

# Modelación estadística y aprendizaje automático para predicción genómica en mejoramiento genético vegetal

*Mónica Balzarini*<sup>1</sup>

<sup>1</sup>mbalzari@gmail.com, Cátedra de Estadística y Biometría. Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba. Unidad de Fitopatología y Modelización Agrícola - Consejo Nacional de Investigaciones Científicas y Técnicas (UFYMA INTA-CONICET), CCT CONICET Córdoba, Argentina.

El mejoramiento genético de caracteres cuantitativos en vegetales es complejo y lento. Los genotipos son evaluados en múltiples años y sitios debido al alto impacto del ambiente. La selección genómica (SG) permite incrementar la tasa de ganancia genética mediante la integración de datos fenotípicos y genómicos en poblaciones de entrenamiento y su posterior uso en la predicción del valor estimado de mejora (GEBV) sólo desde información genómica. La predicción del GEBV se realiza con gran cantidad de datos moleculares; consecuentemente, muchas de las variables predictoras son, a menudo, irrelevantes y redundantes. El objetivo de este trabajo es comparar procedimientos para ajustar modelos de SG con vasta cantidad de marcadores moleculares. Cuatro métodos estadísticos para ajustar modelos predictivos (Bayes A, Bayes B, Bayes C y Regresión Ridge) más Random Forest como algoritmo de aprendizaje automático fueron comparados, con y sin selección previa de marcadores basada en la identificación de marcadores significativos con GWAS (Genome Wide Association Studies). Los modelos se entrenaron con dos poblaciones de genotipos de caña de azúcar (*Sacharum* spp.) y en un panel diverso de genotipos de maíz (*Zea mays* L.) con gran cantidad de datos moleculares y evaluaciones fenotípicas multi-ambientales. La eficiencia fue evaluada a través de la correlación entre los GEBV predichos y los BLUP de los efectos genotípicos luego de descontar efectos ambientales desde los datos fenotípicos. Se observó similar eficiencia entre los métodos de ajuste del modelo de SG, en todas las poblaciones. La eficiencia de los modelos estadísticos de SG se incrementó considerablemente con la selección previa de marcadores por GWAS. La precisión de las predicciones augura la selección de genotipos desde información molecular.

**Keywords:** Marcadores Moleculares, Modelos, Predicción genómica, Eficiencia.