



Dr. Allan Orozco Solano. Director del Programa Institucional de la Sociedad de la Información y Conocimiento (Prosic-UCR).

# Voz experta: Infraestructura de secuenciación genómica para el coronavirus SARS-CoV-2 en Costa Rica

Una invitación a potenciar la Red Nacional de Secuenciación genómica

27 JUL 2021 Ciencia y Tecnología

La secuenciación genómica es un área de trabajo de la biotecnología y biología molecular que se refiere a la lectura de la información genética de los organismos. Requiere una parte de trabajo en 'wet lab', (laboratorio húmedo), es decir, experimental, y otra en 'dry lab' (laboratorio seco), que es la parte computacional. En la primera fase, se preparan las muestras biológicas siguiendo protocolos experimentales para posteriormente secuenciarlos en equipos que son denominados NGS ('Next Generation Sequencing', secuenciadores de próxima generación). Una vez obtenidos los datos a partir de los mismos equipos, estos se pasan a una computadora para filtrarlos, normalizarlos y prepararlos para efectos de obtener un recurso de procesamiento, los 'datasets'. Con estos se realizan análisis de datos empleando distintas herramientas bioinformáticas. La secuenciación genómica forma la base de la medicina de precisión y la vigilancia de epidemias y pandemias, entre otros aspectos. Con el propósito de avanzar en el proyecto genómico de Costa Rica, se fundó por decreto nacional el Consejo Técnico de Bioinformática Clínica (CTBC) adjunto al Ministerio de Salud compuesto por 10 Instituciones nacionales. Aproximadamente hace dos años se aprobó la creación de la Red de Secuenciación Nacional y Supercomputación, posteriormente denominada 'RESCUE' (REd de SeCUenciación) con el propósito de unir distintos actores y aumentar las capacidades de secuenciación genómica a nivel nacional.

## LEA MÁS: Voz experta UCR

En la época actual de la pandemia del SARS-CoV-2 se están usando ampliamente las técnicas de secuenciación genómica en Costa Rica y el mundo para obtener las lecturas del genoma del virus. El genoma del virus SARS-CoV-2 posee un tamaño aproximado de 30 mil bases, un valor relativamente pequeño comparado con otros organismos. El genoma humano diploide contiene 6,4 mil millones de bases, por ejemplo. Según los archivos de GISAID (<https://www.gisaid.org/>), una base de datos pública que registra datos genómicos principalmente del virus de la influenza y el SARS-CoV-2 mundialmente, tenemos al momento 792 registros de lecturas genómicas del coronavirus SARS-CoV-2 procedentes de Costa Rica. Dichas muestras fueron tomadas de distintas personas (con un pre-tamizaje) provenientes de varias zonas de Costa Rica. Este valor no es representativo desde un punto de vista epidemiológico y estadístico, y es insuficiente para realizar un control epidemiológico efectivo y facilitar la detección e identificación de variantes del

virus para efectos de una población de poco más de 5 millones de personas. Por tanto, nos surge la gran pregunta: ¿Cuál es la infraestructura tecnológica de secuenciación genómica que dispone Costa Rica actualmente?

A la fecha, Costa Rica tiene una infraestructura activa de secuenciación genómica compuesta de nueve equipos NGS, con tecnología SBS ('Sequencing by Synthesis', secuenciación por síntesis). Todos son equipos provenientes del fabricante norteamericano Illumina, líder mundial en el área de secuenciación genómica. Los equipos identificados y localizados en Costa Rica son los siguientes:

Lugar de ubicación	Modelo del equipo	Capacidad	Número de equipos	Fabricante
DC LAB-UCR	NextSeq 550	alta	1	Illumina
INCIENSA	MiSeq	media	2	Illumina
CIHATA-UCR	MiSeq	media	1	Illumina
CIBCM-UCR	MiSeq	media	1	Illumina
OIJ	MiSeq	media	1	Illumina
SENASA	MiSeq	media	1	Illumina
HNN-CCSS	MiSeq	media	1	Illumina
HSJD	iSeq 100	baja	1	Illumina

Con respecto a la capacidad máxima total de secuenciación genómica disponible en Costa Rica a nivel de público, disponemos de unos 226,2 Gb de rendimiento teórico en salida genómica (datos de secuenciación en bruto). Por supuesto, deben considerarse la profundidad de secuenciación y otros factores con lo cual estos valores deben tomarse con discreción. Con esta capacidad combinada, una adecuada coordinación logística, con recursos económicos asignados, personal contratado y dedicado, y reactivos a tiempo, fácilmente podríamos secuenciar más de 3000 genomas mensuales en modalidad de la RESCUE, combinando esfuerzos de forma conjunta. La mayoría de todos estos secuenciadores se encuentran dentro de instituciones y centros que son miembros del Consejo Técnico de Bioinformática del Ministerio de Salud lo cual podría proporcionar un mayor nivel de cooperación y colaboración en RESCUE. Todos los secuenciadores anteriormente señalados se encuentran en centros, laboratorios, organismos o instituciones públicas del país.

"Tenemos la capacidad de crear una red colaborativa fuerte, ya que disponemos de profesionales capacitados y con experiencia tanto como los recursos tecnológicos en el país. Por tanto, disponemos de todos los elementos necesarios para consolidar la red RESCUE constituida por nodos periféricos (espacios donde haya secuenciadores genómicos) para apoyar al INCIENSA y crear una mejor dinámica en la investigación conjunta a nivel nacional".

En Costa Rica tenemos una distribución de equipos en tres categorías diferentes en función de su rendimiento: baja, media y alta capacidad para procesar datos genómicos. A continuación, se detallan las capacidades de los tres secuenciadores genómicos más frecuentes en Costa Rica:

Modelo	Longitud de lectura	Tiempo de corrida	Capacidad Máxima Salida (Gb)	Capacidad Máxima de lecturas x corrida
<b>NextSeq 550</b>	2 x 150	12-30 horas	120 Gb	400 millones
<b>MiSeq</b>	2 x 300	4-55 horas	15 Gb	25 millones
<b>iSeq 100</b>	2 x 150	9.5-19 horas	1.2 Gb	4 millones

**Solamente el equipo NextSeq 550 de la Universidad de Costa Rica posee cuatro veces la capacidad de procesamiento de datos genéticos que las dos unidades MiSeq disponibles en el Instituto Costarricense de Investigación y Enseñanzas en Nutrición y Salud (INCIENSA).** Tiene un tiempo de corrida de 12-30 horas, un máximo de salida de 120 Gb, y dispone también de un máximo de lecturas por corrida de 400 millones en función de un límite dimensional de lecturas cortas pareadas de 2 x 150 bp. Por otro lado, dicha unidad tiene 100 veces la capacidad en términos de Gb de salida del equipo de secuenciación más pequeño disponible (iSeq 100), y ocho veces la capacidad de la unidad individual más popular en el país, el modelo denominado MiSeq. Por ejemplo, los reactivos que permiten realizar 100 secuenciaciones genómicas del virus SARS-CoV-2 en ese equipo tiene un costo aproximado de \$16 a 18 mil dólares (incluyendo los kits Illumina, librerías, etc.) en función del protocolo implementado, es decir un valor de un poco más del 7% del valor de compra del equipo. En este proceso no solamente se debe tener en cuenta el tiempo de secuenciación, sino también el tiempo de preparación analítica de las muestras (librerías, protocolo empleado, tipo de kit, métodos, otros reactivos base, etc.). Algunas veces se emplea un protocolo específico (KIT Nextera, index y un cartucho de secuenciación) para un proceso de carga relativamente lento (24/48 muestras por celda) pero todo depende también que tan rápido se hagan las librerías requeridas experimentalmente (en función del personal disponible y su eficiencia) y cómo reacciona un laboratorio al avance

tecnológico y adaptación de los nuevos protocolos y test comerciales empleados para SARS-CoV-2.

**Así mismo, el NextSeq 550 de la UCR podría emplear un kit diseñado para obtener 3 072 genomas (COVIDSeq), y correr unos 384 genomas cada vez a un costo en reactivos de \$26 por genoma secuenciado** (un poco menos de un tercio que el obtenido por un MiSeq) del SARS-CoV-2. En un MiSeq empleando el COVIDSeq solamente puede correr tandas de 96 genomas (48/56 hrs) por cada uso. Es decir, con una buena coordinación entre instituciones y con los recursos adecuados podríamos tener la capacidad de secuenciar unos 3000 genomas por mes, con un costo base total de reactivos de alrededor de \$80K (excluyendo el costo de personal y demás). Una adecuada coordinación con los proveedores de los reactivos en modalidad de 'entradas programadas' puede favorecer la secuenciación y evitar el desequilibrio de algún componente (por ejemplo, el index u otra cosa complementaria) del proceso.

Debemos mencionar que no todos los laboratorios o centros donde están localizados los secuenciadores cuentan con las condiciones para realizar la preparación de las muestras, por ejemplo muchos no tienen cámara B2 de bioseguridad para gestionar muestras infectocontagiosas, aunque habría formas reducibles e inactivas en la gestión o transferencia del material genético. Siempre existe la opción de recibir el material genético extraído de las muestras del SARS-CoV-2 y por tanto no se requieren de cámaras de bioseguridad tipo B2. La secuenciación genómica es una actividad de alto costo y requiere suficientes recursos económicos y humanos.

Por ejemplo, si comparamos la capacidad del Reino Unido en secuenciación genómica para una vigilancia relativamente eficiente de las variantes, dicho país ha compartido 587 705 genomas de SARS-CoV-2 de 5 697 916 de casos reportados de COVID-19 (fecha de corte a la edición de la nota). El Reino Unido formó un consorcio liderado por el Departamento de Salud, el Instituto Wellcome Sanger (por ejemplo, dispone de unidades de secuenciación como el NovaSeq 6000 del fabricante Illumina que puede generar hasta 5-6 TB/13-44 horas en una sola corrida con las muestras ya preparadas) y más de 15 universidades públicas y privadas, todas secuenciando de una forma coordinada. Dinamarca es el país que tiene un factor de proporción más alta de genomas compartidos en el mundo, con 126 767 genomas producidas contra 311 520 casos reportados (<https://www.gisaid.org/index.php?id=208>). USA tiene 650 543, pero contra unos 34 223 487 casos reportados. Así mismo, según registros de GISAID en América Latina, por ejemplo, Panamá tiene 896 genomas compartidos de 429 083 casos reportados, y es el país que más ha secuenciado al momento en toda la región centroamericana. Costa Rica ha compartido 792 genomas contra 398 608 de casos de COVID-19 reportados. En las islas del caribe, Trinidad y Tobago tiene 485 genomas registrados contra 37 365 casos reportados.

**Finalmente, en una situación de emergencia nacional la creación de redes o consorcios de secuenciación que incluya compartir equipos, recursos humanos y colaboración de grupos pueden unir nuestros investigadores, compartir ciencia y tecnología, y potenciar esfuerzos para luchar todos juntos contra un enemigo común que afecta la salud de nuestra población costarricense.** Tenemos la capacidad de crear una red colaborativa fuerte, ya que disponemos de profesionales capacitados y con experiencia tanto como los recursos tecnológicos en el país. Por tanto, disponemos de todos los elementos necesarios para consolidar la red RESCUE constituida por nodos periféricos (espacios donde haya secuenciadores genómicos) para apoyar al INCIENSA y crear una mejor dinámica en la investigación conjunta a nivel nacional.

En este momento, la UCR ya está proporcionando un apoyo en análisis de datos bioinformáticos. Necesitamos voluntad de cooperación de todos los actores, coordinación y logística, recursos económicos, financiamiento de personal dedicado y asignado, y la llegada a tiempo de los distintos reactivos que usan los equipos en tiempos de alta demanda comercial a nivel mundial. Crucialmente, con la estrategia de RESCUE podemos secuenciar genómicamente en mayor grado y rendimiento, y generar al menos unos 3000 genomas por mes, estableciendo un sistema de monitoreo más eficiente del SARS-CoV-2 en nuestro país. Las tecnologías de información deben compartirse cuando son recursos públicos, especialmente en casos de emergencia nacional con el propósito de mantener la integridad y salud de nuestra sociedad.

**¿Desea enviar sus artículos a este espacio?**



Los artículos de opinión de *Voz experta UCR* tocan temas de coyuntura en textos de 6 000 a 8 000 caracteres con espacios. La persona autora debe estar activa en su respectiva unidad académica, facilitar su correo institucional y una línea de descripción de sus atestados. Los textos deben

dirigirse al correo de la persona de la Sección de Prensa a cargo de cada unidad. En el siguiente enlace, puede consultar los correos electrónicos del personal en periodismo: <https://odi.ucr.ac.cr/prensa.html>

[Dr. Allan Orozco Solano](#)

Director de Prosic-UCR y miembro del Consejo Técnico de Bioinformática Clínica (CTBC) del Ministerio de Salud

[ALLAN.OROZCOSOLANO@ucr.ac.cr](mailto:ALLAN.OROZCOSOLANO@ucr.ac.cr)

**Etiquetas:** [#vozexperta](#).