Recibido: 15-jul-2025 Aceptado: 01-may-2026 e-ISSN 2954-4890

Sección: *¿Cómo funciona?*

¿Cómo 'traducimos' el ADN? La ciencia detrás de la lectura de nuestro código

How do we 'translate' DNA? The science behind reading our genetic code

*Valentín Pérez-Hernández**
Mario Hernández-Guzmán

Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Baja California, México.

**Autor para la correspondencia: vperezhdez@hotmail.com*

RESUMEN

La información hereditaria, que nos transmiten nuestros padres biológicos, como la estatura o la predisposición a ciertas enfermedades, se almacena en el ADN de nuestras células. Nuestro ADN funciona como un manual de instrucciones escrito en un alfabeto de cuatro letras llamadas nucléotidos (adenina (A), citocina (C), guanina (G) y timina (T)), y que está protegido en el núcleo de cada célula: Para leerlo e interpretarlo requerimos de técnicas moleculares y cómputo avanzado (bioinformática); sólo así nos es posible entender las instrucciones para identificar posibles enfermedades hereditarias, y potencial uso de tratamientos tempranos.

Palabras clave: Bioinformática, secuenciación, salud humana.

SUMMARY

Hereditary information, such as height or predisposition to diseases, is contained in our DNA. DNA, made of four chemical building blocks called (A)denine, (C)ytose, (G)uanine, and (T)hymine, is stored safely inside the nucleus, a small compartment found in each cell. To read the "instructions" stored in DNA, scientists first collect it from cells, then use special tools and computers to examine the order of its building blocks to understand what the DNA says. The final goal is to identify whether our DNA have mutations associated with hereditary diseases, which in turn might enable future early human treatments.

Keywords: Bioinformatic, sequencing, human health.

Introducción

El ADN se descubrió a finales del siglo XIX, sin embargo, su estructura en espiral o doble hélice no fue descrita hasta 1953 gracias al trabajo de Rosalind Franklin, Watson y Crick. Hoy en día, el ADN se ha convertido en una herramienta cotidiana con su regular empleo que va desde la detección temprana de enfermedades y pruebas de paternidad, hasta su aplicación en asuntos forenses (resolución de crímines). Esto nos lleva a preguntas fascinates: ¿qué es el ADN? ¿cómo leemos la información que resguarda? y, más importante aún, ¿cómo podemos interpretarla?

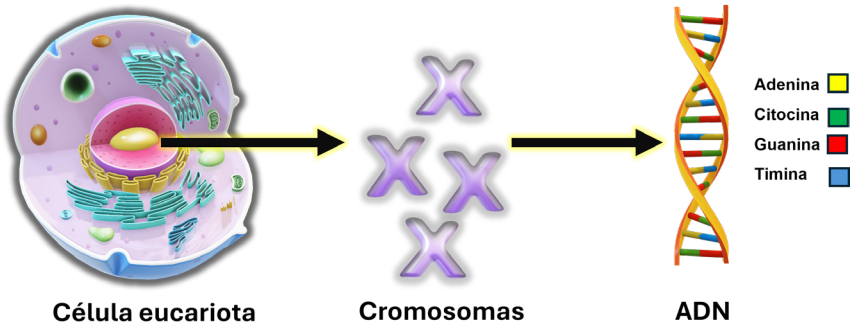

Figura 1. *La célula y la estructura de doble hélice de su ADN. Diagrama de localización de los cromosomas y el ADN en el núcleo de una célula eucariota.*

El ADN está contenido en las células de todos los organismos vivos, desde los microscópicos como las bacterias y arqueas, hasta animales, plantas y hongos. El ADN contiene la información necesaria para el desarrollo y funcionamiento de los seres vivos [1]. Es una molécula compleja (o "macromolécula"), cuya composición está basada en cuatro moléculas referidas como "nucleótidos": adenina (A), citosina (C), guanina (G) y timina (T); estos se emparejan de manera complementaria: la adenina se une con timina (A–T), y la guanina con citosina (G–C), formando una estructura de doble hélice (Figura 1). El orden específico de los nucleótidos (secuencia) determina nuestros genes, las instrucciones para construir y mantener a una célula viva, y el conjunto de todos ellos conforma nuestro genoma. Para darnos una idea, dos genes de importancia mayor en los seres humanos son "BRCA1" (por sus siglas en inglés: "*breast cancer gene* 1") y "BRCA2". Ambos están involucrados en la producción de proteínas que reparan errores en el ADN; en caso de daño, errores, o ausencia de alguno, aumenta la probabilidad de desarrollar diferentes tipos de cáncer, por ejemplo, cáncer de mama [2].

El ADN está dentro de las células, pero ¿cómo logramos sacarlo de ahí con el mínimo daño?

El ADN se encuentra en el núcleo de las células eucariotas o en el nucleoide en las células procariotas. Para poder recuperarlo, se suelen emplear diversos compuestos químicos que permiten la ruptura de la célula y liberación del ADN (Figura 2). Al romper la membrana, que está conformada por lípidos, fosfolípidos y proteínas, se liberan los organelos celulares, y moléculas varias; esta mezcla se conoce como debris o residuos celulares. Estos residuos necesitan ser eliminados para poder separar el material genético necesario. En este paso logramos "purificar" el ADN, recuperándolo con la menor cantidad de contaminantes posibles.

Comercialmente se pueden adquirir *kits* (conjunto de reactivos y material de plástico de uso en biología molecular) que permiten la extracción semi-automatizada del ADN. La elección de éste depende del tipo de material biológico inicial que se está manejando; por ejemplo, para extraer ADN de células (o una muestra) de cabello o tejidos se puede usar el "Kit de extracción de tejido y cabello" de la empresa Promega. Por su parte, para la extracción de ADN de células bacterianas se puede usar el kit para ADN genómico bacteriano "GenElute" de la empresa Sigma-Aldrich.

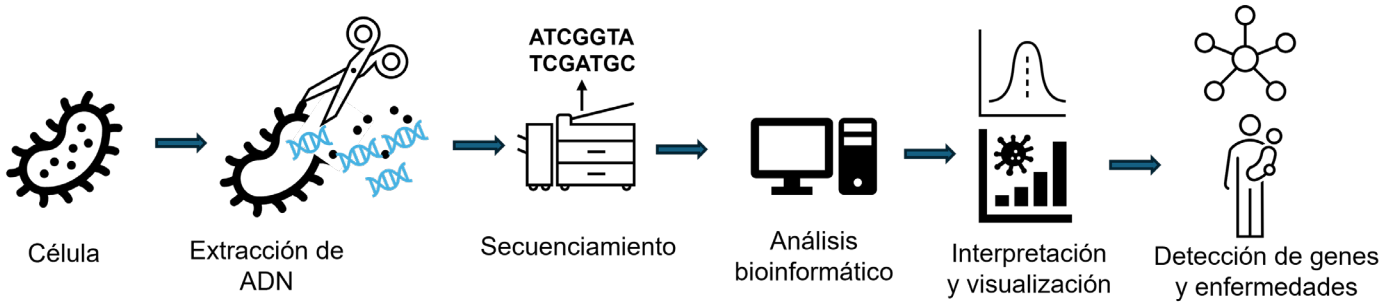

Figura 2. *Flujo general del procesamiento para análisis de ADN: desde una célula (o cultivo), hasta su análisis e interpretación. El diagrama simplifica cómo a apartir de una célula (o tejido) logramos extraer el DNA, secuenciarlo, analizarlo e interpretarlo.*

Leyendo el ADN ¿Qué nos dicen los "nucleótidos"?

Tras obtener el material genético de un organismo, o de una muestra biológica (tejido, cabello, de un hospedero o del ambiente), el siguiente paso es procesarlo para conocer el orden (o secuencia) exacto de los nucleótidos. Este proceso es conocido como secuenciación y se lleva a cabo en equipos conocidos como secuenciadores, quienes emplean un proceso similar al de la replicación biológica del ADN [3].

En la actualidad, compañías como Illumina®, Oxford Nanopore™ Technologies, y Pacific Biosciences® (PacBio) dominan el mercado global de tecnologías de secuenciación. La secuenciación es un proceso que varía dependiendo de la plataforma y tecnología empleados. Por ejemplo, una de las razones para seleccionar una plataforma es el tamaño de gen objetivo o la aplicación o fin del porqué se está secuenciando (por ej. la secuenciación del genoma completo de un microorganismo, o la identificación taxonómica de un microorganismo o grupo microbiano mediante uno o más genes).

La información genómica está formada desde decenas a millones de nucleótidos que se almacenan en archivos de texto plano denominados archivos "FASTA" o "FASTQ" [4]. Esta información no puede ser leída con herramientas computacionales comunes. Son archivos con gran tamaño regularmente, que van desde los megabytes (MB) hasta los gigabytes (un GB es aprox. 1024 MB) por unidad, y dependiendo su tamaño, puede contener una cantidad enorme de información que suele congelar herramientas como Microsoft© Word o el block de notas en una computadora personal. Para poder acceder, leer y traducir las secuencias de ADN contenida en estos archivos requerimos programas especializados; ¡aquí da inicio parte del área de la "bioinformática"!

De letras (nucleótidos) a funciones y taxones con ayuda de la bioinformática

A través de la bioinformática logramos traducir los patrones del ADN (ATGC---TGAT--GGCAT..., etc), o secuencias de nucleótidos (A, T, G, C), a información entendible por el humano. En este paso podemos identificar una secuencia y averiguar que función cumple o qué gen es, por ejemplo, puede ser una región que codifica una enzima o proteína. Este proceso es conocido como "anotación"; si identificamos uno o más genes funcionales, le decimos "anotación funcional". Si anotamos o identificamos un grupo taxonómico, le llamamos "anotación taxonómica". En la actualidad, existen diferentes programas (y metodologías) bioinformáticas para analizar secuencias de ADN. Herramientas como SPAdes, CANU, Kraken2, Prodigal, y RStudio son algunos ejemplos de software que se emplean

para analizar ambos tipos de anotación. En este sentido, es importante visualizar que estas herramientas se emplean sobre datos de ADN ya secuenciados, es decir, un paso previo a la interpretación biológica final.

Para lograr la anotación, las secuencias son comparadas contra otras ya conocidas y que están almacenadas en bases de datos públicas (ver Tabla 1); este proceso, que se repite con cada una de nuestras miles (hasta millones) de secuencias, es conocido como "alineamiento" y se basa en la búsqueda de homologías (similitud) entre las secuencias ya conocidas en las bases de datos y nuestra secuencia desconocida o de interés. El alineamiento de un número grande de secuencias contra una base de datos de referencia, que también suele ser grande, es un proceso exhaustivo que requiere software especializado y gran esfuerzo computacional, es decir, equipo de cómputo con grandes niveles de procesamiento (memoria RAM y número de procesadores) y almacenamiento.

Mediante alineamientos múltiples logramos entonces identificar casos especiales. Por ejemplo, para identificar la secuencia de los 23 pares de cromosomas humanos (proyecto conocido como "Genoma humano", en el año 2003) se necesitó de cerca de 13 años para su éxito; este dio inició en 1990 y finalizó en 2003. Para lograr su ensamble y anotación, se necesitó de supercomputadoras con más de 500 procesadores, y de forma global, el proyecto tuvo un costo aproximado de $2,700 millones de dólares [5].

Actualmente, las tecnologías de secuenciación han reducido su costo substancialmente que, a su vez, ha permitido ampliar las áreas de estudio en biología molecular y (bio) medicina. Además, dichas tecnologías han mejorado en las últimas décadas. Como ejemplo, en los años 1990 se empleaba como única tecnología la secuenciación conocida como "Sanger". Ésta fue substituida por la secuenciación de nueva generación ("NGS", por sus siglas en inglés: *next generation sequencing*) como la "pirosecuenciación" (454 *Life Sciences sequencing*) en 2005 y posteriormente por Oxford Nanopore en 2014 [6].

Tabla 1. *Bases de datos publicas empleados en el análisis de secuencias de ADN.*

Base de datos	País	Tipo de datos biológicos	Página web
GenBank (NCBI)	EE.UU.	ADN, ARN y proteínas	https://www.ncbi.nlm.nih.gov/genbank/
UniProt (EMBL-EBI)	Suiza	Secuencias y funciones de proteínas	https://www.ebi.ac.uk/uniprot/index
ClinVar (NCBI)	EE.UU.	Variantes genéticas y su relación con enfermedades	https://www.ncbi.nlm.nih.gov/clinvar/
dbSNP (NCBI)	EE.UU.	Polimorfismos de un sólo nucleótido (SNPs)	https://www.ncbi.nlm.nih.gov/snp/
GISAID	Internacional	Datos genómicos de virus de la influenza	https://gisaid.org/

Las bases de datos de referencia usadas en los análisis bioinformáticos contienen la información de los genes (o genomas completos) de un número grande de organismos (incluyendo bacterias, animales, plantas, hongos, eucariontes microscópicos, etc.) que los científicos alrededor del mundo han estudiado, secuenciado y publicado hasta la actualidad. En caso de que las secuencias que estamos estudiando correspondan a un nuevo organismo, como lo fue el virus causante de COVID-19, estas se comparan con las variantes más cercanas y se propone un nuevo organismo (o virus, como fue el caso). Para identificar el virus causante de COVID-19, primero se obtuvieron muestras biológicas de personas con síntomas de la enfermedad (p. ej. exudados nasofaríngeos) para así recuperar el material genético viral para posterior secuenciación. En algunos casos, el virus fue previamente cul-

tivado (con el fin de enriquecer la carga viral) en distintas líneas celulares para su posterior análisis [7]. Al comparar, mediante alineamiento con variantes de virus cercanas y virus ya conocidos como el SARS-CoV-1 (cuyas siglas en inglés se traducen como "síndrome respiratorio agudo grave"), se detectó que era una nueva variante y se nombró SARS-CoV-2 [8]. Los genomas de los virus de SARS-CoV-2 se publicaron en bases de datos públicas como GISAID o GenBank (Cuadro 1).

Conclusión: unión de la biología molecular y la bioinformática

La bioinformática ha permitido la identificación y estudio de una gran variedad de microorganismos, genes y enzimas de interés para la humanidad. Esta disciplina está en constante crecimiento con la creación de nuevos programas y algoritmos para mejorar el proceso de análisis. Se espera que, en un futuro próximo, a través del secuenciamiento y análisis del ADN humano, se logre personalizar tratamientos médicos y la posible prevención de enfermedades. La adición, y actualización, de herramientas con inteligencia artificial (p. ej. AlphaFold-3) permitiría el modelado y síntesis de nuevas proteínas y fármacos *in silico*.

Referencias

[1] Bencurova, E., Akash, A., Dobson, R. C. J., & Dandekar, T. (2023). DNA storage—From natural biology to synthetic biology. *Computational and Structural Biotechnology Journal*, 21, 1227–1235. https://doi.org/10.1016/j.csbj.2023.01.045

[2] Oh, M., McBride, A., Yun, S., Bhattacharjee, S., Slack, M., Martin, J. R., Jeter, J., & Abraham, I. (2018). BRCA1 and BRCA2 Gene Mutations and Colorectal Cancer Risk: Systematic Review and Meta-analysis. *JNCI: Journal of the National Cancer Institute*, 110(11), 1178–1189. https://doi.org/10.1093/jnci/djy148

[3] Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and Protein Extraction: The Past and The Present. *BioMed Research International*, 2009(1), 574398. https://doi.org/10.1155/2009/574398

[4] Pucker, B., Schilbert, H. M., & Schumacher, S. F. (2019). Integrating Molecular Biology and Bioinformatics Education. *Journal of Integrative Bioinformatics*, 16(3). https://doi.org/10.1515/jib-2019-0005

[5] Hood, L., & Rowen, L. (2013). The Human Genome Project: Big science transforms biology and medicine. *Genome Medicine*, 5(9), 79. https://doi.org/10.1186/gm483

[6] Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., & Malonia, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, 12(7). https://doi.org/10.3390/biology12070997

[7] Wurtz, N., Penant, G., Jardot, P., Duclos, N., & La Scola, B. (2021). Culture of SARS-CoV-2 in a panel of laboratory cell lines, permissivity, and differences in growth profile. *European Journal of Clinical Microbiology & Infectious Diseases*, 40(3), 477–484. https://doi.org/10.1007/s10096-020-04106-0

[8] Zhu Na, Zhang Dingyu, Wang Wenling, Li Xingwang, Yang Bo, Song Jingdong, Zhao Xiang,et al. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*, 382(8), 727–733. https://doi.org/10.1056/NEJMoa2001017