Abstract

Autoencoding Genetic Markers to Predict the Value of Ecophysiological Model Parameters - Proof of Concept Using a Sorghum Diversity Panel †

Florian Larue ^{1,*}, Grégory Beurier ¹, Lauriane Rouan ¹, David Pot ¹, Jean-François Rami ¹ and Delphine Luquet ¹

CIRAD, 34980 Montferrier-sur-Lez, France; gregory.beurier@cirad.fr (G.B.); lauriane.rouan@cirad.fr (L.R.); david.pot@cirad.fr (D.P.); jean-francois.ramy@cirad.fr (J.-F.R.); delphine.luquet@cirad.fr (D.L.)

* Correspondence: florian.larue@cirad.fr;

† Presented at the third International Tropical Agriculture Conference (TROPAG 2019), Brisbane, Australia, 11–13 November 2019.

Published: 19 January 2020

Abstract: Crop growth modelling formalizes the biological processes at which genotype X environment interactions (GxE) are expressed. It has the potential to evaluate, *in silico*, the effect of elementary traits and related genetic factors on phenotype and yield elaborations. Recent studies showed that, by driving the value of crop model parameters using a genomic selection model, yield was predicted more accurately than by a classical genetic model. However these studies dealt with few, integrative parameters and a narrow genetic diversity, i.e., a reduced number of molecular markers. This contrasts with the necessity to make crop models more responsive to climate change variables and thus increase the number of physiological parameters, while studying wider genetic diversity to seek for adaptive markers. With this respect, methods that reduce the dimensionality of the problem are needed. The autoencoder, a semi-supervised machine learning method, can reduce the number of predictors (markers) without prior information, by compressing input data into an encoded neural network layer. Applied to genetic diversity, it should ease modelling and predicting the genetic value of crop model parameters underlying GxE and yield variability, compared to classical regression methods. This study aimed at testing the autoencoding of the genetic data (ca. 1.5M markers) within a West-African sorghum diversity panel of 200 individuals. It evaluated then the relevance of autoencoded data to predict the genetic value of the dozen of crop/plant parameters controlling growth and plasticity in Ecomeristem model, estimated using data from an experiment in the Phenoarch platform.

Keywords: autoencoder; SNP; gene-based modeling; ecophysiological model; parameter estimation

Funding: This work was supported by the projects Sorghum Genomics Toolbox (funded by Bill & Melinda Gates Foundation) and Biomass for the Future (ANR-11-BTBR-0006-BFF, funded by the French National Research Agency (ANR)).

Conflicts of Interest: The authors declare no conflict of interest.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).