, secuencia-

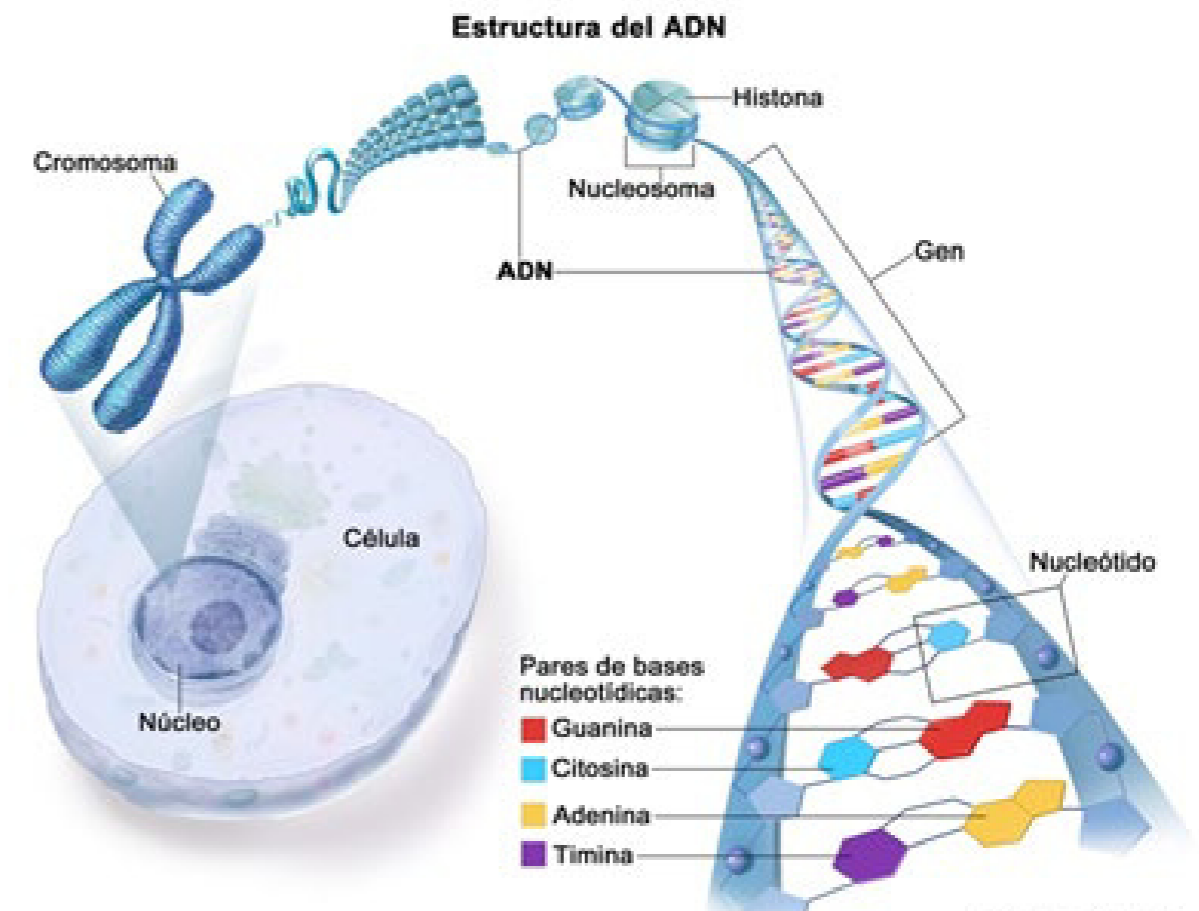
ción y análisis de las funciones de genomas completos.

## Descifrar y conocer el ADN; un proyecto a escala global

Una de las principales características que define la estructura del ADN es su secuencia nucleotídica, la cual almacena toda la información necesaria para la formación de un organismo vivo. Por este motivo, conocer el orden de dicha secuencia permite poder identificar los genes contenidos y estudiar su función de forma detallada. Desde mediados de los años 70 del siglo 20, los biólogos moleculares habían tratado de averiguar la secuencia de genes individuales de muchos organismos, así como genomas de entidades subcelulares como, por ejemplo, algunos virus o plásmidos. Previo a esas décadas, se desarrollaron las metodologías necesarias para emprender el reto de analizar las secuencias proteicas de los primeros organismos. Frede-

rick Sanger y E.O. Thompson publican la primera secuencia de aminoácidos de la insulina (The Amino-acid Sequence in the Glycyl Chain of Insulin. Biochem J. 1952 Sep; 52(1):iii) en 1952. Es decir, un año antes de que se publicaran los tres trabajos fundamentales sobre la elucidación de la estructura del ADN por James Watson, Francis Crick, Maurice Wilkins, Rosalind Franklin y Robert Goslin (ver una versión más detallada y desde una perspectiva histórica del descubrimiento del ADN en Klug, A. J. Mol. Biol. (2004) 335, 3-26).

Desde entonces, diferentes grupos de investigación trataban de buscar financiación para la secuenciación de organismos superiores. En 1986, el Departamento (Ministerio) de Energía (DOE), en EEUU, durante un congreso en Santa Fe (Nuevo México), planteó la idea de destinar una buena partida presupuestaria para



la secuenciación del primer genoma humano, como medio para evaluar el efecto de las radiaciones sobre el material hereditario. Un año más tarde, otro organismo público, el Instituto Nacional de Salud (NIH), se unía a la idea de que el Proyecto Genoma Humano (PGH) se convirtiera no solo en un emblema tecno-científico de finales de siglo XX, sino que también en la primicia tecnológica y comercial del país en el siglo XXI. Finalmente, el Proyecto Genoma Humano se lanzó en 1990, con el objetivo de ser completado en vista a 15 años.

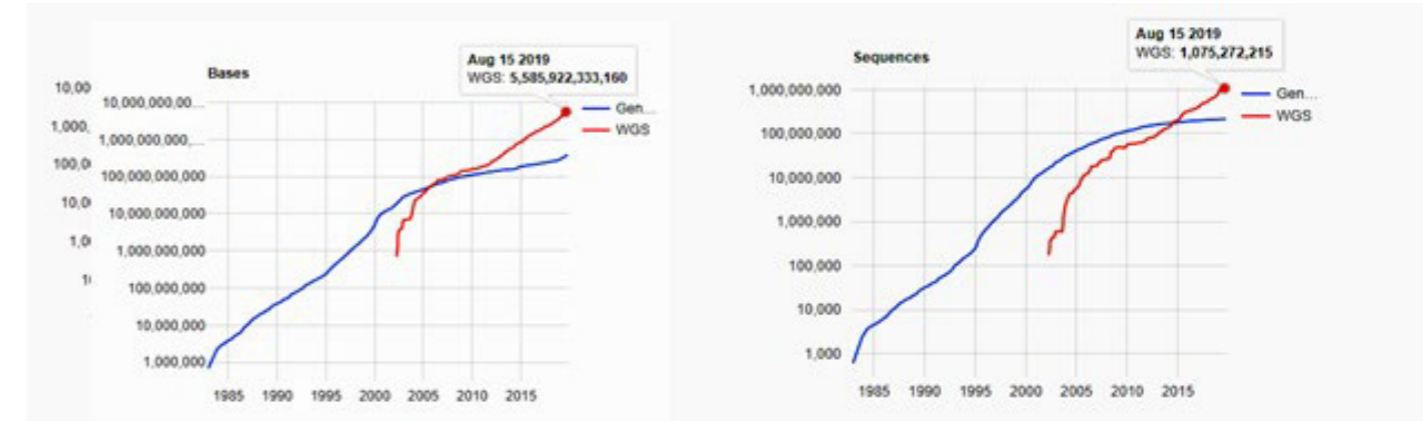
A partir de esa fecha, comenzó una carrera a contrarreloj entre las dos compañías que competían por alcanzar el objetivo (HUGO; Human Genome Organization, una iniciativa académica internacional y Celera Genomics, iniciativa privada con fines netamente comerciales), para que, finalmente en febrero de 2001, publicaran los primeros borradores, conjuntamente en dos prominentes revistas especializadas: Na-

ture y Science. La secuencia cubría únicamente el 90% del genoma, siendo el 10% restante no secuenciado lo correspondiente a heterocromatina (regiones de ADN muy compactado que están asociado a proteínas). Este hecho, propició la creación de nuevos consorcios dedicados a la secuenciación de otros genomas eucariotas, sobre todo, organismos modelos utilizados ampliamente en estudios básicos.

La era de la Genómica, es hoy en día una realidad palpable, convirtiendo esta disciplina en un soporte central y unificador de distintas áreas de investigación biológica. A diferencia de la Genética, que se centra en estudiar la herencia de forma aislada, mediante el estudio de la función y composición de genes individuales, en lugar de analizar la complejidad de su conjunto -lo cual incluye tanto las secuencias codificadoras como las no codificadoras del ADN. La Genómica lleva a cabo el estudio de la información hereditaria de uno o múltiples organismos de forma integral

y simultánea; es decir, a gran escala. No obstante, el estudio de toda esta inmensa información genera una cantidad tan importante de datos que se requieren de grandes capacidades computacionales para poder ser analizados, almacenados y gestionados. Es aquí donde entraría en juego, otra disciplina, cada vez en más auge, conocida como BIOINFORMÁTICA.

La Bioinformática puede definirse como un campo interdisciplinario que desarrolla y aplica metodologías computacionales que permiten analizar gran cantidad de datos biológicos, así como secuencias genéticas y proteicas, con la finalidad de hacer nuevas predicciones o descubrir nuevos significados a procesos biológicos. Los términos bioinformática, biología computacional, informática biológica y, en ocasiones, bio-computación, son utilizados en muchas situaciones como sinónimos, y hacen referencia a campos de estudios interdisciplinarios muy vinculados. Unidas y trabajando en conjunto, la Genómica y la Bioin-



"La genómica plantea claramente algunos de los desafíos computacionales más severos que enfrentamos en la próxima década. La genómica es una "bestia de cuatro cabezas"; considerando las demandas computacionales a lo largo del ciclo de vida de un conjunto de datos (adquisición, almacenamiento, distribución y análisis), la genómica está a la par o es la más exigente de los dominios de Big Data " (Stephens, ZD y col., Big Data: Astronomical or Genomical? PLoS Biol 13(7): e1002195. <https://doi.org/10.1371/journal.pbio.1002195>).

## Unidas y trabajando en conjunto, la Genómica y la Bioinformática han revolucionado la comprensión que tenemos sobre la vida y los fenómenos que subyacen a los seres vivos.

ha transformado la vigilancia realizada en los laboratorios de salud pública, proporcionando así datos para monitorear tendencias, realizar investigaciones y dar respuesta a brotes de enfermedades infecciosas. Pero también han cambiado la forma de abordaje de los estudios puesto que se ha pasado de hacer análisis mediante métodos convencionales (PGFE; electroforesis de geles mediante campos pulsados y MLST; tipificación de secuencias multilocus), que implican muchas horas y días para diagnóstico, a pasar en las últimas décadas a las nuevas tecnologías de secuenciación masiva de genomas (NGS, Next Generation Sequencing), las cuales permiten hacer análisis rápidos y precisos en poco tiempo.

Estas nuevas tecnologías fueron diseñadas para estudiar millones o billones de moléculas de ADN, que pueden ser secuenciadas en paralelo, mejorando sustancialmente el rendimiento y minimizando el uso fragmentos clonados, normalmente usados en la secuenciación de genomas mediante la técnica clásica de

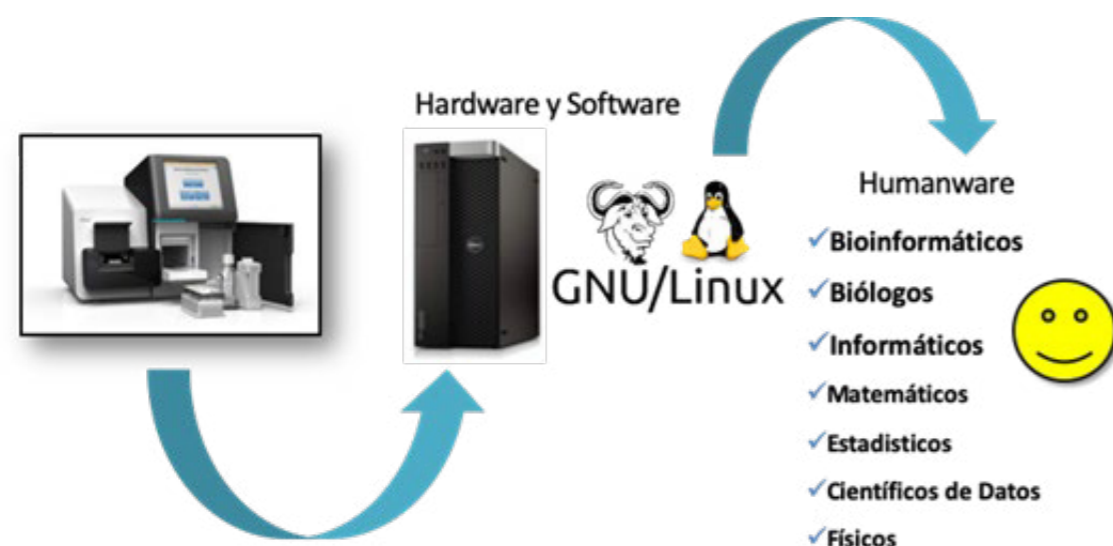
Sanger. El actual "cuello de botella" de los proyectos de secuenciación de genomas completos (WGS; whole-genome sequencing) y de los exomas (WES; whole-exome sequencing), no es la secuenciación del ADN en sí, sino en la forma de estructurar el manejo y almacenamiento de la data y los sofisticados análisis computacionales de los datos experimentales.

Para darnos una idea de la cantidad de información que se maneja actualmente en estas tecnologías, basta con los siguientes datos: para el mes de abril 2017 se habían producido más de 450 millones de secuencias, y para agosto del 2019, más de 1000 millones de secuencias se habían enviado a las grandes bases de datos. Las secuencias subidas a Genbank (la mayor base de datos genómicos en la actualidad) se duplican cada 18 meses (<https://www.ncbi.nlm.nih.gov/genbank/statistics/>).

### Aplicaciones de la Genómica y Bioinformática

Un ejemplo de cómo los nuevos conocimientos en estas

formática han revolucionado la comprensión que tenemos sobre la vida y los fenómenos que subyacen a los seres vivos. Pero ante todo nos ha permitido tener herramientas para analizar e intentar comprender las diversas relaciones que sobre la salud humana y animal tenemos hoy día. Este campo analiza como la tecnología y los conocimientos sobre la genómica pueden ser integrados y aplicados de forma efectiva en servicios de salud y políticas públicas en beneficio de la población. Su aplicación al estudio de patógenos





áreas y el manejo de herramientas bioinformáticas para analizar esta avalancha de datos biológicos nos ayudan

de mortalidad entre un 20 a 30%, más alta que casi todas las restantes toxicoinfecciones alimentarias. Al combinar la secuenciación del genoma completo en tiempo real con datos de pacientes sobre los alimentos que comieron y datos sobre listeria en los alimentos, los funcionarios de salud pública pueden: detectar más grupos (posibles brotes) de infecciones por *Listeria*, vincular estos casos a una fuente probable, identificar fuentes no reconocidas de listeria, detener los brotes de *Listeria* mientras aún sean pequeños (<https://www.cdc.gov/listeria/surveillance/whole-genome-sequencing.html>). Con la secuenciación del genoma completo, los brotes de *Listeria* se pueden detectar cuando tan solo dos personas se han enfermado. Determinar que la misma cepa de *Listeria* está enfermando a las personas es una indicación de que estas enfermedades pueden provenir de la misma fuente, por ejemplo, el mismo alimento contaminado o la misma instalación de procesamiento de alimentos contaminados. Por lo tanto, la secuenciación del genoma completo puede ayudar a los detectives de enfermedades a resolver brotes. Este ejemplo nos muestra como es el abordaje actual de los estudios sobre genómica de patógenos relacionados con enfermedades infecciosas su control y prevención.

En Venezuela, se han aplicado estas técnicas para el estudio de organismos que causan enfermedades como la leptospirosis, la cual es capaz de afectar a seres humanos por contaminación tras contacto indirecto

con roedores y otro tipo de animales, y es así como en el año 2018 se publicó un trabajo que logró identificar una nueva especie biológica (Int J Syst Evol Microbiol 2018 Feb 14;68(2):513-517. Epub 2017 Dec 14), la *Leptospira venezuelensis*, mediante el empleo de secuenciación masiva (NCBI Reference Sequence: NZ\_NFUP000000000.1). Esta nueva especie fue aislada de un paciente con leptospirosis de moderada a severa, una rata y una vaca perteneciente a una granja situada a 40 km de la zona donde vivía el paciente. El genoma de esta bacteria tiene aproximadamente 4,3 millones de pares de bases.

La caracterización inicial de estas cepas basadas en genes ribosomales (altamente conservados entre especies bacterianas, ARNr 16S) sugirió su

designación como una nueva especie dentro del grupo de "intermedios" del género *Leptospira*. La caracterización filogenómica adicional basada en genes de núcleo de copia única fue coherente con su separación en una nueva especie. La identidad promedio de nucleótidos entre estas tres cepas fue >99%, pero por debajo del 89% con respecto a cualquier especie leptospiral descrita anteriormente, lo que también apoya su designación como una especie nueva. Dada esta evidencia, se consideró que estos tres aislamientos representaban una especie novedosa, para la cual se propuso el nombre de *Leptospira venezuelensis* sp. nov.

La leptospirosis afecta a humanos y a animales domésticos y silvestres como perros y ganado, entre otros. Causa-

das por bacterias del género *Leptospira*, en humanos, las infecciones ocurren por contacto a través de lesiones en la piel o mucosas, o bien indirectamente, por agua o tierra contaminada, por orina o tejidos de animales infectados. La gravedad de la enfermedad está dada por el nivel de virulencia de la especie de bacteria. Anualmente, se estima que la leptospirosis afecta a un millón de personas, sobre todo en países tropicales y en desarrollo, y causa la muerte de 60.000, pero aún existe poca información y está subestimada. En América Latina, su incidencia varía entre 1 y 22 casos por 100.000 habitantes.

Los ejemplos anteriormente descritos muestran de forma tangible y práctica como la conjunción de dos disciplinas emergentes dentro de las

ciencias biológicas, la genómica y bioinformática, permiten dar soluciones concretas a diversos problemas en un tiempo de respuesta relativamente corto. Además, no son tecnologías ni conocimientos inaccesibles en nuestro país, actualmente se desarrollan en el Instituto Venezolano de Investigaciones Científicas (IVIC) el análisis de datos generados de secuenciación masiva y la genómica comparativa de bacterias como líneas de investigación pioneras aplicadas a la resolución de problemas de salud pública.

**No son tecnologías ni conocimientos inaccesibles en nuestro país, actualmente se desarrollan en el Instituto Venezolano de Investigaciones Científicas (IVIC) el análisis de datos generados de secuenciación masiva y la genómica comparativa de bacterias como líneas de investigación pioneras aplicadas a la resolución de problemas de salud pública.**

a plantearnos los retos entre las enfermedades infecciosas, lo podemos evidenciar en el siguiente ejemplo: *Listeria monocytogenes* es una bacteria que se desarrolla intracelularmente y es causante de la listeriosis. Es uno de los patógenos causante de infecciones alimentarias más violentos, con una tasa

