



UN NUEVO ALGORITMO PERMITE DETECTAR LOS ERRORES EN LA SECUENCIACIÓN MASIVA DE ÁCIDOS NUCLEICOS

26-11-2015

Científicos del Centro de Astrobiología (CSIC-INTA) y del Centro Nacional de Biotecnología (CNB, CSIC) han desarrollado un nuevo algoritmo que permite filtrar, con mucha precisión, las secuencias erróneas en estudios de secuenciación masiva de ácidos nucleicos. El algoritmo PBF –Poisson Binomial Filtering, llamado así porque utiliza la distribución binomial de Poisson para detectar y filtrar los datos inexactos en la secuenciación–, minimiza el problema en las lecturas de nucleótidos mediante el cálculo de la distribución de probabilidad de errores de una secuencia a partir de sus parámetros de calidad.

Con la secuenciación se consigue leer la información contenida en las moléculas de ADN o ARN, es decir, se obtiene la lista de bases (adenina, citosina, guanina, timina/uracilo) que compone el segmento leído. Gracias a las actuales plataformas de secuenciación masiva, se pueden producir millones de lecturas en poco tiempo y por un bajo coste, lo cual ha supuesto una revolución en diversos campos de la biología. Sin embargo, algunas de estas lecturas pueden contener errores de secuenciación que comprometerían los resultados obtenidos en estas plataformas, lo cual llevaría a interpretaciones imprecisas.

Los errores, que son frecuentes en estos estudios, repercuten en mayor medida en los referidos a poblaciones microbianas. Cuando los errores alcanzan, al menos, el 3% en toda una secuencia leída, que se suponía de una misma especie, aparecen lo que el investigador principal, Fernando Puente Sánchez, llama “especies fantasmas”. Por tanto, se produce una sobreestimación muy acusada de la diversidad microbiana presente en la muestra. “Se trata de un problema que se aprecia sobre todo en la secuenciación masiva, porque al haber muchas más secuencias, el número total de errores también es mayor”, comenta el científico.

Precisamente, Puente Sánchez necesitaba analizar poblaciones microbianas para su tesis y los errores que surgían de la secuenciación lo llevaron a desarrollar este estudio, que comenzó hace dos años, para identificarlos. “Al hacer el análisis, el secuenciador te indica cuánto ‘se fía’ de cada base analizada. El algoritmo utiliza esa información para descartar las secuencias que tengan más probabilidad de contener errores, y obtener así un resultado más preciso”, asegura el investigador.

Validación del método

El método se validó con 37 conjuntos de datos públicos de secuenciación masiva de comunidades microbianas artificiales y ambientales, con las plataformas de secuenciación 454-Roche, Illumina MiSeq y IonTorrent PGM. Los resultados obtenidos

con este nuevo algoritmo se compararon con los obtenidos con otros métodos al uso, como los incluidos en las distribuciones de software mothur, QIIME y USEARCH, tres referentes en el campo de la ecología microbiana.

El algoritmo PBF descarta sustancialmente menos lecturas que sus predecesores, pero produce representaciones más fidedignas, tanto cuantitativa como cualitativamente, de la verdadera diversidad microbiana presente en las muestras estudiadas. Además, el algoritmo produce resultados óptimos para todas las plataformas de secuenciación existentes y requiere de poca potencia de cálculo, siendo posible ejecutarlo en ordenadores de sobremesa, incluso sobre conjuntos de datos de gran tamaño.

Asimismo, el algoritmo PBF está diseñado para que sea fácil de integrar en los protocolos de análisis de datos ya existentes. Esto ha llevado a su adopción temprana por proveedores de *pipelines* de análisis integrado, como LotuS. Por todas estas razones, los autores confían en que el uso de su método se extienda rápidamente en el campo de la ecología microbiana.

Sobre el CAB

El Centro de Astrobiología (CAB) es un centro de investigación mixto del Consejo Superior de Investigaciones Científicas (CSIC) y del Instituto Nacional de Técnica Aeroespacial (INTA). Creado en 1999, y asociado al NASA Astrobiology Institute (NAI), es el primer centro del mundo dedicado específicamente a la investigación astrobiológica. Su objetivo es estudiar, desde una perspectiva transdisciplinar, el origen, presencia e influencia de la vida en el universo.

En el centro trabajan biólogos, químicos, geólogos, astrofísicos, planetólogos, ingenieros, informáticos, físicos y matemáticos, entre otros. Además de todo lo que tiene que ver con la comprensión del fenómeno de la vida tal y como lo conocemos (su emergencia, condiciones de desarrollo, adaptabilidad a ambientes extremos, etc.), también involucra la búsqueda de vida fuera de la Tierra (exobiología) y sus derivaciones, como son la exploración espacial (planetología) y la habitabilidad. El desarrollo de instrumentación avanzada es también uno de sus objetivos fundamentales.

Actualmente, más de 150 investigadores y técnicos trabajan en el CAB en diferentes proyectos científicos tanto nacionales como internacionales. En el CAB se ha desarrollado el instrumento REMS (Rover Environmental Monitoring Station), una estación medioambiental a bordo de la misión Mars Science Laboratory (MSL) de la NASA que explora actualmente Marte. También participa en las próximas misiones a Marte tanto de la NASA (instrumentos TWINS para InSight y MEDA para Mars2020) como de la ESA (instrumento RAMAN/LIBS para ExoMars).

Más información

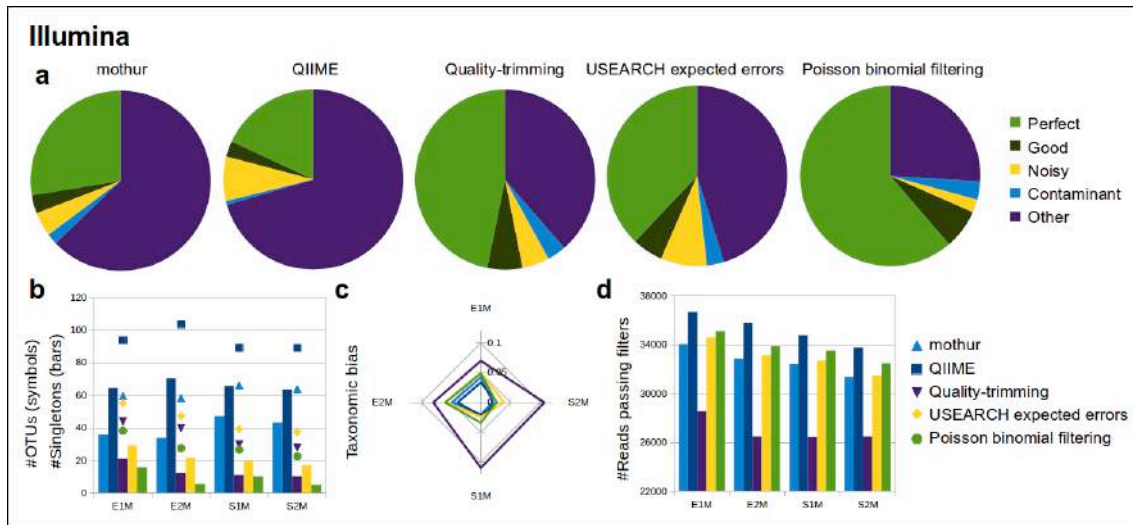
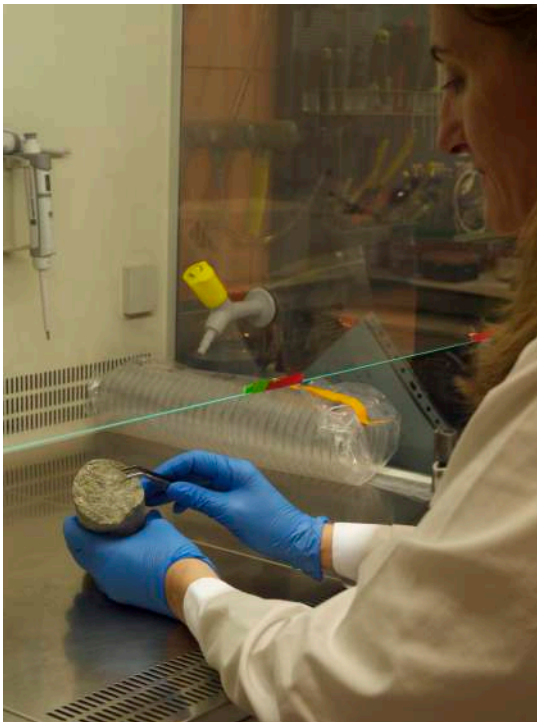


Figura. Comparación de distintos algoritmos de filtrado en datos generados con la plataforma Illumina:

- Fidelidad de las secuencias obtenidas.
- Número total de especies (incluyendo especies fantasma).
- Sesgos causados.
- Número de lecturas retenidas.



Fotografías. Testigo tomado a 150 metros de profundidad de la faja pirítica ibérica en Río Tinto. El ADN extraído de esta roca se ha analizado con el algoritmo PBF.

Artículo científico, publicado en *Nucleic Acids Research*

“A novel conceptual approach to read-filtering in high-throughput amplicon sequencing studies”, por Fernando Puente-Sánchez, Jacobo Aguirre y Víctor Parro, *Nucleic Acids Research*, octubre 2015.

Contacto

Investigador principal:

Fernando Puente-Sánchez: puentesf (+@cab.inta-csic.es)

UNIDAD DE CULTURA CIENTÍFICA DEL CAB

Cristina Delgado: cdelgado (+@cab.inta-csic.es)
Juan Ángel Vaquerizo: jvaquerizo (+@cab.inta-csic.es)

(+34) 915206438

