



REVIEW ARTICLES

The 4.(n) wheat breeding era: genomic-based predictions becoming part of the daily routine and beyond

Eduardo Venske^a, Cezar Augusto Verdi^a, Cesar Daniel Petroli^b, Camila Pegoraro^a, Luciano Carlos da Maia^a, Antonio Costa de Oliveira^{a*}

^aPlant Genomics and Breeding Center, Eliseu Maciel School of Agronomy, Federal University of Pelotas, Capão do Leão campus, PO box 354, zip code 96160-000, capão do Leão, RS, Brazil

^bGenetic Analysis Service for Agriculture - SAGA, International Maize and Wheat Improvement Center - CIMMYT, Carretera México - Veracruz Km 45, C.P. 56237, El Batán, Texcoco, México

Received 20 August 2021. Revised 04 October 2021. Accepted 03 December 2021.

KEYWORDS

Genomic Selection;
Quantitative
Genomics;
Machine Learning;
Data Science;
Breeding 4.0.

Abstract:Wheat (*Triticum aestivum* L.) is one of the most important crops worldwide, also as a role model in plant breeding and genomics. Along with gene edition, genomic selection (GS) is the most advancing area in plant breeding, and several developments are in course in this field, which soon will result in the recently termed “breeding 4.0” stage. The aim of this review is to present and discuss the most recent and impacting advances in GS applied to wheat breeding. Wheat shows particular features, e.g. a crop with narrow genetic variability and a large and complex genome, which motivates especial discussions on the present theme. The advances in enviromics and phenomics are presented and also the way they literally enter the genomic prediction models. The most breeder-advantageous wheat genotyping platforms currently available are also presented. Regarding data analysis within the genomic selection scheme, machine learning and deep learning methods are the most advancing approaches for predicting phenotypes, and improvements on these algorithms are nowadays in the center of the debate. Several advances in course will move plant breeding from the current 3.0 to the 4.0 stage, so achieving the status of “rocket science” and requiring highly skilled breeders and multidisciplinary teams.

Introduction

Wheat is one of the most important crops worldwide, being part of the diet of a large fraction of the human population. The cereal has also especial importance as a role model in biotechnology, genomics and plant breeding. This is due to several features the crop has, such as the general narrow

genetic variability and the large, hexaploid, redundant and complex genome sequence, which make numerous methods a challenge when applying to wheat (Borrill et al., 2019). In other words, the crop deserves special studies and discussion when applying any novel breeding approach.

Although uncountable methods have been (and are) important in plant breeding, nowadays undoubtedly gene edition and genomic selection, including all related tools, are well-acknowledged as the cutting-edge approaches in crop improvement. Genomic selection (GS), which gives

*Corresponding author:

E-mail: acostol@terra.com.br (A. Costa de Oliveira).



regard to the theme of this review, has firstly been proposed about 20 years ago, and basically aims to predict unobserved phenotypes using genomic information, saving time and costs in breeding programs (Meuwissen et al., 2001). In fact, the idea of predicting phenotypes from genotyping profiles is even older (Bernardo, 1994). Although the very basic backbone structure of the genomic selection scheme virtually did not change across two decades, giving the very fortunate sight of the proposers, several improvements have been seen in all “gears of this machine”. The aim of this review is to present and discuss the most recent advances in genomic selection, both in theory and in practical terms, for wheat breeding. Several breeding programs already “got in” breeding 3.0, i.e., integrating genomic data in their selection routines (Wallace et al., 2018), but what about surpassing this barrier? In summary, this article will focus on the latest advances in the basic pillars of GS, i.e., envirotyping, phenotyping, genotyping and data analysis (Figure 1).

Envirotyping

While breeders know very well that both the environment as the genotype x environment interaction can never be neglected in a breeding program, it seems that currently the environment has received a more especial attention in this context. This is probably due to the extremely large number of sites (multi-environmental scale) each breeding program has to deal routinely and even due to the recent climate change outcomes. Environmental typing or “Envirotyping” and “Enviromics” are newly developed concepts in crop research, coming out as a the third “-typing” pillar, alongside phenotyping and genotyping, and gives regard to a deep, and most times, historical, environmental characterization (Xu, 2016; Costa-Neto et al., 2021a; Crossa et al., 2021). It aims to use the similarity among sites in an “omic” scale for forecasting unobserved phenotypes (Resende et al., 2021).

Nevertheless, it is not enough, though, just collecting this large amount of environmental information, but this data must be integrated in the genomic selection analysis. This is the game-changer point, thus especial models and software are being developed, such as the package *EnvirType* (Costa-Neto et al., 2021b), among others. Environmental factors affect each crop differently, even in the context of GS (Li et al., 2021). The integration of this type of data in the GS models are somehow recent and promising. As an example, interaction models which include environmental covariates, alongside pedigree and genomic information, have allowed for predicting wheat hybrid phenotypes, which is probably the hardest task (Basnet et al., 2019). Looking at these results, a question arises: in addition to climate variables, would be now time to start including soil parameters in the prediction models?

Phenotyping

Phenomics is also a recent theme, and came to stay in plant breeding. It gives regard to the use of high-throughput phenotyping technologies and matches very well in the scenario where genomic selection works, in which large numbers of breeding lines and experimental sites are common and saving time and costs are priorities (Crain et al., 2018; Juliana et al., 2019; Singh et al., 2019; Zhao et al., 2019). One way phenomics in fact enters in GS is through adding secondary correlated traits in the models, as covariates or multivariates (e.g. Shabannejad et al., 2020). In this scenario, “special traits”, such as those obtained through images and temperature measurements, gain special relevance.

Adding physiological traits, i.e., vegetation index and canopy temperature, a study has found a range of increments in the GS grain yield prediction, from -7 to 33% (Crain et al., 2018). Similarly, an average of 146% accuracy increase was found, when similar correlated traits were considered (Sun et al., 2019). Days to maturity was predicted three to

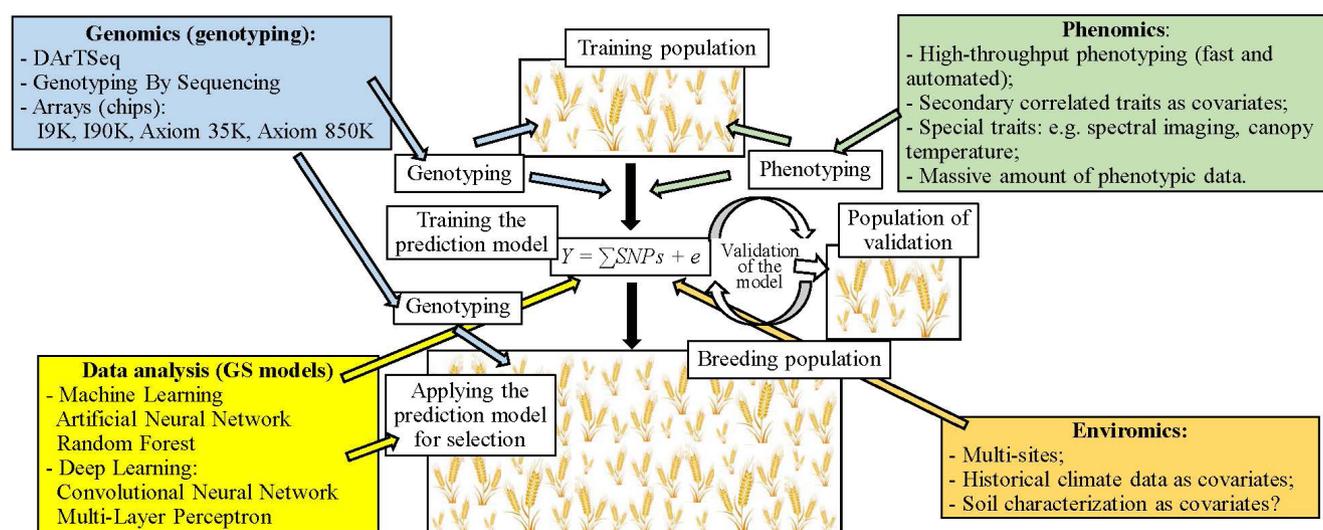


Figure 1. Genomic Selection backbone (in the center) and how the current advances (colored boxes) improve the scheme.

four times more accurately when drought tolerance ratios obtained from machine learning-treated images were added as covariates in the GS equation (Shabannejad et al., 2020). Obviously, each trait is impacted differently by the use of correlated traits in GS. In this regard, a study has shown that the GS prediction accuracy of grain protein and grain yield in spring wheat were increased on average by 12% and 20%, respectively, when spectral reflectance indices were included in the prediction models (Sandhu et al., 2021b).

Genotyping

First of all, it is important to clarify that GS is about quantitative genetics. So, any “qualitative” marker technology, such as the promising ones KASP, Taqman and rhAmp (Ayalew et al., 2019), in general terms does not fit for GS. GS requires wide genomic marker coverage, thus in this regard wheat predictions could be affected due to such large and complex genome. However, the marker numbers required strongly depend on features of the population assayed (e.g. related or non-related individuals) (Haile et al., 2021). Moreover, redundancies do not bring any help (Bassi et al., 2016). A wide assay, using an Axiom™ Affymetrix array, has shown that the prediction accuracies for several traits in a large panel of wheat lines showed substantial response up to just over 5,000 SNPs (Norman et al., 2018).

In fact, there is an increasingly variety of technologies for wheat genotyping, from arrays (chips) to sequencing-based, all delivering a dense number of SNP markers, which may bring a dilemma to breeders (Rasheed et al., 2017). To cite a few, I9K, I90K, Axiom 35K and Axiom 850K are among the most used genotyping arrays for wheat, and genotyping-by-sequencing and DArTSeq, the sequencing-based technologies mostly applied (Bassi et al., 2016). The decision on which tool to use will depend on quality, which ultimately, would be reflected in the final prediction accuracy obtained, but also costs must be kept in mind. The comparison of different technologies on spring wheat, has found that DArTseq was the best option for genomic selection, as it allowed for a similar prediction accuracy comparing to the I90K array (Illumina Infinium BeadChip), but at a lower cost (Liu et al., 2020).

Data analysis

The prediction model is the heart of any GS program, as this equation will allow for all phenotypic forecasting. This is not an easy step, giving the complexity of the large datasets considered (Crossa et al., 2021). Linear models, such as Genomic Best Linear Unbiased Prediction (GBLUP) and Ridge Regression BLUP (rrBLUP) have been the most popular methods, followed by LASSO, Bayesian approaches and few others such as Reaction Norm (Haile et al., 2021). All these methods are still being highly used and improved.

However, other advances are taking place in this regard. Machine Learning (ML) approaches are being incorporated to GS. Among the ML methods already applied in wheat GS, one can cite Random Forest and Artificial Neural Network, as the most commonly used (González-Camacho et al., 2018). Deep

Learning (DL) is a type of Machine Learning approach which applies artificial intelligence and non-parametric statistics, to “deeply” learn data patterns and build models (Montesinos-López et al., 2021). Currently, this type of approach is being applied to several scientific and technological fields, from robots to oncology treatment. The main advantages of DL are its ability to deal with unknown and complex data patterns, which seems to be the case of GS. DL methods are so recent in GS that very few studies have already been performed on wheat. A work on spring wheat has found that DL based methods outperformed classical rrBLUP for all predicted traits (including grain yield) and in all simulated scenarios of number of markers (Sandhu et al., 2021a). There is, at the same time, contrasting performance among DL methods. For instance, Multi-Layer Perceptron showed 5% better results than Convolutional Neural Network (Sandhu et al., 2021a). Although there are several *pros* of ML and DL, further requirements are also demanded, specially highly skilled professionals.

Final considerations and future perspectives

GS has been proposed two decades ago and although the general scheme of the method did not change substantially over time, substantial advances have been seen on each gear of the approach. Enviromics, Phenomics and Genomics make up the modern breeding triangle (Crossa et al., 2021). When applying these factors to GS, it is compulsory to add at least two further key-factors, which connect the corners of the triangle, i.e., advanced knowledge (e.g. genetics) and improved methods for data analysis. In this review we have shown, in brief, the latest advances on these fields applied to wheat.

There is still space for several improvements in GS for wheat, especially regarding data analyses, with the integration of enviromics and phenomics to genomic data, in advanced prediction models. Gradually, breeding 3.0 is getting democratized (Wallace et al., 2018), but the current stage has firstly to be strongly built to allow moving to the next level, which would be breeding 4.0 and beyond. Uncountable advances have been seen in crop breeding since its beginning (Bernardo, 2016). It becomes visible that the breeder of the near future has to be a highly skilled professional, and surrounded by a multidisciplinary team, to deal with so many complex advances. Breeding is getting the status of “rocket science”.

Conflicts of interest

The authors declare no conflicts of interest.

Funding

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and

Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), and FAO/IAEA.

Acknowledgements

The authors acknowledge the Brazilian funding agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), and FAO/IAEA for grants and scholarships provided.

References

- Ayalew, H., Tsang, P. W., Chu, C., Wang, J., Liu, S., Chen, C., & Ma, X. F. (2019). Comparison of TaqMan, KASP and rhAmp SNP genotyping platforms in hexaploid wheat. *PLoS One*, *14*(5), e0217222. <http://dx.doi.org/10.1371/journal.pone.0217222>. PMID:31116793.
- Basnet, B. R., Crossa, J., Dreisigacker, S., Pérez-Rodríguez, P., Manes, Y., Singh, R. P., Rosyara, U. R., Camarillo-Castillo, F., & Murua, M. (2019). Hybrid wheat prediction using genomic, pedigree, and environmental covariables interaction models. *The Plant Genome*, *12*(1), 180051. <http://dx.doi.org/10.3835/plantgenome2018.07.0051>. PMID:30951082.
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science*, *242*, 23-36. <http://dx.doi.org/10.1016/j.plantsci.2015.08.021>. PMID:26566822.
- Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Science*, *34*(1), 20-25. <http://dx.doi.org/10.2135/cropsci1994.0011183X003400010003x>.
- Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and Applied Genetics*, *129*(12), 2323-2332. <http://dx.doi.org/10.1007/s00122-016-2772-5>. PMID:27681088.
- Borrill, P., Harrington, S. A., & Uauy, C. (2019). Applying the latest advances in genomics and phenomics for trait discovery in polyploid wheat. *The Plant Journal*, *97*(1), 56-72. <http://dx.doi.org/10.1111/tpj.14150>. PMID:30407665.
- Costa-Neto, G., Fritsche-Neto, R., & Crossa, J. (2021a). Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. *Heredity*, *126*(1), 92-106. <http://dx.doi.org/10.1038/s41437-020-00353-1>. PMID:32855544.
- Costa-Neto, G., Galli, G., Carvalho, H. F., Crossa, J., & Fritsche-Neto, R. (2021b). *EnvRtype*: a software to interplay enviromics and quantitative genomics in agriculture. *G3 (Bethesda, Md.)*, *11*(4), jkab040. <https://doi.org/10.1093/g3journal/jkab040>.
- Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., & Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *The Plant Genome*, *11*(1), 170043. <http://dx.doi.org/10.3835/plantgenome2017.05.0043>. PMID:29505641.
- Crossa, J., Fritsche-Neto, R., Montesinos-Lopez, O. A., Costa-Neto, G., Dreisigacker, S., Montesinos-Lopez, A., & Bentley, A. R. (2021). The modern plant breeding triangle: Optimizing the use of genomics, phenomics, and enviromics data. *Frontiers in Plant Science*, *12*, 651480. <http://dx.doi.org/10.3389/fpls.2021.651480>. PMID:33936136.
- González-Camacho, J. M., Ornella, L., Pérez-Rodríguez, P., Gianola, D., Dreisigacker, S., & Crossa, J. (2018). Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. *The Plant Genome*, *11*(2), 170104. <http://dx.doi.org/10.3835/plantgenome2017.11.0104>. PMID:30025028.
- Haile, T. A., Walkowiak, S., N'Diaye, A., Clarke, J. M., Hucl, P. J., Cuthbert, R. D., Knox, R. E., & Pozniak, C. J. (2021). Genomic prediction of agronomic traits in wheat using different models and cross-validation designs. *Theoretical and Applied Genetics*, *134*(1), 381-398. <http://dx.doi.org/10.1007/s00122-020-03703-z>. PMID:33135095.
- Juliana, P., Montesinos-López, O. A., Crossa, J., Mondal, S., González Pérez, L., Poland, J., Huerta-Espino, J., Crespo-Herrera, L., Govindan, V., Dreisigacker, S., Shrestha, S., Pérez-Rodríguez, P., Pinto Espinosa, F., & Singh, R. P. (2019). Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theoretical and Applied Genetics*, *132*(1), 177-194. <http://dx.doi.org/10.1007/s00122-018-3206-3>. PMID:30341493.
- Li, X., Guo, T., Wang, J., Bekele, W., Sukumaran, S., Vanous, A. E., McNellie, J. P., Tibbs-Cortes, L. E., Lopes, M. S., Lamkey, K. R., Westgate, M. E., McKay, J. K., Archontoulis, S. V., Reynolds, M. P., Tinker, N. A., Schnable, P. S., & Yu, J. (2021). An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. *Molecular Plant*, *14*(6), 874-887. <http://dx.doi.org/10.1016/j.molp.2021.03.010>. PMID:33713844.
- Liu, C., Sukumaran, S., Jarquin, D., Crossa, J., Dreisigacker, S., Sansaloni, C., & Reynolds, M. (2020). Comparison of array-and sequencing-based markers for genome-wide association mapping and genomic prediction in spring wheat. *Crop Science*, *60*(1), 211-225. <http://dx.doi.org/10.1002/csc2.20098>.
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829. <http://dx.doi.org/10.1093/genetics/157.4.1819>. PMID:11290733.
- Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., & Crossa, J. (2021). A review of deep learning applications for genomic selection. *BMC Genomics*, *22*(19), 1-23. <http://dx.doi.org/10.1186/s12864-020-07319-x>. PMID:33407114.
- Norman, A., Taylor, J., Edwards, J., & Kuchel, H. (2018). Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. *G3 (Bethesda, Md.)*, *8*(9), 2889-2899. <http://dx.doi.org/10.1534/g3.118.200311>. PMID:29970398.
- Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., & He, Z. (2017). Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. *Molecular Plant*, *10*(8), 1047-1064. <http://dx.doi.org/10.1016/j.molp.2017.06.008>. PMID:28669791.
- Resende, R. T., Piepho, H. P., Rosa, G. J. M., Silva-Junior, O. B., E Silva, F. F., de Resende, M. D. V., & Grattapaglia, D. (2021). Enviromics in breeding: Applications and perspectives on envirotypic-assisted selection. *Theoretical and Applied Genetics*, *134*(1), 95-112. <http://dx.doi.org/10.1007/s00122-020-03684-z>. PMID:32964262.
- Sandhu, K. S., Lozada, D. N., Zhang, Z., Pumphrey, M. O., & Carter, A. H. (2021a). Deep learning for predicting complex traits in spring wheat breeding program. *Frontiers in Plant Science*, *11*, 613325. <http://dx.doi.org/10.3389/fpls.2020.613325>. PMID:33469463.
- Sandhu, K. S., Mihalyov, P. D., Lewien, M. J., Pumphrey, M. O., & Carter, A. H. (2021b). Combining genomic and phenomic information for predicting grain protein content and grain yield in spring wheat. *Frontiers in Plant Science*, *12*, 613300. <http://dx.doi.org/10.3389/fpls.2021.613300>. PMID:33643347.
- Shabannejad, M., Bihanta, M. R., Majidi-Hervan, E., Alipour, H., & Ebrahimi, A. (2020). A simple, cost-effective high-throughput image analysis pipeline improves genomic prediction accuracy for days to maturity in wheat. *Plant Methods*, *16*(1), 146. <http://dx.doi.org/10.1186/s13007-020-00686-2>. PMID:33292394.
- Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., Singh, R. P., & Poland, J. (2019). High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. *Frontiers in Plant*

- Science*, 10, 394. <http://dx.doi.org/10.3389/fpls.2019.00394>. PMID:31019521.
- Sun, J., Poland, J. A., Mondal, S., Crossa, J., Juliana, P., Singh, R. P., Rutkoski, J. E., Jannink, J. L., Crespo-Herrera, L., Velu, G., Huerta-Espino, J., & Sorrells, M. E. (2019). High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. *Theoretical and Applied Genetics*, 132(6), 1705-1720. <http://dx.doi.org/10.1007/s00122-019-03309-0>. PMID:30778634.
- Wallace, J. G., Rodgers-Melnick, E., & Buckler, E. S. (2018). On the road to breeding 4.0: Unraveling the good, the bad, and the boring of crop quantitative genomics. *Annual Review of Genetics*, 52(1), 421-444. <http://dx.doi.org/10.1146/annurev-genet-120116-024846>. PMID:30285496.
- Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *Theoretical and Applied Genetics*, 129(4), 653-673. <http://dx.doi.org/10.1007/s00122-016-2691-5>. PMID:26932121.
- Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., Wang, J., & Fan, J. (2019). Crop phenomics: Current status and perspectives. *Frontiers in Plant Science*, 10, 714. <http://dx.doi.org/10.3389/fpls.2019.00714>. PMID:31214228.