

13/07/2020

Applications of 'Deep Learning' to the Genetic Improvement of Polyploids



An article recently published by CRAG (Centre de Recerca en Agrigenòmica) and researchers of the University of Florida shows how modern computational methods can help in the genetic improvement of plant species with more than two sets of chromosomes.

The predictive accuracy of the deep learning techniques has been evaluated in blueberries and strawberries (Credit: Muffet / CC BY)

The development of new varieties in plants typically consists of a trial and error process whereby 'elite' consolidated lines are crossed, and the performance of their descendants is evaluated. Eventually, some of these descendants replace the elite lines when they outperform them in at least some of the traits of interest, say disease resistance or flavor. This is a continuous process but, unfortunately, a slow one. For instance, the time to develop a new strawberry genetic variety is over eight years.

Many traits of interest in plants are complex

For this reason, breeders have strived to accelerate this process using modern genomic technologies. One possibility is to carry out genetic tests to identify the most favorable crosses and individuals. This technique requires that the genes and the causative mutations behind the trait are known. Unfortunately, and contrary to what is commonly thought, the genes that

determine traits of economic interest are only partially known. In fact, relatively few causative mutations have been discovered so far. Besides, the expression of traits, say flavor, depend not only on the genes but also on the environmental conditions on which the plants grow. For instance, the same variety can grow very well if irrigation is enough but, at the same time, it can be very sensitive to drought. Breeders then say that such traits are 'complex', because they depend on the environment and on many genes which are only partially characterized.

What to do, then? Again, molecular methods can help, but using a complementary approach called 'genomic prediction'. This procedure consists in utilizing all genetic markers available in order to 'predict' future performance of plant varieties. This is typically done using variants of the well-known linear regression method.

How Deep Learning can help

There exist numerous genomic prediction methods, but most of them assume a relatively simple relation between genetic markers and the trait of interest. However, some new methods for prediction have been developed recently that are called 'deep learning'. Deep learning comprises a kind of algorithms that are inspired on how the human brain works and break the whole computation procedure into small units called 'neurons'. These methods are extremely popular today and have numerous applications, ranging from automatic translation to video and sound analysis.

The interest of deep learning in genomics is that it is extremely flexible in the relation it assumes between the markers and the traits of interest. This is important in species like the strawberry because they are polyploids –they have more than two copies of the same chromosome–, and it is precisely in this type of species where we can expect that interactions between genes become more important than usual. Deep learning can be a promising tool for genomic prediction in this scenario.

The work by Zingaretti et al., carried out by scientists at the Centre for Research in Agricultural Genomics (CRAG) in collaboration with the University of Florida, addresses precisely this issue. In this study, we show, in one of the first applications of deep learning to genomic prediction, that this method can actually be quite effective in the presence of interactions between genes, that is, when the whole trait cannot be predicted simply by considering the genes individually. This has been demonstrated in two important soft fruit commodities: strawberries, an octoploid, and blueberries, a tetraploid. This work is of particular relevance for the Spanish industry since, as of 2018, Spain is the first European strawberry producer and the sixth worldwide.

Miguel Pérez-Enciso

ICREA Research Professor at the UAB.

Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB.

miguel.perez@cragenomica.es

References

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](#)