

ENSAMBLAJE DE NOVO DEL GENOMA DE DOS GALLINAS ESPAÑOLAS AUTÓCTONAS: CASTELLANA NEGRA Y ANDALUZA FRANCISCANA

Peiró-Pastor^{1*}, R., Rauw¹, W.M. y Gómez¹, L.

¹INIA-CSIC, Crta. La Coruña Km. 7,5, 28040 Madrid

*ramon.peiro@inia.csic.es

INTRODUCCIÓN

El proyecto GALLIBER tiene como objetivo principal evaluar el impacto del calentamiento global en la eficiencia productiva de las gallinas ponedoras. Para ello se pretende investigar la variabilidad fenotípica, genotípica y de expresión genética y su relación con situaciones de estrés térmico de varias razas autóctonas españolas. Sin embargo, los genomas de referencia existentes en las bases de datos, *Red Jungle Fowl* (GRCg6a) y *White Leghorn Layer* (GRCg7w), están incompletos, careciendo incluso de algunos de sus cromosomas. Por evitar el uso de mapas de referencia incompletos hemos construido mapas de referencia propios basados en la secuenciación y ensamblado *de novo* de 2 ejemplares de gallinas autóctonas españolas: Castellana Negra y Andaluza Franciscana.

MATERIAL Y MÉTODOS

El ADN de ambos ejemplares se secuenció con la tecnología PacBio Sequel II/Ile en su modo CCS (*Circular Consensus Sequencing*), que produce una secuencia consenso a partir de múltiples sub-lecturas de la misma molécula de ADN con una gran exactitud. Se obtuvieron en cada muestra más de 5M de lecturas con longitudes (N50) superiores a 19Kb, y con un total secuenciado en cada muestra de más de 100Gb. Además de las lecturas originales, se obtuvieron 2 conjuntos de lecturas adicionales aplicando filtros por longitud y por calidad media de lectura ($L > 10Kb - Q > 15$ y $L > 15Kb - Q > 25$). Los tres conjuntos de lecturas correspondientes a cada ejemplar fueron ensamblados con Canu (Nurk *et al.*, 2020) y con Hifiasm (Cheng *et al.*, 2021), obteniéndose un total de 12 ensamblajes diferentes. De cada uno de ellos se obtuvieron diversas métricas, continuando el análisis posterior sólo en el mejor ensamblaje para cada ejemplar. Las potenciales contaminaciones en los ensamblajes fueron identificadas con Diamond (Buchfink *et al.*, 2021) y BlobTools (Laetsch y Blaxter, 2017), eliminándose los *contigs* problemáticos. Seguidamente se obtuvieron los *scaffolds* con la ayuda de los genomas de referencia existentes (GRCg6a y GRCg7w) usando RagTag (Alonge *et al.*, 2022). Finalmente, se anotaron los genomas y se analizó su calidad después del proceso de *scaffolding* con BUSCO (Manni *et al.*, 2021).

RESULTADOS Y DISCUSIÓN

Para cada ensamblaje se han obtenido diferentes métricas de calidad y completitud. Dichas métricas indican una completitud del genoma para la Castellana Negra del 91,4% y para la Andaluza Franciscana del 96,5%, y la proporción de genoma fragmentado fue del 0,6% en ambos casos. Por otro lado, la longitud de los cromosomas ensamblados en ambos ejemplares fue en general similar a los correspondientes genomas de referencia publicados anteriormente.

CONCLUSIÓN

El mejor método de ensamblaje es usando el ensamblador Hifiasm, con un filtrado previo de las lecturas a $L > 10Kb$ y $Q > 15$. Los genomas obtenidos tras el proceso de ensamblaje y depuración muestran buenas métricas de calidad, aunque se necesitan refinamientos manuales posteriores para mejorar las calidades de los genomas y de su anotación. Estos genomas obtenidos podrán ser usados en fases posteriores del proyecto para el análisis de variantes y de expresión génica.

REFERENCIAS BIBLIOGRÁFICAS

• Nurk, S. *et al.* 2020. biorXiv. DOI: 10.1101/2020.03.14.992248. • Cheng, H. *et al.* 2021. Nat. Methods 18: 170-175. • Buchfink B. *et al.* 2021. Nat. Methods 18: 366-368. • Laetsch, D.R. y Blaxter, M.L. 2017. F1000Research 6: 1287. • Alonge, M. *et al.* 2022. Genome Biol. 23: 258. • Manni, M. *et al.* 2021. Current Protocols 1: e323.

Agradecimientos: Agradecemos al personal de la finca El Encín su trabajo de cuidado y cría de las gallinas. Asimismo, agradecemos a los equipos técnicos de los centros de computación FinisTerra3 (CESGA) y Drago (CSIC) las facilidades prestadas en la realización de este estudio.