



Selección genómica como herramienta para agilizar los procesos de mejoramiento en palma de aceite

Iván Ayala, Juan Malagón, David Botero, Leonardo Araque, Andrés Tupaz, Jenny Rodríguez, Carmenza Montoya y Hernán Mauricio Romero

Programa de Biología y Mejoramiento, Cenipalma. Correo: hmromero@cenipalma.org

Introducción

El mejoramiento de diferentes rasgos productivos constituye una herramienta importante para la optimización del rendimiento y manejo de diferentes especies vegetales [1]. No obstante, su uso por medio de técnicas clásicas (selección fenotípica) puede llegar a ocupar bastante tiempo, tomando alrededor de 14 años por generación en palma de aceite [2]. Por tal motivo, la búsqueda de métodos alternativos se ha extendido, siendo la selección genómica (SG) una de las propuestas más interesantes, pues a través del análisis del ADN es posible realizar una selección temprana, reduciendo el tiempo de mejoramiento a 4 años por generación [2].

No obstante, la implementación del SG supone varios retos. Uno de los más importantes es la identificación de patrones genéticos que permitan estimar el comportamiento de las plantas, antes de que estas lo desarrollen [3]. Es en este punto, donde herramientas de aprendizaje de máquina generan un aporte de valor, al utilizar modelos que permiten manejar el volumen de datos masivo e identificar los patrones de interés para generar un ciclo de mejoramiento corto y eficaz [4].

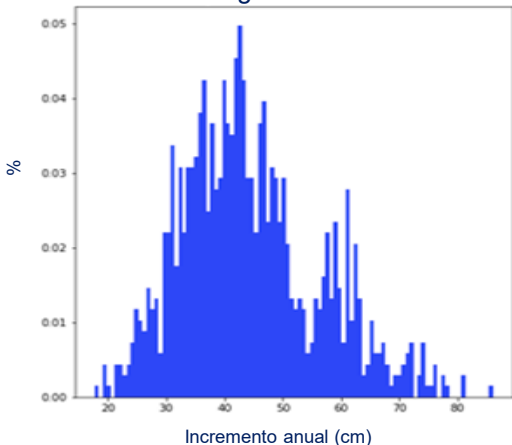
Población inicial



Origen	Cantidad
Angola	485
Bogor	21
Camerún	298
DxD-1	36
DxD-2	60
TxP	14
TxT	17
Variedades	73

El diseño experimental contó con una población inicial de 1.006 materiales vegetales de *Elaeis guineensis*. Dichas palmas fueron seleccionadas de diferentes lugares de África.

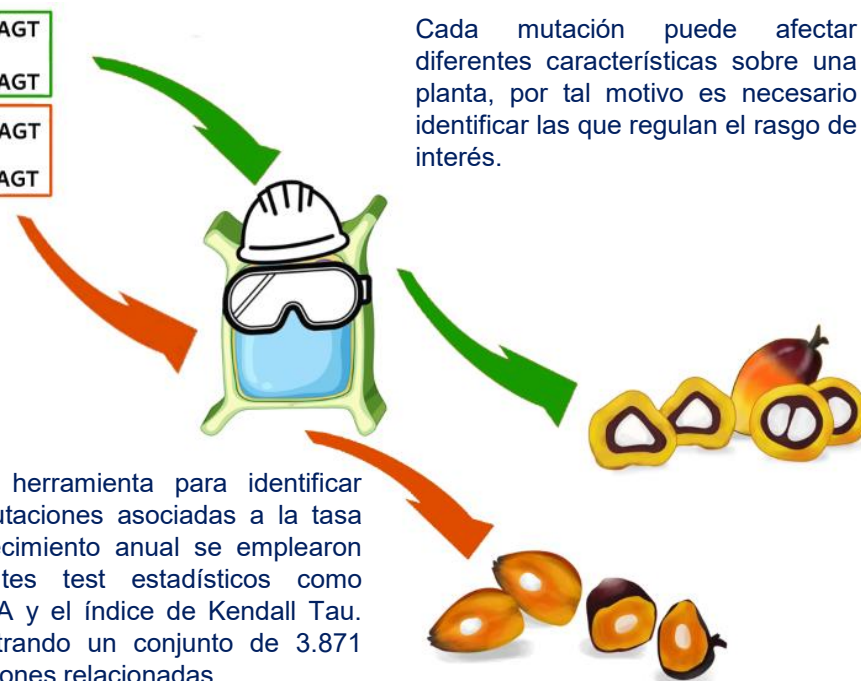
Histograma de altura



Uno de los rasgos de mayor interés es la tasa de crecimiento anual (CA), puesto que puede prolongar el tiempo de vida productivo de la palma.

Selección de marcadores

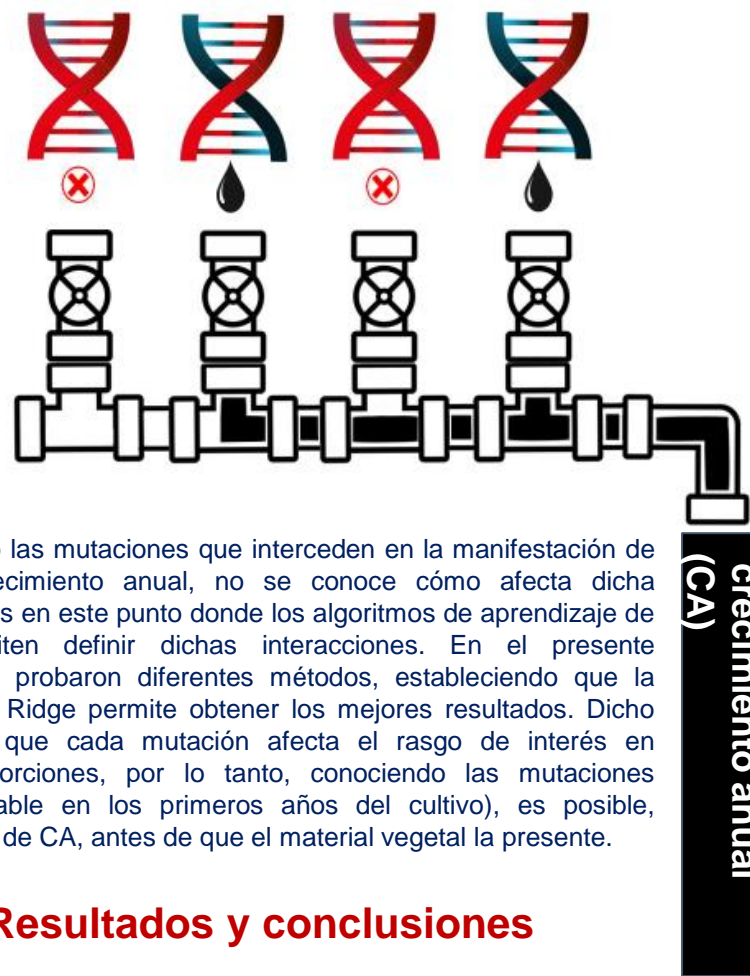
ACTATCAGT
ACTATCAGT
ACTGTCAGT
ACTGTCAGT



Cada mutación puede afectar diferentes características sobre una planta, por tal motivo es necesario identificar las que regulan el rasgo de interés.

Como herramienta para identificar las mutaciones asociadas a la tasa de crecimiento anual se emplearon diferentes test estadísticos como ANOVA y el índice de Kendall Tau. Encontrando un conjunto de 3.871 mutaciones relacionadas.

Inteligencia artificial al servicio de la selección genómica



Aún conociendo las mutaciones que interceden en la manifestación de la tasa de crecimiento anual, no se conoce cómo afecta dicha característica. Es en este punto donde los algoritmos de aprendizaje de máquina permiten definir dichas interacciones. En el presente experimento se probaron diferentes métodos, estableciendo que la regresión lineal Ridge permite obtener los mejores resultados. Dicho modelo define que cada mutación afecta el rasgo de interés en diferentes proporciones, por lo tanto, conociendo las mutaciones (proceso realizable en los primeros años del cultivo), es posible, estimar las tasas de CA, antes de que el material vegetal la presente.

Obtención de ADN

Si bien dos organismos pueden parecer similares y manifestar características semejantes, también pueden tener diferencias significativas en el genoma (mutaciones). Dichas mutaciones pueden servir como biomarcadores, útiles para la selección artificial de especímenes. Con base en esto, se utilizó la técnica de genotipado por secuenciación, como método para establecer el ADN de cada material vegetal presente en el experimento.

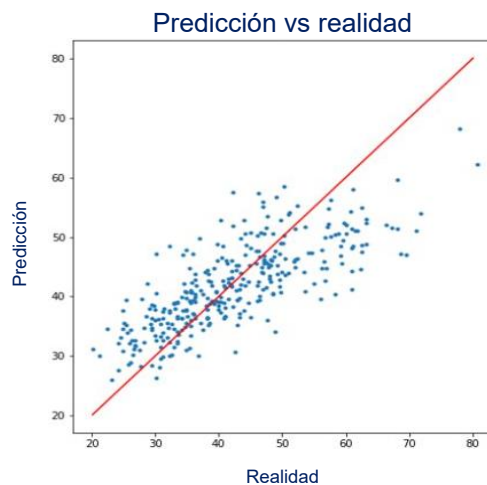


Identificación de mutaciones

ACTGTCAGT
ACTGTCAGT
ACTATCAGT
ACTATCAGT

Usando el ADN obtenido de cada uno de los especímenes de *Elaeis guineensis* se identifican las mutaciones como las secciones con posibles variaciones de organismo a organismo. Dicha identificación se realizó empleando DeepVariante [5]. Una herramienta de aprendizaje profundo que permite aprovechar las bondades de la inteligencia artificial para obtener un listado de mutaciones con confiabilidad. En el presente experimento se encontró un conjunto de 68.953 mutaciones.

Resultados y conclusiones



El modelo Ridge permite realizar una selección temprana de palmas de aceite *Elaeis guineensis* de acuerdo con su tasa de crecimiento anual, con una precisión del 76,13 %.

La selección genómica constituye una herramienta de alto valor para la aceleración de los ciclos de mejoramiento y obtención precisa de posibles parentales.

Bibliografía

- [1] Y. Xu, K. Ma, Y. Zhao, X. Wang, K. Zhou, G. Yu, C. Li, P. Li, Z. Yang, C. Xu, *et al.*, "Genomic selection: A breakthrough technology in rice breeding", *The Crop Journal*, vol. 9, no. 3, pp. 669-677, 2021.
- [2] D. Cros, B. Tchounke, L. Nkague-Nkamba, "Training genomic selection models across several breeding cycles increases genetic gain in oil palm in silico study", *Molecular Breeding*, vol. 38, no. 7, pp. 1-12, 2018.
- [3] M. A. Nadeem, M. A. Nawaz, M. Q. Shahid, Y. Doğan, G. Comertpay, M. Yıldız, R. Hatipoğlu, F. Ahmad, A. Alsaleh, N. Labhane, *et al.*, "DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing", *Biotechnology & Biotechnological Equipment*, vol. 32, no. 2, pp. 261-285, 2018.
- [4] Nayeri, S., Sargolzaei, M., Tulpan, D. (2019). A review of traditional and machine learning methods applied to animal breeding. *Animal health research reviews*, 20(1), 31-46
- [5] Deep variant. Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., ... & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature biotechnology*, 36(10), 983-987.

Agradecimientos al Fondo de Fomento Palmero (FFP)