

Caracterización de las tecnologías de secuenciación genética de segunda y tercera generación

Characterization of second- and third-generation genomic sequencing technologies

Dorian Rojas-Villalta

Instituto Tecnológico de Costa Rica, Costa Rica

rojasvillaltadorian@gmail.com

 <https://orcid.org/0000-0002-2939-1485>

Daniela Benavides-Villegas

Instituto Tecnológico de Costa Rica, Costa Rica

elabenavides.v@gmail.com

 <https://orcid.org/0000-0002-4649-4101>

Belén Angulo-Hidalgo

Instituto Tecnológico de Costa Rica, Costa Rica

mangulobelen08@gmail.com

 <https://orcid.org/0000-0002-8421-9514>

Luis Muñoz-Solórzano

Instituto Tecnológico de Costa Rica, Costa Rica

luchomusol@gmail.com

 <https://orcid.org/0000-0002-6960-377X>

Chiara Consumi-Tubito

Instituto Tecnológico de Costa Rica, Costa Rica

chiaraconsumi@gmail.com

 <https://orcid.org/0000-0002-3256-1975>

Recepción: 22 Febrero 2023

Aprobación: 27 Junio 2023



Acceso abierto diamante

Resumen

El avance en la secuenciación de material genético ha favorecido el estudio de nuevas áreas de investigación relacionadas a la genómica, transcriptómica y demás ciencias -ómicas. Esto creó la necesidad de desarrollar nuevas tecnologías más accesibles y eficientes, como las tecnologías de segunda y tercera generación. Por ello, esta revisión caracterizó las principales metodologías de secuenciación de segunda y tercera generación. Las tecnologías de segunda generación se ven representadas por la secuenciación por síntesis fragmentos cortos con alta calidad y precisión. En la tercera generación se especializan en lecturas largas secuenciadas por nanoporos y secuenciación de una sola molécula en tiempo real. Las empresas Illumina y PacBio poseen una mayor calidad y precisión en las lecturas, en comparación a Oxford Nanopore Technologies (ONT). Sin embargo, las lecturas de ONT solucionan problemas del ensamblaje genómico propios de lecturas cortas. Todas las tecnologías analizadas poseen secuenciadores para proyectos de producción a escala; Illumina y ONT también poseen secuenciadores sobremesa, y solamente la última presenta dispositivos portátiles. Con respecto a la accesibilidad, los países centroamericanos poseen poca cobertura por parte de estas empresas, siendo Illumina la única con una distribuidora radicada en la región. Cada tecnología de secuenciación posee ventajas y desventajas asociadas, lo cual impulsa a su uso en conjunto, lo que permite análisis más convenientes y completos. Por lo cual el desarrollo de

nuevos programas bioinformáticos para la mejora del análisis corresponde a consideraciones que deben ser abordadas con la intención de fortalecer el uso de las técnicas de secuenciación.

Palabras clave: Illumina, Oxford Nanopore, PacBio, plataformas de secuenciación, bioinformática.

Abstract

The progress in genetic material sequencing enhanced diverse research opportunities related to genomics, transcriptomics and other -omics areas. A need in the development of more accessible and efficient techniques resulted in the second- (NGS) and third-generation technologies (TGS). In this paper we aimed to characterize the main sequencing methods of second- and third-generation. NGS technology is based on the sequencing by biosynthesis of short fragments with high quality and precision, while TGS are specialized in long fragment sequencing by nanopores and single-molecule, real-time technologies. Illumina and PacBio have more precise and high-quality data, compared to Oxford Nanopore Technologies (ONT). However, nanopores long reads address errors in genome assembly, commonly associated with small lectures. All analyzed technologies have production-scale sequencers, but only ONT and Illumina developed benchtop products. ONT also presents portable sequencers for small projects. Latinoamerican countries have less coverage and technical support of sequencing technologies, and only Illumina is sited in this region. Each sequencing technology have advantages and limitations regarding their techniques and their coupled use is promoted to make a deeper analysis of the data, therefore the development of new bioinformatics programs remain as one of the principal obstacles to overcome to enhance the application of these technologies.

Keywords: Illumina, Oxford Nanopore, PacBio, sequencing platforms, bioinformatics.

Introducción

Posterior al desarrollo de la metodología de secuenciación de Frederick Sanger [1], se ha incentivado el crecimiento de la genómica y demás ciencias -ómicas. La secuenciación de Sanger consiste en el uso de didesoxinucleótidos (ddNTPs); análogos químicos de los nucleótidos sin el grupo hidroxilo en el extremo 3', lo cual inhibe la extensión de la cadena de ADN en la replicación [2]. Esta técnica emplea una combinación de cada ddNTP individual marcado radioactivamente junto a nucleótidos tradicionales, permitiendo el alargamiento de la cadena de ADN hasta la incorporación de un ddNTP; los resultados son analizados por medio de electroforesis capilar para determinar la secuencia según las distancias de corrida [1].

En 1990, la técnica de Sanger fue automatizada y utilizada para la secuenciación del genoma humano, como parte del Proyecto del Genoma Humano (HGP, por sus siglas en inglés), el cual se finalizó en el 2001 [3]. Los resultados obtenidos favorecieron a la identificación de alelos deficientes asociados a enfermedades como la fibrosis quística, lo que permitió enfocar investigaciones al desarrollo de su tratamiento [4]. En la actualidad, el genoma humano sigue siendo actualizado agregando zonas de heterocromatina y otras regiones cromosómicas no representadas en el primer genoma de referencia, desbloqueando nuevas secuencias para estudios funcionales [5].

En este sentido, el desarrollo de las diferentes ciencias -ómicas ha tenido múltiples aplicaciones como la caracterización y entendimiento de las rutas metabólicas para la manipulación y mejoramiento genético, planteamiento de plataformas industriales de biomanufactura, detección de predisposiciones genéticas, procesos fisiopatológicos y epidemiología de enfermedades, entre otras [6]–[8]. Dada su aplicabilidad en diversos aspectos, el desarrollo de nuevas tecnologías de menor costo y mayor eficiencia ha sido uno de los principales enfoques recientes, llevando a las siguientes generaciones en técnicas de secuenciación [9].

La primera generación consiste en la secuenciación de Sanger y Maxam Gilbert; la segunda generación o next generation sequencing (NGS) engloba técnicas de empresas como Illumina Inc., principalmente. Finalmente, la secuenciación de tercera generación se relaciona con los sistemas de empresas como Pacific Biosciences (PacBio) y Oxford NanoPore Technologies Inc. (ONT) [9]. En esta revisión se analiza y compara las principales metodologías de secuenciación de segunda y tercera generación, describiendo su fundamento bioquímico, aspectos de lectura, dispositivos y accesibilidad.

Tecnologías de Secuenciación

La secuenciación genética consiste en la determinación de la secuencia de nucleótidos de secciones de material genético [9]. Cada tecnología de secuenciación bajo su propio mecanismo bioquímico da lugar a lecturas de información (reads), que reflejan los resultados de la secuenciación [10]. Estos resultados consisten en un conjunto complejo de múltiples caracteres que son transformados y analizados posteriormente mediante programas bioinformáticos [11]. Asimismo, la información generada se interpreta de diferente manera según el objetivo del estudio, ya sea en metagenómica, transcriptómica o secuenciación de genomas completos.

Se han desarrollado gran variedad de interfaces para el análisis, organización y visualización de los resultados de secuenciación. Programas como NanoPlot y FastQC generan resúmenes y representaciones gráficas de los datos. NanoPlot se emplea en el procesamiento de secuencias de lectura larga [12], mientras que FastQC se utiliza para lecturas cortas [13].

Asimismo, los resultados de una secuenciación conllevan un proceso de recorte genómico (genomic trimming); el cual consiste en una eliminación de lecturas de baja calidad o longitud indeseada [14]. Para ensamblar los genomas se emplean plataformas como Unicycler [15], que realizan el acomodamiento de las lecturas cortas y largas para la generación de un archivo con el genoma completo en formato fasta; este ensamblaje se puede realizar de novo o mediante un genoma de referencia. Una vez obtenido el genoma, este

puede analizarse de diversas formas dependiendo de la finalidad del estudio, como alineamiento con otros genomas, estudiar relaciones filogenéticas y realizar anotaciones sobre los distintos genes presentes (genome mining) [16], [17].

Fundamento bioquímico de las técnicas de secuenciación

Segunda generación

La secuenciación de segunda generación consiste en plataformas que producen gran cantidad de lecturas cortas (25-400 bp) de secuencias de ADN [18]. En esta generación destaca la empresa Illumina Inc., actualmente dominante en el mercado. Asimismo, otras empresas como Roche 454, AB SOLiD y Ion Torrent fueron desarrolladas en torno a este modelo; sin embargo, fueron desplazadas y/o descontinuadas por la innovación de la tecnología de Illumina.

El concepto detrás de la tecnología NGS es similar a la técnicas de secuenciación de Sanger, previamente descrita [19]. La tecnología de Illumina utiliza el método de secuenciación por síntesis (sequencing by synthesis, SBS), permitiendo la lectura paralela de millones de fragmentos por medio de la detección de las bases individuales a medida que se incorporan a las cadenas de ADN en crecimiento [19], [20]. La ADN polimerasa cataliza la incorporación de desoxirribonucleótidos trifosfato (dNTP) marcados con fluorescencia a una hebra molde de ADN durante ciclos secuenciales de síntesis. Durante cada ciclo los nucleótidos se identifican mediante excitación del fluoróforo [19], creando una imagen de un terminador reversible marcado fluorescentemente a medida que se añade cada dNTP, y luego se separa para el ingreso de la siguiente base. Al finalizar, se obtiene una secuenciación base por base con datos de alta precisión y calidad, para una gran variedad de usos, como el ensamblaje de novo, la resecuenciación, la identificación de mutaciones de interés clínico y la metagenómica [19], [21].

Tercera generación

Las técnicas de secuenciación masiva de ácidos nucleicos de segunda generación presentan debilidades, principalmente en cuanto a la capacidad de generar lecturas extensas (>1 kb). Para satisfacer estas necesidades se desarrollaron las tecnologías de tercera generación (third-generation sequencing, TGS), las cuales permiten extensiones de hasta 2.3 Mb y no requieren de una amplificación preliminar, disminuyendo la incidencia de pérdidas de marcas epigenéticas [22].

Existen dos proveedores altamente utilizados para la secuenciación de lecturas largas: Pacific Biosciences y Oxford Nanopore Technologies Inc. PacBio tiene la capacidad de generar lecturas HiFi (High Fidelity) con una resolución alta por base. Por su parte, ONT puede generar lecturas con una extensión de hasta 2 Mb [23]. Además, existen otras plataformas emergentes de TGS, como Molecu y 10X Genomics, basadas en el ensamblaje de lecturas cortas para generar lecturas largas sintéticas [24]. En este artículo se exploran las tecnologías de PacBio y ONT, dado que son las más establecidas y comercializadas.

La tecnología de Oxford Nanopore Technologies fue lanzada en el 2014 y utiliza la secuenciación por nanoporos, los cuales emplean el perfil eléctrico de cada nucleótido de ADN y ARN para su identificación [25]. Estos canales permiten el paso de una molécula a la vez, detectando cambios en la corriente eléctrica característicos de cada nucleótido [22]. El fundamento de la secuenciación de ONT permite la lectura directa de secuencias de ARN, siendo esta la única tecnología con dicha capacidad [26], [27].

Por otro lado, la empresa Pacific Biosciences es líder de la tercera generación, con la tecnología de secuenciación en tiempo real de una sola molécula (Single-Molecule, Real-Time, SMRT) lanzada en el 2011 [28]. Esta produce lecturas largas de 3000-15000 bp, se utiliza ampliamente para la elaboración de perfiles filogenéticos de microbiomas por medio de una clasificación taxonómica a nivel de género; así bien, también es

utilizada en estudios genómicos, transcriptómicos, epigenómicos, metagenómicos y de ensamblaje de novo [24], [29].

La estrategia de secuenciación de PacBio consiste en una modificación de la SBS de Illumina, en la cual se logra capturar una sola molécula de ADN que ingresa a celdas de flujo con tecnología de guía de onda de modo cero (zero-mode waveguide, ZMW) [22]. La ZMW es un dispositivo focalizador de la luz hacia un punto específico, lo que permite el registro preciso de la señal lumínica emitida por un fluoróforo unido a un nucleótido específico [22], [30].

Comparación entre tecnologías de segunda y tercera generación

Características de la secuenciación

Las tecnologías de secuenciación de Illumina se caracterizan por la lectura de fragmentos cortos de ADN [31]. Este tipo de secuenciación se recomienda para análisis de variantes clínicas, ya que es altamente preciso y eficiente, permitiendo la identificación de variaciones genéticas de menor frecuencia [32], [33]. Su rendimiento se considera más alto que las tecnologías de tercera generación, y la preparación de la muestra, más compleja al requerir una amplificación por PCR [27], [34], [35]. A pesar de una baja tasa de error (0.01%), cuenta con algunos sesgos y limitaciones. En primer lugar, la lectura de secuencias cortas se ve limitada ante las repeticiones en tándem del ADN. Esto puede causar errores en la lectura en forma de gaps y complicaciones en el ensamblaje [36]. Se han reportado errores en sustituciones; es decir, el reemplazo de una base por otra no correspondiente, y errores menos frecuentes (p. ej. inserciones, deleciones, errores específicos) a ciertos motifs e interferencias entre muestras multiplexadas [21], [37]. Algunos de estos se adjudican a sesgos por el contexto de la secuencia, así como procesos químicos relacionados a la tecnología de secuenciación y la preparación de las librerías genómicas [31], [38].

Seguidamente, Oxford Nanopore Technologies Inc. ha desarrollado tecnologías de secuenciación para longitudes de fragmentos especialmente largos, siendo esta característica una de sus principales ventajas. La secuenciación de fragmentos largos brinda ventajas al momento de realizar el ensamblaje genómico, como una mayor repetitividad de fragmentos, cobertura y probabilidad para el cierre del genoma, en comparación con ensamblajes de secuencia cortas [39], [40]. Asimismo, ONT es la única tecnología en el mercado que permite la secuenciación de ARN de forma directa. Dicho estilo de secuenciación favorece el estudio de organismos cuyo material genético se encuentra basado en ARN, como algunos virus, y colabora con el análisis de estructuras genéticas relacionadas, como los ARN largos no codificantes y retrotransposones [41], [42]. Además, es destacable la mejora en el análisis de organismos cuyo genoma completo no se encuentra reportado o ensamblado en una sola secuencia (contig) [43]. Por ejemplo, una nueva actualización del genoma humano con regiones que no habían sido previamente secuenciadas fue reportado recientemente con ayuda de las tecnologías de secuenciación de ONT [5], y el ensamblaje de cromosomas bacterianos y de microorganismos en un único contig ha sido favorecido de la misma forma [44]–[46].

Sin embargo, una de las principales desventajas de las tecnologías de secuenciación de fragmentos largos se relaciona a la calidad de las lecturas. La amplia longitud de las secuencias leídas aumenta la tasa de error, tendencia mantenida a lo largo de las secuenciaciones para ADN y ARN de forma directa e indirecta [32], [35], [47]. En las tecnologías de ONT, la tasa de error se ha presentado con una media de <0.1% [25].

Finalmente, la tecnología de secuenciación de PacBio permite la medición en tiempo real de la incorporación de nucleótidos de acuerdo con señales de luz emitidas por el sistema [28], [30]. Este es utilizado para la lectura de cadenas largas de ADN y ARN indirectamente, sin que se comprometa la precisión. Por lo cual, su secuenciación es libre de error sistemático al contar con más de un 99.999% de precisión por consenso [28]. Asimismo, la secuenciación cuenta con bajo sesgo por contexto de la muestra, un mapeo preciso de las lecturas y cobertura uniforme al no haber desviaciones por el contenido de bases GC. Una característica

distintiva de este tipo de secuenciación es su análisis epigenómico sin la necesidad de amplificar por PCR; el sistema es capaz de detectar directamente modificaciones a las bases durante la secuenciación, especialmente la metilación 5mC [30]. Esto sirve para eliminar la necesidad de alguna modificación química para realizar esta clase de estudios, reduciendo costos, tiempo y aumentando la eficiencia experimental [30]. Adicionalmente, PacBio ha sido empleada para la secuenciación a partir de una célula (single-cell sequencing), técnica que favorece el estudio de haplotipos y mutaciones de ultra baja frecuencia posiblemente asociadas a patologías [48]–[50].

Dispositivos

Illumina cuenta con una línea de secuenciadores de sobremesa (benchtop sequencers) y otra de secuenciadores de producción a escala (production-scale sequencers) [51]. Los secuenciadores sobremesa (cuadro 1) poseen la capacidad de realizar secuenciaciones de genoma completos, secuenciación dirigida, ARN y análisis metagenómicos de 16S. Por su parte, la línea de secuenciadores de producción a escala (cuadro 2) poseen las mismas funciones que los secuenciadores benchtop con un mayor poder de procesamiento. Los precios de los distintos dispositivos son protegidos por la compañía. Sin embargo, las ventas de servicios para MiSeq se estiman en \$185.71 por muestra al secuenciar 24 muestras multiplexadas, tomando en cuenta los costos de preparación de las librerías genómicas [35]. Una característica distintiva de estos equipos es tener un rendimiento masivo con un costo de miles de dólares para una corrida sencilla, resultando en un costo por muestra bajo, pero costo por corrida alto [27].

Cuadro 1

Dispositivos de secuenciación de sobremesa benchtop de las tecnologías de segunda y tercera generación 51–54

	Illumina			ONT		
	iSeq 100	MiniSeq	MiSeq	NextSeq 550	MinION	GridION
Portabilidad	No	No	No	No	Sí	No
Tiempo de corrida (h)	9.5-19	4-24	4-55	12-30	~72	~72
Longitud máxima (bp)	150	150	300	150	10000	10000
Lecturas totales (mill)	4	25	25	400	110	110*
Uso en DIV	No	No	Sí**	Sí**	No	No
Costo	No disp.	No disp.	\$128000	No disp.	\$1000-4900	\$50000

DIV = diagnóstico in vitro. No disp. = No disponible. *Por celda, GridION permite hasta 5 celdas por ejecución. **Las versiones Dx de los dispositivos. Pacific Bioscience no posee ningún dispositivo de sobremesa comercializado.

Con respecto a los distintos dispositivos de ONT, en el mercado hay disponibles cuatro equipos principales, entre estos se pueden mencionar los secuenciadores de sobremesa: MiniION y GridION (cuadro 1), y un equipo de producción a escala: PromethION (cuadro 2) [53], [54]. Dichos equipos poseen la capacidad de secuenciar genomas y transcriptomas completos de forma directa e indirecta y secuenciaciones dirigidas; así bien, el equipo a escala permite realizar análisis poblacionales y metagenómicos. El MinION y GridION son tecnologías que emplean celdas de flujo específicas con 512 nanoporos; mientras que, el PromethION utiliza celdas con 2675 nanoporos, brindándole 6 veces mayor capacidad de procesamiento simultáneo. Las celdas de 512 nanoporos poseen un valor de \$900, y las de 2675 nanoporos presentan un costo de \$2400 [54]. El cuarto equipo desarrollado corresponde a un adaptador para MinION y GridION llamado Flonge, el cual permite la

realización de secuenciaciones pequeñas con celdas de flujo de un solo uso; dichas celdas poseen la capacidad de secuenciación de ~24 horas en condiciones óptimas. Este dispositivo presenta un valor de \$1460, donde se incluyen adicionalmente 12 celdas de flujo. Así bien, sus celdas pueden ser compradas de forma independiente en paquetes de nueve y cuentan con un precio de \$810 [53], [54].

Cuadro 2

Dispositivos de secuenciación de producción a escala production scale de las tecnologías de segunda y tercera generación 51 53–55

	ILLUMINA	ONT	PACBIO
	NextSeq 1000 & 2000 NovaSeq X	NovaSeq 6000	PromethION 2, 24 y 48
Tiempo de corrida (h)	11-48	13-48	13-44
Longitud máxima (bp)	150	150	250
Lecturas totales (bill)	1.2	1.6-52	20
Output	360 Gb	165-16000 Gb	6 Tb – 20 B
Costo	No disp.	No disp.	No disp.

No disp. = No disponible. *Por celda, existen tres versiones de PromethION (2, 24 y 48), que difieren en la cantidad de celdas por ejecución (dos, 24 o 48 celdas).

En relación con PacBio, esta empresa posee una única línea desarrollada de secuenciadores a escala llamada Sequel, cuyos los principales dispositivos utilizados son: Sequel II y IIe (cuadro 2) [55]. Estos permiten la secuenciación de todo tipo de genomas, transcriptomas y epigenomas. Un servicio de la secuenciación de 30 horas por medio de Sequel II presenta un valor mayor a los \$3000 por muestra; esto según las cotizaciones generadas por la Universidad de Washington (University of Washington) [56].

Disponibilidad de servicios

Illumina Inc. cuenta con oficinas comerciales en Estados Unidos, Australia, Alemania, China, Francia, Japón, América Latina, Corea, Holanda, Singapur y el Reino Unido; paralelamente, tiene dos centros de manufactura, investigación y desarrollo en California (Estados Unidos) y Woodlands (Singapur) [57]. Por su parte, Oxford Nanopore Technologies Inc. presenta sitios de contacto en Estados Unidos, Reino Unido, Singapur, China y Japón, así como una gran cantidad de distribuidoras a nivel mundial [58]. Sin embargo, solamente existe una distribuidora encargada de la cobertura de los países centroamericanos y ciertos países de América Latina (Interprise) [59]. Finalmente, PacBio posee diferentes sedes en Estados Unidos, Reino Unido, China, Singapur y Japón, con su principal ubicada en Menlo Park (California, Estados Unidos) [60]. De la misma forma posee múltiples distribuidores ubicados a lo largo del pacífico de Asia, Europa, Medio Oriente y América. En este sentido, ninguna de las sedes actuales cubre a los países centroamericanos, y el único país latinoamericano con distribuidora para esta empresa es México [61].

Consideraciones futuras

La creación y actualización de programas bioinformáticos, añadiendo funciones prácticas para el estudio de las secuencias y evitando la polarización del tipo de datos (reads largos o cortos) para el análisis híbrido,

representa otra área de desarrollo. Por ejemplo, Next Generation Sequencing Eclipse Plugin (NGSEP) agrega la función de “Estadísticas sobre poblaciones genotipadas”, útil en el análisis de la estructura poblacional, construcción de mapas genéticos, mapeo genético de rasgos complejos y predicción de fenotipos para la selección genómica, así complementando las posibilidades de análisis genómico [61]. De la misma forma, los ensambladores híbridos de Unicycler [15] y SPAdes [62] son tecnologías emergentes que permiten la utilización de datos de distintas generaciones de secuenciación, favoreciendo a los ensamblajes en un solo contig de mayor calidad y completitud [63].

Asimismo, es relevante la expansión del uso de estas tecnologías por medio del aumento del soporte técnicos y la disponibilidad de servicios, especialmente a países latinoamericanos. Esto considerando los beneficios asociados en espacios científicos, clínicos y demás áreas [64], [65]. Algunos obstáculos relacionados a la inequidad, el estado de desarrollo, la falta de cobertura de las empresas y la migración de profesionales, deben de ser abordados con la finalidad de permitir el fortalecimiento de dichas tecnologías [66]–[68].

Conclusiones

En conclusión, las tecnologías de secuenciación representan una herramienta importante para el desarrollo de la investigación, la industria y la clínica. Illumina, Oxford Nanopore Technologies y PacBio son las empresas de mayor auge a nivel mundial con tecnologías de secuenciación de segunda y tercera generación. Los fundamentos de estas secuenciaciones emplean la secuenciación por síntesis, nanoporos o una única molécula en tiempo real. Así bien, Illumina se concentra en lecturas de fragmentos cortos, mientras que ONT y PacBio, en lecturas de fragmentos largos. Cada una posee ciertas desventajas relacionadas, como baja calidad en las lecturas largas o complicaciones en el ensamblaje de lecturas cortas. Asimismo, estas empresas difieren en otros aspectos relacionados a los dispositivos disponibles, las posibles aplicaciones y su accesibilidad (cuadro 3).

Referencias

- [1] F. Sanger, S. Nicklen, and A. R. Coulson, “DNA sequencing with chain-terminating inhibitors,” *Proceedings of the National Academy of Sciences*, vol. 74, no. 12, pp. 5463–5467, Dec. 1977, doi: 10.1073/PNAS.74.12.5463.
- [2] C. R. Gade, M. Dixit, and N. K. Sharma, “Dideoxy nucleoside triphosphate (ddNTP) analogues: Synthesis and polymerase substrate activities of pyrrolidiny nucleoside triphosphates (prNTPs),” *Bioorg Med Chem*, vol. 24, no. 18, pp. 4016–4022, Sep. 2016, doi: 10.1016/J.BMC.2016.06.043.
- [3] J. Craig Venter et al., “The sequence of the human genome,” *Science* (1979), vol. 291, no. 5507, pp. 1304–1351, Feb. 2001, doi: 10.1126/SCIENCE.1058040/SUPPL_FILE/1058040S3-7_LARGE.JPEG.
- [4] R. A. Gibbs, “The Human Genome Project changed everything,” *Nat Rev Genet*, vol. 21, no. 10, pp. 575–576, Aug. 2020, doi: 10.1038/S41576-020-0275-3.
- [5] S. Nurk et al., “The complete sequence of a human genome,” *Science* (1979), vol. 376, no. 6588, pp. 44–53, Apr. 2022, doi: 10.1126/SCIENCE.ABJ6987.
- [6] B. Amer and E. E. K. Baidoo, “Omics-Driven Biotechnology for Industrial Applications,” *Front Bioeng Biotechnol*, vol. 9, p. 30, Feb. 2021, doi: 10.3389/FBIOE.2021.613307/BIBTEX.
- [7] R. Bhattacharjya, A. Tiwari, T. K. Marella, H. Bansal, and S. Srivastava, “New paradigm in diatom omics and genetic manipulation,” *Bioresour Technol*, vol. 325, p. 124708, Apr. 2021, doi: 10.1016/J.BIORTECH.2021.124708.
- [8] M. A. Wörheide, J. Krumsiek, G. Kastenmüller, and M. Arnold, “Multi-omics integration in biomedical research – A metabolomics-centric review,” *Anal Chim Acta*, vol. 1141, pp. 144–162, Jan. 2021, doi: 10.1016/J.ACA.2020.10.038.
- [9] J. M. Heather and B. Chain, “The sequence of sequencers: The history of sequencing DNA,” *Genomics*, vol. 107, no. 1, pp. 1–8, Jan. 2016, doi: 10.1016/J.YGENO.2015.11.003.
- [10] W. Huang, L. Li, J. R. Myers, and G. T. Marth, “ART: a next-generation sequencing read simulator,” *Bioinformatics*, vol. 28, no. 4, pp. 593–594, 2012.
- [11] S. Goodwin, J. D. McPherson, and R. McCombie, “Coming of age: ten years of next-generation sequencing technologies,” *Nat Rev Genet*, vol. 17, no. 6, pp. 333–351, 2016, doi: 10.1038/nrg.2016.49.
- [12] W. de Coster, S. D’hert, D. T. Schultz, M. Cruets, and C. van Broeckhoven, “NanoPack: visualizing and processing long-read sequencing data,” *Bioinformatics*, vol. 34, no. 15, pp. 2666–2669, 2018.
- [13] J. Brown, M. Pirrung, and L. A. McCue, “FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool,” *Bioinformatics*, vol. 33, no. 19, pp. 3137–3139, 2017.
- [14] S. F. Yang, C. W. Lu, C. te Yao, and C. M. Hung, “To Trim or Not to Trim: Effects of Read Trimming on the De Novo Genome Assembly of a Widespread East Asian Passerine, the Rufous-Capped Babbler (*Cyanoderma ruficeps* Blyth),” *Genes (Basel)*, vol. 10, no. 10, Oct. 2019, doi: 10.3390/GENES10100737.
- [15] R. R. Wick, L. M. Judd, C. L. Gorrie, and K. E. Holt, “Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads,” *PLoS Comput Biol*, vol. 13, no. 6, Jun. 2017, doi: 10.1371/JOURNAL.PCBI.1005595.
- [16] M. Alser et al., “Accelerating genome analysis: A primer on an ongoing journey,” *IEEE Micro*, vol. 40, no. 5, 2020.

- [17] M. T. Swain, I. J. Tsai, S. A. Assefa, C. Newbold, M. Berriman, and T. D. Otto, “A Post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs,” *Nat Protoc*, vol. 7, no. 7, p. 1260, Jul. 2012, doi: 10.1038/NPROT.2012.068.
- [18] M. Imelfort and D. Edwards, “De novo sequencing of plant genomes using second-generation technologies,” *Brief Bioinform*, vol. 10, no. 6, pp. 609–618, Nov. 2009, doi: 10.1093/BIB/BBP039.
- [19] Illumina, “An introduction to Next-Generation Sequencing Technology.” www.illumina.com/technology/next-generation-sequencing.html (accessed Oct. 03, 2022).
- [20] E. R. Mardis, “Next-Generation Sequencing Platforms,” *Annual Review of Analytical Chemistry*, vol. 6, pp. 287–303, Jun. 2013, doi: 10.1146/ANNUREV-ANCHEM-062012-092628.
- [21] M. Molnar and L. Ilie, “Correcting Illumina data,” *Brief Bioinform*, vol. 16, no. 4, pp. 588–599, Jul. 2015, doi: 10.1093/bib/bbu029.
- [22] J. M. Valderrama Martín, F. Ortigosa, and R. Cañas, “Métodos de secuenciación: tercera generación,” *Encuentros en la Biología*, vol. 13, no. 175, pp. 15–21, 2020.
- [23] D. Lang et al., “Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore,” *Gigascience*, vol. 9, no. 12, pp. 1–7, Nov. 2020, doi: 10.1093/GIGASCIENCE/GIAA123.
- [24] D. Buck et al., “Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis,” *F1000Res*, vol. 6, 2017, doi: 10.12688/F1000RESEARCH.10571.2.
- [25] C. Delahaye and J. Nicolas, “Sequencing DNA with nanopores: Troubles and biases,” *PLoS One*, vol. 16, no. 10, p. e0257521, Oct. 2021, doi: 10.1371/JOURNAL.PONE.0257521.
- [26] B. Lin, J. Hui, and H. Mao, “Nanopore Technology and Its Applications in Gene Sequencing,” *Biosensors 2021*, Vol. 11, Page 214, vol. 11, no. 7, p. 214, Jun. 2021, doi: 10.3390/BIOS11070214.
- [27] E. E. Schadt, S. Turner, and A. Kasarskis, “A window into third-generation sequencing,” *Hum Mol Genet*, vol. 19, no. R2, pp. R227–R240, Oct. 2010, doi: 10.1093/hmg/ddq416.
- [28] J. Eid et al., “Real-Time DNA Sequencing from Single Polymerase Molecules,” *Science* (1979), vol. 323, no. 5910, pp. 133–138, Jan. 2009, doi: 10.1126/science.1162986.
- [29] E. B. Fichot and R. S. Norman, “Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform,” *Microbiome*, vol. 1, no. 1, pp. 1–5, Mar. 2013, doi: 10.1186/2049-2618-1-10/FIGURES/4.
- [30] PacBio, “How HiFi sequencing works - PacBio.” <https://www.pacb.com/technology/hifi-sequencing/how-it-works/> (accessed Oct. 02, 2022).
- [31] M. Schirmer, R. D’Amore, U. Z. Ijaz, N. Hall, and C. Quince, “Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data,” *BMC Bioinformatics*, vol. 17, no. 1, p. 125, Dec. 2016, doi: 10.1186/s12859-016-0976-y.
- [32] D. Tshiabuila et al., “Comparison of SARS-CoV-2 sequencing using the ONT GridION and the Illumina MiSeq,” *BMC Genomics*, vol. 23, no. 1, pp. 1–17, Dec. 2022, doi: 10.1186/S12864-022-08541-5/FIGURES/8.
- [33] K. Thys, P. Verhasselt, J. Reumers, B. M. P. Verbist, B. Maes, and J. Aerssens, “Performance assessment of the Illumina massively parallel sequencing platform for deep sequencing analysis of viral minority variants,” *J Virol Methods*, vol. 221, pp. 29–38, 2015, doi: <https://doi.org/10.1016/j.jviromet.2015.04.022>.

- [34] S. S. T. Mak et al., “Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing,” *Gigascience*, vol. 6, no. 8, Jun. 2017, doi: 10.1093/gigascience/gix049.
- [35] A. Pecman et al., “Systematic Comparison of Nanopore and Illumina Sequencing for the Detection of Plant Viruses and Viroids Using Total RNA Sequencing Approach,” *Front Microbiol*, vol. 13, 2022, doi: 10.3389/fmicb.2022.883921.
- [36] A. P. Heikema et al., “Comparison of Illumina versus Nanopore 16S rRNA Gene Sequencing of the Human Nasal Microbiota,” *Genes (Basel)*, vol. 11, no. 9, 2020, doi: 10.3390/genes11091105.
- [37] E. S. Wright and K. H. Vetsigian, “Quality filtering of Illumina index reads mitigates sample cross-talk,” *BMC Genomics*, vol. 17, no. 1, p. 876, 2016, doi: 10.1186/s12864-016-3217-x.
- [38] A. Cacho, E. Smirnova, S. Huzurbazar, and X. Cui, “A Comparison of Base-calling Algorithms for Illumina Sequencing Technology,” *Brief Bioinform*, vol. 17, no. 5, pp. 786–795, Sep. 2016, doi: 10.1093/bib/bbv088.
- [39] S. Goldstein, L. Beka, J. Graf, and J. L. Klassen, “Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing,” *BMC Genomics*, vol. 20, no. 1, pp. 1–17, Jan. 2019, doi: 10.1186/s12864-018-5381-7/FIGURES/7.
- [40] M. A. Madoui et al., “Genome assembly using Nanopore-guided long and error-free DNA reads,” *BMC Genomics*, vol. 16, no. 1, pp. 1–11, Dec. 2015, doi: 10.1186/s12864-015-1519-Z/TABLES/3.
- [41] I. Kirov et al., “Nanopore RNA Sequencing Revealed Long Non-Coding and LTR Retrotransposon-Related RNAs Expressed at Early Stages of Triticale SEED Development,” *Plants*, vol. 9, no. 12, p. 1794, Dec. 2020, doi: 10.3390/PLANTS9121794.
- [42] T. Wongsurawat et al., “Rapid sequencing of multiple RNA viruses in their native form,” *Front Microbiol*, vol. 10, no. FEB, p. 260, 2019, doi: 10.3389/FMICB.2019.00260/BIBTEX.
- [43] A. Bayega et al., “Nanopore long-read RNA-seq and absolute quantification delineate transcription dynamics in early embryo development of an insect pest,” *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/S41598-021-86753-7.
- [44] G. Wang et al., “A strategy for complete telomere-to-telomere assembly of ciliate macronuclear genome using ultra-high coverage Nanopore data,” *Comput Struct Biotechnol J*, vol. 19, pp. 1928–1932, Jan. 2021, doi: 10.1016/J.CSB.2021.04.007.
- [45] Y. Wang, X. Cai, and Y. Mao, “The first complete genome sequence of species *Shewanella decolorationis*, from a bioremediation competent strain Ni1-3,” *G3 Genes|Genomes|Genetics*, vol. 11, no. 10, Sep. 2021, doi: 10.1093/G3JOURNAL/JKAB261.
- [46] E. L. Moss, D. G. Maghini, and A. S. Bhatt, “Complete, closed bacterial genomes from microbiomes using nanopore sequencing,” *Nat Biotechnol*, vol. 38, no. 6, pp. 701–707, Jun. 2020, doi: 10.1038/S41587-020-0422-6.
- [47] A. Bruno, J. M. Aury, and S. Engelen, “BoardION: real-time monitoring of Oxford Nanopore sequencing instruments,” *BMC Bioinformatics*, vol. 22, no. 1, pp. 1–8, Dec. 2021, doi: 10.1186/S12859-021-04161-0/FIGURES/5.
- [48] J. R. Revollo, J. A. Miranda, and V. N. Dobrovolsky, “PacBio sequencing detects genome-wide ultra-low-frequency substitution mutations resulting from exposure to chemical mutagens,” *Environ Mol Mutagen*, vol. 62, no. 8, pp. 438–445, Oct. 2021, doi: 10.1002/EM.22462.
- [49] Y. Zeng, X. Wang, and J. Zhang, “Single-cell biomedicine: roles of single-cell nuclear elements,” *Cell Biol Toxicol*, vol. 36, no. 1, Feb. 2020, doi: 10.1007/S10565-020-09515-7.

- [50] S. B. Kingan et al., “A High-Quality De novo Genome Assembly from a Single Mosquito Using PacBio Sequencing,” *Genes (Basel)*, vol. 10, no. 1, p. 62, Jan. 2019, doi: 10.3390/GENES10010062.
- [51] Illumina, “Sequencing Platforms | Compare NGS platform applications & specifications.” <https://www.illumina.com/systems/sequencing-platforms.html> (accessed Oct. 03, 2022).
- [52] M. A. Quail et al., “A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers,” *BMC Genomics*, vol. 13, no. 1, p. 341, 2012, doi: 10.1186/1471-2164-13-341.
- [53] Oxford Nanopore Technologies, “Products | Oxford Nanopore Technologies.” <https://nanoporetech.com/products> (accessed Oct. 01, 2022).
- [54] Oxford Nanopore Technologies, “Store | Oxford Nanopore Technologies.” <https://store.nanoporetech.com/devices.html> (accessed Oct. 01, 2022).
- [55] PacBio, “Sequel systems - PacBio.” <https://www.pacb.com/technology/hifi-sequencing/sequel-system/> (accessed Oct. 01, 2022).
- [56] University of Washington, “University of Washington PacBio Sequencing Services.” <https://pacbio.gs.washington.edu/> (accessed Oct. 01, 2022).
- [57] Illumina, “Office Locations | Contact information.” <https://www.illumina.com/company/contact-us/locations.html> (accessed Oct. 03, 2022).
- [58] Oxford Nanopore Technologies, “Contact us | Oxford Nanopore Technologies.” <https://nanoporetech.com/contact> (accessed Oct. 01, 2022).
- [59] Oxford Nanopore Technologies, “Nanopore Distributors directory.” <https://nanoporetech.com/services/distributors> (accessed Oct. 01, 2022).
- [60] PacBio, “Contact us - PacBio.” <https://www.pacb.com/contact-us/> (accessed Oct. 01, 2022).
- [61] C. Perea et al., “Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP,” *BMC Genomics*, vol. 17, Aug. 2016, doi: 10.1186/S12864-016-2827-7.
- [62] A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, and A. Korobeynikov, “Using SPAdes De Novo Assembler,” *Curr Protoc Bioinformatics*, vol. 70, no. 1, p. e102, Jun. 2020, doi: 10.1002/CPBI.102.
- [63] A. Khezri, E. Avershina, and R. Ahmad, “Hybrid Assembly Provides Improved Resolution of Plasmids, Antimicrobial Resistance Genes, and Virulence Factors in *Escherichia coli* and *Klebsiella pneumoniae* Clinical Isolates,” *Microorganisms*, vol. 9, no. 12, Dec. 2021, doi: 10.3390/MICROORGANISMS9122560.
- [64] M. Santos et al., “Increasing access to next-generation sequencing in oncology for Brazil,” *Lancet Oncology*, vol. 20, pp. 20–23, 2019, doi: 10.1016/S1470-2045(18)30822-2.
- [65] W. A. Mantilla, M. C. Sanabria-Salas, A. M. Baldion, L. F. Sua, D. M. Gonzalez, and M. Lema, “NGS in Lung, Breast, and Unknown Primary Cancer in Colombia: A Multidisciplinary Consensus on Challenges and Opportunities,” *JCO Glob Oncol*, no. 7, pp. 1012–1023, Dec. 2021, doi: 10.1200/go.21.00046.
- [66] M. Helmy, M. Awad, and K. A. Mosa, “Limited resources of genome sequencing in developing countries: Challenges and solutions,” *Appl Transl Genom*, vol. 9, pp. 15–19, Jun. 2016, doi: 10.1016/J.ATG.2016.03.003.
- [67] S. A. Vishnopolska et al., “Genetics and genomic medicine in Argentina,” *Mol Genet Genomic Med*, vol. 6, no. 4, pp. 481–491, Jul. 2018, doi: 10.1002/MGG3.455.

- [68] M. Chand and R. L. Tung, “Skilled immigration to fill talent gaps: A comparison of the immigration policies of the United States, Canada, and Australia,” *Journal of International Business Policy*, vol. 2, no. 4, pp. 333–355, Dec. 2019, doi: 10.1057/S42214-019-00039-4/TABLES/3.



Disponible en:

<https://www.redalyc.org/articulo.oa?id=699878260009>

Cómo citar el artículo

Número completo

Más información del artículo

Página de la revista en redalyc.org

Sistema de Información Científica Redalyc
Red de revistas científicas de Acceso Abierto diamante
Infraestructura abierta no comercial propiedad de la
academia

Dorian Rojas-Villalta, Daniela Benavides-Villegas,
Belén Angulo-Hidalgo, Luis Muñoz-Solórzano,
Chiara Consumi-Tubito

**Caracterización de las tecnologías de secuenciación
genética de segunda y tercera generación**

Characterization of second- and third-generation genomic
sequencing technologies

Tecnología en marcha

vol. 37, núm. 2, p. 70 - 81, 2024

Instituto Tecnológico de Costa Rica, Costa Rica

revistatm@itcr.ac.cr

/ ISSN-E: 2215-3241