

13/07/2020

Aplicaciones del 'aprendizaje profundo' (*Deep Learning*) para la mejora genética de los poliploides



Un artículo recientemente publicado por autores del Centre de Recerca en Agrigenòmica (CRAG) y de la Universidad de Florida muestra cómo los métodos computacionales modernos pueden ayudar en el mejoramiento genético de las especies vegetales con más de dos juegos de cromosomas.

La capacidad predictiva de las técnicas de aprendizaje profundo ha sido evaluada en arándanos y fresas (Crédito: Muffet / CC BY)

El desarrollo de nuevas variedades en las plantas suele consistir en un proceso de prueba y error en el que se cruzan líneas 'élite' consolidadas y se evalúa el rendimiento de sus descendientes. Con el tiempo, algunos de estos descendientes reemplazan a las líneas élite cuando las superan en al menos algunos de los caracteres de interés, tales como la resistencia a las enfermedades o el sabor. Éste es un proceso continuo, pero, por desgracia, lento. Por ejemplo, el tiempo para desarrollar una nueva variedad genética de fresa es de más de ocho años.

Muchos caracteres de interés en las plantas son complejos

Por esta razón, los mejoradores se han esforzado en acelerar este proceso utilizando tecnologías genómicas modernas. Una posibilidad es llevar a cabo pruebas genéticas para identificar los cruces e individuos más favorables, cosa que requiere conocer los genes y las mutaciones responsables del carácter. Lamentablemente, y en oposición a lo que se piensa comúnmente, los genes que determinan los rasgos de interés económico sólo se conocen parcialmente; de hecho, hasta ahora se han descubierto relativamente pocas mutaciones causales. Además, la expresión de un carácter –como puede ser el sabor– depende no sólo de los genes, sino también de las condiciones ambientales en las que se cultivan las plantas. Por ejemplo, la misma variedad puede crecer muy bien si la irrigación es suficiente, y a la vez ser muy sensible a la sequía. El personal encargado de la mejora de plantas describe tales rasgos como ‘complejos’, ya que dependen del medio ambiente y de muchos genes que sólo están caracterizados parcialmente.

¿Qué hacer, entonces? Una vez más, los métodos moleculares pueden ayudar, pero utilizando un enfoque complementario llamado ‘predicción genómica’. Este procedimiento consiste en utilizar todos los marcadores genéticos disponibles, incluyendo candidatos que no se han validado previamente, para ‘predecir’ los resultados de los cruces entre distintas variedades. Esto se hace típicamente usando variantes del conocido método de regresión lineal.

Cómo puede ayudar el 'aprendizaje profundo'

Existen numerosos métodos de predicción genómica, pero la mayoría de ellos consideran una relación relativamente simple entre los marcadores genéticos y el carácter de interés. Sin embargo, recientemente, se han desarrollado algunos nuevos métodos de predicción que se denominan como ‘aprendizaje profundo’, *deep learning* en inglés. El aprendizaje profundo comprende un conjunto de algoritmos que se inspiran en el funcionamiento del cerebro humano y que dividen todo el procedimiento de cálculo en pequeñas unidades llamadas ‘neuronas’. Estos métodos son muy populares hoy en día y tienen numerosas aplicaciones que van desde la traducción automática hasta el análisis de vídeo y sonido.

El aprendizaje profundo es de gran interés en genómica, pues es extremadamente flexible en la relación que asume entre los marcadores y los rasgos de interés. Esto es importante en especies como la fresa y el arándano porque son poliploides –tienen más de dos copias del mismo cromosoma–, y es precisamente en este tipo de especies donde podemos esperar que las interacciones entre los genes sean más importantes de lo habitual. El aprendizaje profundo puede ser una herramienta prometedora para la predicción genómica en este escenario.

El trabajo de Zingaretti et al., realizado por el Centro de Investigación en Agrigenómica (Crag) en colaboración con la Universidad de Florida, aborda precisamente esta cuestión. En este estudio mostramos, en una de las primeras aplicaciones del aprendizaje profundo para la predicción genómica, que este método puede ser realmente eficaz en presencia de interacciones entre genes, es decir, cuando un carácter no puede predecirse simplemente considerando los genes individualmente. Lo demostramos con dos especies de interés hortícola: la fresa, que es octoploide, y el arándano, tetraploide. También proporcionamos programas informáticos que pueden ser utilizados por el personal investigador para aplicar el aprendizaje profundo a la predicción genómica. Este trabajo es de especial relevancia para la industria española, ya que, según datos de 2018, España es el primer productor europeo de

fresas y el sexto a nivel mundial.

Miguel Pérez-Enciso

Profesor de investigación ICREA en la UAB.

Centro de Investigación en Agrigenómica (CRAG) CSIC-IRTA-UAB-UB.

miguel.perez@cragenomica.es

Referencias

L.M. Zingaretti, S.A. Gezan, L.F. Ferrão, L.F. Osorio, A. Monfort, P.R. Muñoz, V.M. Whitaker, M. Pérez-Enciso. 2020. **Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species**. *Frontiers in Plant Science* doi.org/10.3389/fpls.2020.00025.

Pérez-Enciso M, Zingaretti LM. 2019. **A Guide on Deep Learning for Complex Trait Genomic Prediction**. *Genes* doi.org/10.3390/genes10070553.

[View low-bandwidth version](#)