

Evaluación de la precisión en predicciones genómicas usando marcadores de densidad variada y tag-SNPs en ganado vacuno Avileña-Negra Ibérica

J.L. López-Martínez^{1*}, M.J. Carabaño¹, C. Meneses¹, M. Ramón¹, C. Díaz¹

¹ Departamento Mejora Genética Animal, INIA-CSIC, 28040 Madrid

*Corresponding autor: leonardo.lopez@inia.cisc.es

Resumen

La selección genómica es un conjunto de métodos que emplea marcadores moleculares, principalmente polimorfismos de un solo nucleótido (SNPs), para predecir los valores de cría genómicos (GEBV). Este enfoque se fundamenta en el concepto de desequilibrio de ligamiento, que implica la asociación entre alguno de los SNPs y el QTL responsable de la variación fenotípica observada. Con el fin de obtener GEBV confiables, los SNPs deben capturar la mayor parte de la variación genética aditiva del carácter, lo que precisa el uso de decenas de miles de marcadores. Sin embargo, contar con tal cantidad no siempre garantiza mejores predicciones genómicas, ya que unos pocos marcadores capturan dicha variación genética. Por esta razón, hoy en día se busca disminuir el número de marcadores utilizados en las evaluaciones genómicas sin comprometer la capacidad predictiva que aporta este tipo de información. El objetivo de este estudio consistió en evaluar la precisión de la predicción genómica en registros de edad al primer parto, intervalo entre el primer y el segundo parto, y peso al destete, en ganado vacuno de raza Avileña-Negra Ibérica, a través de escenarios en los que se utilizaron marcadores de distinta densidad. Los cuatro escenarios iniciales involucraron marcadores de densidad media (E1) y alta (E2), *tag-SNPs* (E3) y una combinación de *tag-SNPs* y *singletons* (E4). La precisión de la imputación en E1 y E2 se midió con base a métricas de técnicas de validación cruzada. Los grupos de validación se formaron mediante repeticiones de *k* bloques con el mismo número de individuos y definidos de forma aleatoria, o con una agrupación basada en el parentesco entre individuos mediante *k-means*, con el objetivo de aumentar el parentesco dentro del grupo y disminuirlo entre los grupos. Este último sólo fue utilizado en E2. Para el escenario basado en el uso de *tag-SNPs*, se formaron dos escenarios adicionales. A saber, *tag-SNPs* correctamente imputados mediante agrupación en repeticiones de *k* (E5), y *tag-SNPs* correctamente imputados mediante agrupación de *k-means* basado en parentesco (E6). Las predicciones del GEBV se obtuvieron mediante el uso del BLUPF90 y se utilizó el método de regresión lineal para evaluar el sesgo, la dispersión y la precisión del GEBV en cada uno de los escenarios. Para los distintos caracteres en cada uno de los escenarios, se observó una ganancia en precisión al incluir información genómica, siendo este mayor en E6 y menor en escenarios con mayor densidad de marcadores (E2 por ejemplo). En lo que se refiere al sesgo y la dispersión, en todos los escenarios se observaron desviaciones en relación a los valores esperados. Se demostró la viabilidad de usar densidades bajas de marcadores para la predicción precisa de los GEBV.

Keywords: Evaluación genómica; Imputación; Método LR; Validación.