

Sevilla, lunes, 17 de julio de 2023

Científicos del CSIC crean una herramienta bioinformática para obtener genomas de calidad

- Diseñada por científicos del IPBLN-CSIC, puede generar genomas de referencia para parásitos como el de la malaria
- La comunidad científica puede valerse de esta herramienta para sus investigaciones de manera fácil y gratuita



Los genomas de referencia son representaciones fehacientes y fiables del material genético objeto de estudio. / iStock

Investigadores del Instituto de Parasitología y Biomedicina López Neyra, perteneciente al Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), en colaboración con un equipo de la Universidad de Glasgow (Escocia), han creado ILRA, una herramienta bioinformática con un gran potencial para generar genomas de alta calidad en contextos especialmente problemáticos, como en el caso de especies con genomas complejos y cuando las muestras biológicas de origen son insuficientes o de mala calidad. Este estudio, [publicado en la revista *Briefings In Bioinformatics*](#), ha sido posible con el apoyo

de la Fundación “La Caixa”, el Ministerio de Ciencia e Innovación y la organización benéfica británica *Wellcome Trust*.

Los avances más destacados e innovadores en investigación biomédica van de la mano de las llamadas tecnologías de secuenciación masiva y técnicas *ómicas*. Estas técnicas permiten obtener, a partir de una muestra de material biológico, la composición exacta de las secuencias del ADN, desglosada en sus cuatro componentes llamados bases o nucleótidos: adenina (A), citosina (C), guanina (G) y timina (T), las famosas ACGT. Un ejemplo es la genómica, que permite descifrar la composición precisa del ADN, todos los genes, de un organismo, proporcionando una base de datos de enorme valor y utilidad para multitud de estudios posteriores, por ejemplo, la predisposición a padecer una determinada enfermedad o las características genéticas que hacen a un determinado patógeno muy virulento. No sólo eso, varios proyectos internacionales, como el consorcio ERGA y el BioGenoma, pretenden secuenciar 4 millones de eucariotas en un futuro próximo.

Sin embargo, todos estos avances dependen en gran medida de poder contar con genomas de referencia de alta calidad, es decir, que sean una representación fehaciente y fiable del material genético de origen. Así, si los genomas de referencia contienen errores, como secuencias inexactas, las interpretaciones de los estudios que los utilicen pueden ser erróneas, provocando, por ejemplo, que se identifiquen mutaciones que, en realidad, no existen, o ignorando genes o adaptaciones relevantes para una enfermedad o proceso biológico específico. En definitiva, obstaculizando el progreso científico.

En el caso de los parásitos, algunos de ellos causantes de las enfermedades humanas más mortales, la gran mayoría de genomas de referencia presentan problemas y limitaciones. Por ejemplo, gran parte del conocimiento se basa en genomas que se han obtenido a partir de cultivos de parásitos en el laboratorio, los cuales pueden diferir de los parásitos que circulan en entornos naturales. Las dinámicas de transmisión en áreas endémicas de África, Asia o América del Sur pueden moldear y dejar huella en el genoma de estos parásitos, reflejando procesos evolutivos de adaptación al ambiente. Sin embargo, la investigación actual tiende a utilizar genomas de referencia de unas pocas cepas y con ello se pretende representar a todos los parásitos del mundo, pero es necesario evitar esta generalización, produciendo genomas *de novo* y mejorando los de referencia que se utilizan en cada caso, estudio o circunstancia concreta. El investigador del IPBLN y primer autor de la investigación, **José Luis Ruiz Rodríguez**, compara estos genomas con un mapa de carreteras: “Cuanto más detallados y precisos sean, más fáciles nos será conducir por el gran y diverso paisaje de la genética”.

En este contexto, surge la herramienta bioinformática ILRA, gracias a la investigación liderada por **Ruiz Rodríguez** y **Elena Gómez Díaz**, también investigadora del IPBLN-CSIC, así como por el catedrático de Biología Computacional **Thomas Dan Otto**, de la Universidad de Glasgow.

En su flujo de trabajo, ILRA combina programas previamente existentes con nuevas soluciones para mejorar la corrección de los genomas, incluso permitiendo la integración de otros datos de secuenciación del mismo material genético, si están disponibles.

La investigadora **Gómez Díaz** destaca cómo ILRA es parte de un proyecto comprometido con la diversidad y la ciencia igualitaria porque hace posible la corrección automática de genomas de referencia con una implicación mínima del usuario y sin exigir grandes conocimientos bioinformáticos, que no suelen ser accesibles para la mayoría de grupos de investigación: “El objetivo de nuestra herramienta es que cualquier laboratorio, sin importar sus recursos disponibles o su experiencia en análisis de datos, sea capaz de producir genomas de alta calidad, algo crucial dado que la mayoría de la biodiversidad y la investigación en enfermedades infecciosas se focaliza en países en vías de desarrollo”.

Genomas de referencia casi perfectos del parásito de la malaria

En el trabajo, se puso a prueba el nuevo programa comparando genomas obtenidos con múltiples técnicas de secuenciación, incluyendo algunas de menor calidad a modo de control, y diferentes organismos, como secuencias humanas y de varios parásitos. Destaca el caso del parásito *Plasmodium falciparum*, que ostenta el récord de ser uno de los organismos con una composición de nucleótidos más extrema, con una abundancia de las bases A y T muy inusual, superando esta el 80 % de media. Esto provocaba que la secuenciación de este parásito haya sido inusualmente complicada, dificultando su estudio y limitando el conocimiento básico de su biología. Esto es crítico ya que *Plasmodium* es el causante de la malaria, una enfermedad infecciosa que mata a cientos de miles de personas al año para la que aún no hay un medicamento o vacuna plenamente efectivos. La lucha contra la enfermedad se ha visto afectada en gran parte por la falta de buenos genomas de referencia que representen la diversidad cambiante de las cepas que existen en la naturaleza y las comunidades afectadas.

En este estudio, se llevó a cabo un esfuerzo internacional con varios equipos de investigadores de Colombia, Kenia o Ghana, para generar como prueba de concepto genomas de referencia de alta calidad para los parásitos de la malaria de regiones menos representadas, que hasta ahora no tenían genomas disponibles. Quedó así demostrado que el programa permite obtener genomas de referencia casi perfectos que cualquier grupo puede usar en sus investigaciones, ya que representan adaptaciones locales que no están presentes en los genomas de referencia tradicionales.

Para **Otto**, la relevancia de esta investigación es clara: “ILRA es una herramienta que debería ayudar a los grupos de investigación a mejorar sus ensamblados de genoma sin necesidad de conocimientos bioinformáticos profundos”. Además, destaca la importancia de las estancias de investigación, forma mediante la cual surgió este proyecto, recalando que son esenciales para favorecer la transferencia de conocimiento, aprender de diferentes culturas y dar lugar a investigación internacional del más alto nivel.

José Luis Ruiz, Susanne Reimering, Juan David Escobar-Prieto, Nicolas M B Brancucci, Diego F Echeverry, Abdirahman I Abdi, Matthias Marti, Elena Gómez-Díaz, Thomas D Otto. **From contigs towards chromosomes: automatic improvement of long read assemblies (ILRA)**. *Briefings in Bioinformatics*. DOI: [10.1093/bib/bbad248](https://doi.org/10.1093/bib/bbad248)

Delegación del CSIC en Andalucía y Extremadura Comunicación

comunicacion@csic.es