



Genomic selection in tropical perennial crops and plantation trees: a review

Essubalew Getachew Seyum · Ngalle Hermine Bille ·
Wosene Gebreselassie Abteu · Norman Munyengwa · Joseph Martin Bell ·
David Cros 

Received: 11 March 2022 / Accepted: 6 September 2022 / Published online: 20 September 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract To overcome the multiple challenges currently faced by agriculture, such as climate change and soil deterioration, more efficient plant breeding strategies are required. Genomic selection (GS) is crucial for the genetic improvement of quantitative traits, as it can increase selection intensity, shorten the generation interval, and improve selection accuracy for traits that are difficult to phenotype. Tropical perennial crops and plantation trees are of major economic importance and have consequently been the

subject of many GS articles. In this review, we discuss the factors that affect GS accuracy (statistical models, linkage disequilibrium, information concerning markers, relatedness between training and target populations, the size of the training population, and trait heritability) and the genetic gain expected in these species. The impact of GS will be particularly strong in tropical perennial crops and plantation trees as they have long breeding cycles and constrained selection intensity. Future GS prospects are also discussed. High-throughput phenotyping will allow constructing of large training populations and implementing of phenomic selection. Optimized modeling is needed for longitudinal traits and multi-environment trials. The use of multi-omics, haploblocks, and structural variants will enable going beyond single-locus genotype data. Innovative statistical approaches, like artificial neural networks, are expected to efficiently handle the increasing amounts of heterogeneous multi-scale data. Targeted recombinations on sites identified from profiles of marker effects have the potential to further increase genetic gain. GS can also aid re-domestication and introgression breeding. Finally, GS consortia will play an important role in making the best of these opportunities.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11032-022-01326-4>.

E. G. Seyum · N. H. Bille · J. M. Bell
Department of Plant Biology and Physiology, Faculty of Sciences, University of Yaoundé I, Yaoundé, Cameroon

E. G. Seyum · W. G. Abteu
Department of Horticulture and Plant Sciences, College of Agriculture and Veterinary Medicine, Jimma University, P.O. Box 307, Jimma, Ethiopia

N. Munyengwa
Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, QLD 4072, Australia

D. Cros (✉)
CIRAD, UMR AGAP Institut, 34398 Montpellier, France
e-mail: david.cros@cirad.fr

D. Cros
UMR AGAP Institut, CIRAD, INRAE, Univ. Montpellier, Institut Agro, 34398 Montpellier, France

Keywords Genomic predictions · Machine learning · Pangenomes · Genotype-by-environment interaction · Crop growth models · Reaction norms

Abbreviations

BLUP	Best linear unbiased prediction
CGM	Crop growth model
CNV	Copy number variation
GBLUP	Genomic BLUP
GEBV	Genomic estimated breeding value
GEGV	Genomic estimated genetic value
GEI	Genotype-by-environment interactions
GS	Genomic selection
GWAS	Genome-wide association study
HTP	High-throughput phenotyping
LD	Linkage disequilibrium
MAS	Marker-assisted selection
NIRS	Near-infrared spectroscopy
NGS	Next-generation sequencing
QTL	Quantitative trait locus
RKHS	Reproducing kernel Hilbert spaces
rrBLUP	Random regression BLUP
SNP	Single nucleotide polymorphism
SV	Structural variants

Introduction

The steady growth of the world population, expected to reach 9–11 billion by 2050, along with climate change and soil deterioration, are major challenges to achieving world food security (Kopittke et al. 2019; Rööß et al. 2017). Biotic and abiotic stresses caused by pathogens, animals, weeds, drought, extreme temperatures, flooding, salinity, acidic conditions, and nutrient starvation all reduce global agricultural productivity (Tyczewska et al. 2018). Plant breeding represents one of the main ways to alleviate these problems and improve both crop production and productivity (Bhat et al. 2016). Plant breeding uses two main approaches, conventional and molecular breeding. Conventional breeding mainly uses phenotypic data (Borrelli et al. 2015) and has several limitations, including the long time (> 10 years) needed to release a new variety, confounding environmental effects leading to low heritability for many traits of interest, particularly the most complex ones, like yield. Molecular plant breeding using DNA markers includes quantitative trait loci (QTL)-based marker-assisted selection (MAS) that can greatly increase the speed, efficiency, and precision of breeding compared to conventional methods (Gupta et al. 2010). However, QTL-based MAS is efficient only for traits

controlled by a few QTLs that have a major effect on trait expression, whereas for complex quantitative traits governed by a large number of minor QTLs, such as yield, it may be less efficient than conventional phenotypic selection (Bhat et al. 2016). For complex traits, the most efficient molecular breeding strategy available today is genomic selection (GS) (Hickey et al. 2019). GS is a form of MAS in which genetic markers covering the whole genome are used so that all QTL are in linkage disequilibrium (LD) with at least one marker (Goddard and Hayes 2007; Heffner et al. 2009; Isik 2014; Meuwissen et al. 2001). GS has emerged as one of the most promising selection strategies to enhance genetic gain per unit time and/or unit cost for both plant and animal breeding programs (Fugerey-Scarbel et al. 2021; Merrick et al. 2022; Mrode et al. 2019; Voss-Fels et al. 2019; Wartha and Lorenz 2021; Xu et al. 2020). In dairy cattle, GS doubled the rate of genetic progress (Wiggans et al. 2017). In plants, GS is progressively integrated into breeding schemes and is now routinely used for major crops, in particular in the private sector (Merrick et al. 2022; Varshney et al. 2017; Voss-Fels et al. 2019). For instance, GS played a key role in the development of drought-tolerant maize hybrids that gave higher yields under both favorable and water stress conditions in the western US Corn Belt (Merrick et al. 2022; Voss-Fels et al. 2019). GS has also been applied on a large scale at the International Maize and Wheat Improvement Center since 2010, where it is used in spring wheat to discard low-performing lines (Merrick et al. 2022).

The first step in GS is creating a training set (or training population). The training set is genotyped and phenotyped for the targeted traits, and a prediction model is then built using these genotypic and phenotypic data. Several high-throughput next-generation sequencing (NGS) technologies such as SNP arrays (LaFramboise 2009; Wang et al. 1998), genotyping-by-sequencing (Elshire et al. 2011), and whole-genome sequencing (Ni et al. 2017) platforms have facilitated the production of large amounts of single nucleotide polymorphism (SNPs) markers to use in GS, at an affordable cost. The target population is also genotyped but not phenotyped, and the prediction model calculates the genomic estimated breeding values (GEBVs) or, when non-additive effects are taken into account, the total genomic estimated genotypic values (GEGV) of the selection candidates

(Grattapaglia et al. 2018). The efficiency of GS is determined, in particular, by its accuracy, which is defined as the correlation between the predicted and the true (unknown) genetic value of the selection candidates (Lorenz et al. 2011). GS accuracy is affected by the effective size of the population, marker density and type, the size and structure of the training population, the genetic architecture of the traits, relatedness between the training and target population, LD between markers and QTLs, trait heritability, imputation method, etc. (Grattapaglia and Resende 2011; Isik 2014; Robertsen et al. 2019).

Tropical perennial crops and plantation trees are of huge importance for the human population, in particular for use as food, timber, pulp, and stimulant crops (Jamnadass et al. 2016). However, their productivity is generally well below their potential, in particular, due to biotic and abiotic constraints, as shown, for example, in Eucalyptus (Elli et al. 2019), oil palm (Pirker et al. 2016; Woittiez et al. 2017), coffee (Wang et al. 2015), and cocoa (Aneani and Ofori-Frimpong 2013). Applying more efficient breeding approaches to these species will help fill production gaps. Genomic selection is particularly attractive for perennial plant species as they have long generation intervals and low selection intensity. Isik (2014) showed that the impact of GS could be much greater in perennial forest trees than in any other crop or livestock breeding program. A significant number of articles on GS have already been published on a variety of traits of interest in several tropical perennial crops and plantation trees, for instance, yield in oil palm (Cros et al. 2017, 2015), rubber tree (Cros et al. 2019) and guava (Silva et al. 2021), growth in eucalyptus (Bouvet et al. 2016; Denis et al. 2012; Resende et al. 2012) and rubber tree (Souza et al. 2019), fruit quality in citrus (Minamikawa et al. 2017), resistance to diseases in cocoa (McElroy et al. 2018; Romero Navarro et al. 2017), etc. (Supplementary Table S1). However, a review of GS in these species is lacking. The objective of the present article is therefore to review the results of GS research in tropical perennial crops and plantation trees, to discuss the main factors affecting GS accuracy and to highlight the genetic gains expected in these species using this approach. We focus on perennial crops defined as such according to the FAO indicative crop classification (FAO 2015) and on plantation trees both grown in the tropics. The production of the corresponding

species include fruit, timber, pulp, latex, oil, nuts, and stimulants. To our knowledge, the species covered by published articles on GS so far are banana, guava, citrus, *Eucalyptus* species (*E. urophylla*, *E. grandis*, *E. benthamii*, *E. pellita*, and *E. robusta*), rubber tree, oil palm, jatropha, cacao, and coffee.

Factors affecting the accuracy of genomic selection

The correlation between the GEBVs and true breeding values is known as GS accuracy (r_{GS}), and it is a key parameter for breeders due to the linear correlation between selection accuracy and annual genetic gain R_y (Eq. (1)) (Grattapaglia et al. 2018):

$$R_y = \frac{i \times r \times \delta_A}{y} \quad (1)$$

where i is selection intensity, r is selection accuracy, δ_A is the additive genetic standard deviation, and y is the generation interval in years.

GS accuracy is usually obtained by k-fold cross-validation within a single experimental design (with each fold repeatedly used as a validation set and the remaining folds as the training set) or between experimental designs (with one site used for training and the other for validation), the latter being preferable as cross-validations may overestimate accuracy (Lorenz et al. 2011).

Below, we present sequentially the major factors that affect the accuracy of genomic predictions, although most factors are interconnected and their effects are not independent.

Statistical models for genomic prediction and trait genetic architecture

The whole-genome regression models used for genomic predictions deal with the “large p , small n ” problem that, in GS, concerns the number of markers that usually (largely) exceeds the number of data records, in contrast to multiple linear regressions that cannot be used without variable selection, which conflicts with the original goal of GS, i.e., avoiding marker selection and overfitting. Multiple linear regression results in an insufficient degree of freedom leading to poor prediction due to the inability to estimate all marker effects at

the same time, which is exacerbated by multicollinearity. A wide range of statistical methods has been developed for GS to alleviate this constraint (Campos et al. 2013; Jannink et al. 2010; Montesinos-López et al. 2021; Morota and Gianola 2014; Tong and Nikoloski 2021; Wang et al. 2018). They represent two broad categories: (i) parametric approaches, which mainly include methods that rely on the best linear unbiased prediction methodology (genomic BLUP [GBLUP] and random regression BLUP [RRBLUP]) and various Bayesian methods (Bayesian LASSO, BayesA, BayesB, etc.), and (ii) semi- and non-parametric approaches that fall into the machine learning category (reproducing kernel Hilbert spaces [RKHS], artificial neural networks, etc.). These methods differ in several ways: in terms of genetic assumptions and modeling of the genetic architecture of the traits (e.g., purely additive models, models that explicitly model dominance and/or epistatic effects, models with marker effects sampled from a common statistical distribution [RRBLUP, GBLUP], models with marker effects sampled from specific distributions [Bayesian LASSO, BayesB, etc.], models that implicitly model non-additive effects [e.g., RKHS]), in terms of computational approach (relationship-based methods and marker effect-based methods, single trait and multi-trait models, etc.), and in terms of the genomic information used in the model (type of polymorphisms, use of a priori information on markers, a combination of omics data, etc.).

The most widely used statistical approach for GS is GBLUP (Heslot et al. 2015; Montesinos-López et al. 2021), which combines linear mixed model analysis and genomic relationships. GBLUP derives from the first BLUP analyses applied in animal breeding to implement selection based on phenotypes and pedigree and that estimated the breeding values of individuals using the pedigree-based relationship matrix (A) (Henderson 1975), with a model of the form:

$$Y = X\beta + Zu + e \quad (2)$$

where Y is an $n \times 1$ vector of data records, X is an $n \times p$ incidence matrix relating data records with fixed effects, β is a $p \times 1$ vector of fixed effects, and Z is an $n \times q$ incidence matrix. u is a $q \times 1$ vector of random effects (i.e., breeding values), associated

with A , and e is an $n \times 1$ vector of residual effects. This initial approach we term pedigree-based BLUP (PBLUP) paved the way for GBLUP, which uses the genomic relationships (G) matrix, thus capturing existing relationships among individuals rather than expected relationships (Bernardo 1994; VanRaden 2007). An alternative approach to GBLUP is RRBLUP (Meuwissen et al. 2001), which yields GEBVs by estimating marker effects. GBLUP and RRBLUP are equivalent when there are many QTLs, when there is no major QTL, or when the QTLs are evenly distributed along the genome (Bernardo 2020). RRBLUP uses a model of the form:

$$Y = X\beta + Z'm + e \quad (3)$$

where Z' is an $n \times k$ incidence matrix giving the genotypes at k SNPs and m a $k \times 1$ vector of random SNP effects.

The relative performance of the different statistical methods is expected to vary depending on the genetic architecture of the trait considered (Lebedev et al. 2020). Genetic architecture corresponds to the genetic characteristics that determine the genotype–phenotype relationship, in particular, the number of genes that control the trait, the number of alleles per gene, the distribution of the genes along the genome, the distribution of the gene effects, and the mode of gene action (additive, dominant, epistatic) (Momen et al. 2018). Thus, methods in which marker effects are sampled in distributions where variance is the same for all markers (e.g., GBLUP, RRBLUP, Bayesian random regression) are expected to be more suitable for traits following the infinitesimal model, while methods with marker-specific variances (e.g., Bayesian LASSO, BayesB) are expected to be more suitable for traits whose genetic architecture includes major QTLs. Consequently, many GS studies, including those on tropical perennial fruit crops and plantation trees, use a range of statistical prediction methods to identify the most appropriate one for a specific trait. Overall, few variations have been found among statistical approaches, for example, in oil palm yield components (Cros et al. 2015; Kwong et al. 2017a), in eucalyptus growth (Durán et al. 2017; Müller et al. 2017), and in rubber tree latex yield (Cros et al. 2019). This confirms results obtained in empirical evaluations in other species, in which GS statistical methods were seen to perform similarly

(Heslot et al. 2015); however, in some cases, differences were found: e.g., BayesB performed best for several traits including vegetative growth, production, and disease resistance in banana (Nyine et al. 2018) and vegetative growth and oil yield in oil palm (Ithnin et al. 2017). This could mean that, in the populations considered, QTLs with large effects were segregated for these traits.

Similarly, when non-additive effects play a significant role in genetic variation, models that account for non-additive effects are expected to increase GS accuracy. In a simulation study, Denis and Bouvet (2013) showed that modeling dominance for the genomic predictions of the genetic value of eucalyptus clones improved accuracy when dominance effects were preeminent (ratio of dominance to the additive variance of 1.0) and heritability was high ($H^2=0.60$). With empirical data, also in eucalyptus, Resende et al. (2017), Tan et al. (2018), and Paludeto et al. (2021) showed that the use of GS models that account for dominance increased the accuracy of prediction for growth traits, which had high levels of dominance variance, whereas this was not the case for wood traits. In citrus, Minamikawa et al. (2017) showed that considering both additive and dominance effects improved prediction accuracy for acidity and juiciness.

When considering traits correlated with a sufficient magnitude but with contrasting levels of heritability, the use of multi-trait models can increase prediction accuracy for low heritability traits (Tong and Nikoloski 2021). In tropical perennial crops and plantation trees, the results obtained in oil palm (Marchal et al. 2016) and *Eucalyptus robusta* (Rambolarimanana et al. 2018) agreed with this principle. Multivariate models thus offer the opportunity to improve prediction accuracy at no extra cost (apart from increased computational resources), and they should therefore be systematically evaluated when correlations exist among the traits of interest, or between the traits of interest and secondary traits.

Machine learning methods are complex black-box approaches that are of growing interest for genomic predictions as they have several desirable features. They avoid the use of assumptions that are often violated and cannot be verified (Gianola and Van Kaam 2008), and they are particularly suitable to account for non-additive effects in particular in polyploids (Bayer et al. 2021) and to integrate data

from different biological sources for multi-omics predictions (Montesinos-López et al. 2021; Tong and Nikoloski 2021). RKHS is the most often evaluated machine learning approach for GS in tropical perennial crops and plantation trees. In bananas, RKHS was slightly more accurate than parametric approaches for a few traits (Nyine et al. 2018). In a study analyzing eight traits in *E. urophylla* × *E. grandis* eucalyptus hybrids, RKHS proved to be slightly more accurate in predicting low-heritability traits but less accurate in predicting pulp yield (Tan et al. 2017) and performed similarly to GBLUP for three traits in *E. grandis* (Rambolarimanana et al. 2018). A few other machine learning methods have been implemented in tropical perennial crops and plantation trees. Maldonado et al. (2020) compared several parametric prediction models, RKHS and two artificial neural network approaches, deep learning and Bayesian regularized neural networks, in *E. globulus* and maize, and found that predictions made with deep learning methods were significantly more accurate for all the traits considered. Sousa et al. (2020) compared several machine learning approaches and a parametric model to predict resistance to leaf rust in *Coffea arabica* and obtained the best accuracy with artificial neural networks. Several authors used random forest in oil palm and citrus and found that, on average over several traits, random forest performed no better than parametric approaches (Kwong et al. 2017b; Minamikawa et al. 2017). In oil palm, the support vector machine was found to be slightly better on average than other methods (Kwong et al. 2017b). Despite these uneven results in tropical perennial crops and plantation trees, machine learning should be further investigated, in particular as the training populations used so far were possibly not large enough for the optimal training of this type of approach (Montesinos-López et al. 2021). Particular attention should also be paid to artificial neural networks, which have produced promising results.

One limit to the differences among statistical methods and models in perennial fruit and tree crops reported so far is that they were not always supported by a statistical test indicating whether the differences were significant or not. This can be done, for example, using the Hotelling-Williams t -test (Steiger 1980).

Linkage disequilibrium and effective size

Linkage disequilibrium (LD) between markers and QTL and effective size (N_e) have interrelated effects that strongly influence GS accuracy (Heffner et al. 2009; Isik 2014; Lebedev et al. 2020). LD is defined as the non-random association of alleles at two or more loci in haplotypes (Slatkin 2008; Weir 1979). LD between two loci is measured based on the frequency of alleles, using indexes like D , D' , and r^2 (Collins and (Ed.) 2007). A key assumption in GS is that there is LD between QTLs and markers, such that, with dense genome marker coverage, every QTL controlling the phenotype of interest would be in LD with at least one marker. Good knowledge of this parameter in the target population is therefore of particular interest to define the marker density required for GS. It is thus useful to explore historical events, such as bottlenecks, genetic drift, and natural and artificial selection, that may have shaped the LD profile in the target population (Flint-Garcia et al. 2003; Gupta et al. 2005; Mackay and Powell 2007; Slatkin 2008). The LD profile is largely determined by the past N_e , which can be described as the number of randomly mating individuals in a population that would give rise to the observed rate of inbreeding (Falconer and Mackay 1996). There is an inverse relationship between N_e and LD, with high rates of genetic drift and inbreeding in low N_e populations leading to strong LD between markers and QTLs compared to high N_e populations (Grattapaglia 2014; Lin et al. 2014; Thistlethwaite et al. 2020). As N_e decreases and LD increases, pairs of individuals within the population tend to share longer haplotypes, enabling good genomic prediction accuracy (Clark et al. 2012; Heffner et al. 2009; Isik 2014; Lebedev et al. 2020). For a given marker density, training population size, and trait, LD and GS prediction accuracy is higher in populations with low N_e than in populations with high N_e (Grattapaglia 2014; Lin et al. 2014; Solberg et al. 2008).

The crucial role of LD and N_e in GS accuracy has also been underlined in studies on tropical perennial crops and plantation trees. Several studies investigated the LD profile to evaluate whether the marker density was high enough in citrus (Gois et al. 2016; Minamikawa et al. 2017), cocoa (McElroy et al. 2018), eucalyptus (Denis and Bouvet 2013; Durán et al. 2017; Müller et al. 2017), and oil palm

(Kwong et al. 2017a). Many studies in tropical perennial crops and plantation trees also investigated the efficiency of GS in populations with high LD/low N_e . This was possible using populations obtained through specific mating designs among a reduced number of parents (Denis and Bouvet 2013; Resende et al. 2012). In this way, Resende et al. (2012) found that in a population of eucalyptus where $N_e=11$ was obtained with an incomplete diallel, GS accuracy was higher for the four growth and wood quality traits studied than in the population where $N_e=51$, despite a slightly larger number of training individuals in the latter population. In other studies, high LD/low N_e was obtained in full-sib families GS (Cros et al. 2017; de Souza et al. 2018; Gois et al. 2016; Kwong et al. 2017b). This strategy is also applied in other crops as it maximizes GS accuracy, although at the cost of only applying to families comprising the training population (Crossa et al. 2017; Lebedev et al. 2020; Lin et al. 2014).

The fact that GS accuracy reaches a plateau when marker density reaches a certain level (see below) suggests that an appropriate strategy to filter the markers would increase the cost-efficiency of GS. Filtering SNPs on LD has been investigated in several studies, as the SNPs that show very high LD values provide redundant information. In oil palm, Kwong et al. (2017a) evaluated the impact of marker density reduction by LD filtering and noted that, for some traits, it was possible to reach the same GS accuracy as using all the SNPs.

Marker density and marker type

As marker density strongly affects the extent of LD, it also plays a major role in GS accuracy. In GS studies of both plants and animals, increasing the number of markers was shown to improve prediction accuracy until a plateau was reached (Isik 2014; Lin et al. 2014; Meuwissen et al. 2001; Robertsen et al. 2019; Solberg et al. 2008). The same trend was observed in tropical perennial crops and plantation trees, where the density of markers required to reach maximum prediction accuracy depends in particular on the type of population, trait, and marker. Romero Navarro et al. (2017) found increasing prediction accuracy for yield and disease traits in cocoa with increasing marker density before a plateau was reached at around 1000 markers. In the rubber tree,

the prediction accuracy for rubber yield plateaued at around 300 SRRs (Cros et al. 2019). In eucalyptus, the prediction accuracy among five growth and wood property traits reached a plateau between 5000 and 20,000 SNPs (Tan et al. 2017). Among seven production traits in oil palm hybrids, the plateau was reached with 500 to 2000 SNPs (Cros et al. 2017).

GS accuracy is also affected by the type of marker. Thus, in oil palm, GS accuracy for bunch number and average bunch weight plateaued at 160 SSRs in heterotic group A and at 90 SSRs in group B (Marchal et al. 2016) versus 3000 SNPs in group A and 350 SNPs in group B (Cros et al. 2017). This likely resulted from the fact that, as SNPs are biallelic, they are less informative than SSRs. However, in practice, SSRs cannot be used for genomic predictions, as GS relies on dense genotyping of large populations of selection candidates and therefore requires high throughput genotyping approaches at a reasonable cost. If marker density is constrained by the genotyping approach, the GS accuracy may be reduced. Thus, Kwong et al. (2017b) obtained mean GS prediction accuracies of 0.21 over palm oil yield components using 135 SSRs, versus 0.31 with 200 K SNPs.

Two primary options are available to reach the high marker density required for GS: methods that reduce genome complexity and SNP arrays (Edwards et al. 2013; Wiggans et al. 2017). They were made possible by the development of NGS technologies, which became available between 2004 and 2006 (Hu et al. 2021). Less expensive and with much higher throughput than the Sanger method (Sanger and Coulson 1975; Sanger et al. 1977), NGS methods have made it possible to carry out high-density and high-throughput genotyping, i.e., with good genome coverage in large populations, at an affordable cost. SNP arrays have been developed in several tropical perennial crops and plantation trees, with, for example, a 200 K array in oil palm (Kwong et al. 2016), a 60 K array in eucalyptus (Silva-Junior et al. 2015), and a 15 K array in cacao (McElroy et al. 2018). Most SNP genotyping methods based on reducing genome complexity consist of restriction enzyme-based approaches and sequence capture (Uitdewilligen et al. 2013; Zhou and Holliday 2012). These methods do not require specific preliminary investment and can be applied directly to any population. Given their relative simplicity and lower cost compared to SNP

arrays, they became widely used, in particular for introgression breeding, genome-wide association mapping (GWAS), and QTL mapping (see, e.g., Kitony et al. (2021) and Reyes et al. (2021) in rice, Pootakham et al. (2015) in oil palm, or Chia Wong et al. (2022) in cacao). However, they are associated with a higher rate of missing data and genotyping errors than SNP arrays. Despite these differences, it seems that the choice between these two types of approaches has no impact on GS accuracy: The accuracy of genomic prediction of 13 wood quality and growth traits in eucalyptus using SNP genotypes obtained with sequence capture and a 60 K SNP array was similar (de Moraes et al. 2018).

Training and validation population relatedness

The accuracy of GS is positively correlated with the relatedness between the training and test population (Daetwyler et al. 2013; Isidro y Sánchez J, Akdemir D 2021; Pszczola et al. 2012; Wientjes et al. 2013). This is because when pairs of genotypes are closely related, they tend to share long haplotype blocks in the same linkage phase. To limit allele duplication and redundancy, relationships within the training population should be minimized (Isidro y Sánchez J, Akdemir D 2021). The accuracy of GS in tropical perennial crops and plantation trees was also found to be affected by the relatedness between the training and test population. In two eucalyptus species, *E. benthamii* and *E. pellita*, Müller et al. (2017) found that prediction accuracy declined strongly for three growth traits when individuals were randomly assigned to the training and validation populations compared to when they were assigned using a principal component analysis to minimize relatedness between training and validation populations. Similarly, considering eight wood growth and quality traits in *Eucalyptus urophylla* × *E. grandis*, Tan et al. (2017) obtained the worst prediction accuracies when minimizing the relatedness between the training and validation populations using k-means clustering. In another study, a significant positive correlation was found between GS accuracy and the relationship between training and validation populations for various production traits in oil palm (Cros et al. 2015).

Size and design of the training population

The size of the training population is one of the most important factors that determine GS accuracy. Several GS studies have reported that increasing the size of the training population improves GS accuracy (Calleja-Rodriguez et al. 2020; Cericola et al. 2018; Combs and Bernardo 2013; Isidro et al. 2015; Liu et al. 2018; Nielsen et al. 2016; Tan et al. 2017). In a family of full-sibs of *Hevea brasiliensis*, Cros et al. (2019) reported an increase in the accuracy of GS for rubber yield with an increase in the size of the training population up to a plateau of 200 individuals. In Eucalyptus, Denis and Bouvet (2013) also reported an increase in GS accuracy as a result of increasing the size of the training population, and Tan et al. (2017) reported an increase in GS accuracy that followed a diminishing return trend with increasing size of the training population.

The possibility of assembling large training populations among tropical perennial crops and plantation trees is contrasted. Thus, training populations comprising more than 1000 individuals were used in eucalyptus (Mphahlele et al. 2021), cacao (McElroy et al. 2018), and oil palm (Kwong et al. 2017a), whereas only small populations (<600 individuals) have been used so far in banana (Nyine et al. 2018), rubber tree (Cros et al. 2019; Munyengwa et al. 2021; Souza et al. 2019), coffee (Fanelli Carvalho et al. 2020; Ferrão et al. 2019; Sousa et al. 2020, 2019, p. 2), jatropha (Peixoto et al. 2017), and guava (Silva et al. 2021). However, the size of the training population must be considered in relation to the relatedness between training and validation populations. Thus, for GS predictions in a biparental cross, it is better to use a relatively small but highly related training population of full-sibs or half-sibs than a large training population comprising distantly related or unrelated individuals (Brandariz and Bernardo 2019a; Brauner et al. 2020).

For some of the species considered here, breeding relies on a large number of phenotyped individuals, e.g., thousands of individuals for yield components and tolerance to ganoderma disease in oil palm (Cros et al. 2017; Daval et al. 2021) and thousands of individuals for tolerance to pests and diseases in *Eucalyptus grandis* (Mphahlele et al. 2021). In this case, genotyping a sample of the phenotyped population and making the genomic predictions using the single-step

GBLUP approach (Lourenco et al. 2020), i.e., using a training population combining the genomic data of the genotyped individuals and the genealogical data of the others, is an efficient way to maximize the cost-efficiency of GS; see Mphahlele et al. (2021) in *E. grandis*, Cappa et al. (2019) in a complex eucalyptus population, and Imai et al. (2019) in citrus.

The cost of phenotyping is a major constraint in GS, especially now that sequencing costs have dramatically decreased thanks to next-generation sequencing (Akdemir and Isidro-Sánchez 2019). This financial constraint is particularly applicable to perennial crops, as their phenotypic evaluation requires large surface areas over several years. Thus, training populations need to be optimized to improve the cost-effectiveness of GS in these species. Training population optimization is the process of selecting, within a pool of individuals that could be used to train the GS model, a sample of individuals that will best predict the genetic value of the selection candidates (Isidro y Sánchez J, Akdemir D 2021). Several methods have been developed to optimize the training population, including CD-mean, PEV-mean, stratified sampling, or EthAcc (Isidro y Sánchez J, Akdemir D 2021). This aspect has received little attention in tropical perennial crops and plantation trees, although in oil palm, Cros et al. 2015 confirmed the efficiency of training population optimization to improve GS accuracy.

Trait heritability

The broad-sense heritability of a trait (H^2) is defined as the proportion of the phenotypic variance that is genetically controlled. Narrow-sense heritability (h^2) considers only variations due to additive gene action and ignores non-additive (dominance and epistasis) genetic effects (Falconer and Mackay 1996). In GS studies, the heritability of the trait affects the accuracy of GEBV, with higher h^2 leading to greater GS accuracy (Hayes et al. 2009; Lin et al. 2014; Meuwissen et al. 2001, p. 2). This was illustrated by studies in tropical perennial crops and plantation trees where positive correlations were found between h^2 and GS prediction accuracy for a set of disease resistance and yield traits in cacao (Romero Navarro et al. 2017), eight palm oil production traits in the B heterotic group used in oil palm breeding (Cros et al. 2015), 18 Arabica coffee agronomic traits (Sousa et al., 2019),

and 15 vegetative growth, disease resistance, and fruit production traits in banana (Nyine et al. 2018). When simulating GS in eucalyptus, Denis and Bouvet (2013) noted that the prediction accuracy was higher with $H^2=0.6$ than with $H^2=0.1$, regardless of the ratio of dominance to additive variance, modeling dominance or not, or the breeding cycle. However, some studies detected no effect of trait heritability on GS prediction accuracy, but the effect may have been masked by other factors with stronger effects on prediction accuracy than heritability, in particular variations in the size of the training population, among traits, like in Durán et al. (2017).

Genetic gain from genomic selection

Genetic gain from the selection is defined as the improvement in the average genetic value of a population under the effect of selection over breeding cycles (Hazel and Lush 1942). GS has substantially increased genetic gain in animal breeding and plays a central role in many commercial plant breeding programs (Fugeray-Scarbel et al. 2021; Voss-Fels et al. 2019; Wartha and Lorenz 2021; Xu et al. 2020). The main advantages of GS over conventional phenotypic selection are its ability to (i) increase selection intensity and/or to shorten the generation interval by replacing all or part of the phenotyping activities by genotyping in selected breeding cycles and (ii) increase accuracy for traits that are difficult to phenotype (Fugeray-Scarbel et al. 2021; Wartha and Lorenz 2021).

When GS is used to increase selection intensity or to shorten the breeding cycle, an increase in annual genetic gain can be obtained even though GS is less accurate than conventional phenotypic evaluation. This has been illustrated in studies of tropical perennial crops and plantation trees that are promising for GS due to their long generation intervals and challenging phenotypic evaluations. Thus, based on the relative accuracy of GS and phenotypic selection, Resende et al. (2012, 2017) demonstrated that GS could significantly increase annual genetic gain for growth and wood quality traits in eucalyptus, i.e., from +50% to +300%, thanks to the fact that GS can be implemented at the seedling stage (<1 year), i.e., much earlier than phenotypic selection, which cannot be carried out before at least three years old.

Additionally, the possibility of increasing selection intensity by using a bigger population of selection candidates should further increase the advantage of GS over conventional selection. Based on 17 years of *E. grandis* breeding, Mphahlele et al. (2021) reported that the accumulated genetic gain with GS would be from 1.53 to 3.35 times higher than with conventional phenotypic selection, depending on the trait, because GS allows three breeding cycles in a 17-year period versus two with phenotypic selection. In coffee, it was also shown that with GS, 3-year breeding cycles would lead to a higher annual genetic gain in traits for growth, production, and tolerance to biotic stresses than the conventional 6-year phenotypic breeding cycles in *Coffea arabica* (Sousa et al. 2019) and in *Coffea canephora* (Alkimim et al. 2020). Similarly, an increase in annual genetic gain through a reduction in the generation interval with GS has been reported in citrus (Gois et al. 2016) and in rubber tree (Souza et al. 2019).

However, in many cases, the advantage of using GS over phenotypic selection in terms of genetic gain did not concern all the traits of interest. In this case, the interest of GS is its ability to increase selection intensity. This leads to a two-stage breeding scheme, starting with genomic selection, followed by phenotypic selection. In this case, the limiting factor for GS is the number of selection candidates that can be genotyped. In oil palm, using GS for bunch production before conventional phenotypic progeny tests was estimated to improve the performance of the selected A×B hybrids by more than 10% when 4000 A and 4000 B were genotyped (Cros et al. 2017). Similarly, in a full-sib rubber tree family, applying GS to 3000 individuals before clonal trials would have increased the selection response for rubber production by around 10% (Cros et al. 2019).

Some studies on tropical perennial crops and plantation trees also compared GS and QTL-based MAS approaches and the genetic gain expected from GS. For instance, in cacao, McElroy et al. (2018) found that GS largely outperformed GWAS in genetic gain for most of the disease resistance traits considered. In breeding populations of eucalyptus under selection, Müller et al. (2017) showed that GS outperformed GWAS for growth traits, as GS accounted for large proportions of the heritability, whereas GWAS captured very few significant associations. In a study simulating several cycles of within-family oil palm

breeding, Wong and Bernardo (2008) found that GS enabled higher annual genetic gains than marker-assisted recurrent selection for all the family sizes, number of QTLs, and heritability considered.

Future prospects for genomic selection in perennial tropical crops and plantation trees

Promising results have already been obtained with GS in tropical perennial crops and plantation trees. However, different aspects require further investigation to take full advantage of the approach. As mentioned above, statistical approaches for predictions still require attention; in particular, single-step GBLUP and multivariate models need to be more widely used and artificial neural networks need to be investigated in greater detail. Training populations also need optimization. Other promising aspects have hardly or not been studied at all so far for use with GS in tropical perennial crops and plantation trees, and these aspects are discussed below.

High-throughput phenotyping

High-throughput phenotyping (HTP) platforms allow faster phenotyping and reduced labor costs compared to conventional methods (Persa et al. 2021). HTP allows analyses at the field scale with outdoor platforms that use remote sensing and imaging, mostly based on visible/near-infrared and far-infrared spectroscopy, and analyses of the harvestable part of the crop using near-infrared reflectance spectroscopy (NIRS). The use of HTP has already led to significant results in model species such as rice, maize, and wheat, for a wide range of traits, like adaptation, quality, and vegetative growth (Asaari et al. 2019; Blancon et al. 2019; Chattopadhyay et al. 2019; Juliana et al. 2019; Sun et al. 2019; Wu et al. 2019). For GS, HTP is an efficient way to characterize large training populations (Wartha and Lorenz 2021). This is particularly useful for perennial species that require phenotyping over extended periods of time. HTP has already been used in different tropical perennial crops and plantation trees. For instance, multispectral data collected from an unmanned aerial vehicle were used to estimate the height and diameter at the breast height of eucalyptus trees (Borges et al. 2021). NIRS has also been used for rapid quantification of

flavor-related components of cocoa and beverage quality components of Arabica coffee (e.g., Álvarez et al. 2012; dos Santos Scholz et al. 2014). In eucalyptus populations used for GS, NIRS was used to measure chemical and physical wood quality traits (de Moraes et al. 2018; Durán et al. 2017; Rambolari-manana et al. 2018).

In addition to enabling the phenotyping of large populations, HTP data can be used in GS models as covariates associated with the trait of interest to increase prediction accuracy (Persa et al., 2021). To our knowledge, this aspect has not been investigated so far in GS studies on tropical perennial crops and plantation trees, but such studies would be of interest.

Phenomic selection is another approach that relies on spectral data that are usually obtained by NIRS (Rincint et al. 2018). In this case, the prediction of the genetic values is based on spectral data instead of molecular markers, meaning genomic data could no longer be needed. Phenomic selection has been investigated in a few crops, particularly in two temperate perennial species, poplar and grapevine. In poplar, the expected genetic gain using phenomic selection was higher than or the same as using genomic selection, depending on the trait (Rincint et al. 2018). In grapevine, phenomic predictions were reported to be a possible alternative to genomic predictions (Brault et al. 2022).

Longitudinal traits

Longitudinal traits are traits recorded repeatedly over the period of interest in the lifetime of individuals. This is a common case in perennial species. In tropical perennial crops and plantation trees, longitudinal traits are, for instance, growth and production, which are evaluated on each plant at different ages. The random regression model, a standard approach used for the genetic analysis of such traits (Oliveira et al. 2019), is a mixed model that makes it possible to model individual genetic values as a continuous function of time (or environmental covariates, see below), which can lead to more accurate estimates of the genetic values and facilitate the selection of genotypes with an optimal profile over the period of interest. Random regression can link genetic effects and time with complex functions, including nonlinear patterns, without making assumptions about the shape of the curve (Mrode 2014; Oliveira et al. 2019).

The parameters that characterize these functions (e.g., slopes and intercepts for linear functions) are treated as random effects, and the analysis yields genotype-specific parameters. Random regression has already been used for genomic predictions of longitudinal traits in different species, in particular in animals (Oliveira et al. 2019). Surprisingly, even though many traits in tropical perennial crops and plantation trees are longitudinal, random regression has rarely been used in these species. One example is *Jatropha curcas*, where random regression was used to analyze grain yield over the years (Peixoto et al. 2020). However, to our knowledge, this approach has not been used in the context of GS in tropical perennial crops and plantation trees so far.

Leveraging multi-environment trials

Multi-environment trials and GS models that account for environmental effects make it possible to predict the genetic value of new genotypes in known environments, known genotypes in new environments, and new genotypes in new environments (Bustos-Korts et al. 2016; Malosetti et al. 2016). The ability to predict the performances in new environments is of major interest in the context of climate change, in particular for perennial crops where breeding suffers from inertia due to the length of the breeding cycles. Analysis of genotype-by-environment interactions (GEI) helps select genotypes that are stable across environments and can identify the best genotypes for specific target environments. In particular, this has been extensively studied in cereals (Crossa et al. 2017). Considering GEI in GS models can significantly increase prediction accuracy when data from multi-environment trials are available (Tong and Nikoloski 2021; Xu et al. 2020). A variety of approaches have been developed to incorporate environmental data in GS models (Bustos-Korts et al. 2016; Crossa et al. 2017; Malosetti et al. 2016; Tong and Nikoloski 2021; Xu et al. 2020). The most attractive methods enable predictions in new environments using reaction norms (Costa-Neto et al. 2021; Costa-Neto and Fritsche-Neto 2021; Crossa et al. 2021) or crop growth models (CGM) (Crossa et al. 2021; Van Eeuwijk et al. 2019; Xu et al. 2020).

Reaction norms are linear or nonlinear functions that describe the phenotypes produced by a single genotype across an environmental gradient (Li

et al. 2017). They can be incorporated into genetic analyses using random regression (Marchal et al. 2019; Mrode 2014; Oliveira et al. 2019), leading to genotype-specific coefficients that characterize random norms for each environmental covariate. Equivalently, the environmental covariates can be used to build an environmental relationship matrix that identifies putative similarities among the environments considered (Costa-Neto et al. 2021), rather like using SNPs to build the relationship matrix.

CGM relies on plant physiology, soil science, and climatology principles to model plant development. CGMs use equations involving genetic parameters that are specific to the genotypes under consideration and are assumed to be independent of the environment and environmental variables (Boote et al. 2013). Several methods have been developed to incorporate CGM in the context of GS (Crossa et al. 2021; Rincent et al. 2017). CGM can be implemented to predict developmental stages that – along with daily weather data – will be used to compute climate stress covariates according to the plant development stage. CGM can also be used to compute environmental stress covariates that include the response of the crop to environmental conditions. These environmental covariates can then be incorporated in the GS model using, for example, random regression. Alternatively, the genetic parameters of the CGM can be estimated for the genotypes that comprise the training set and the genetic parameters of the selection candidates predicted by a GS model. Using the CGM and environmental covariates makes it possible to predict the phenotype of the selection candidates in the target environment. This approach has been termed gene-based modeling. Another method consists of incorporating a CGM in the GS prediction framework for the joint estimation of marker effects and CGM genetic parameters. This is referred to as CGM-WGP (whole-genome predictions) and relies on the use of approximate Bayesian computation or Bayesian generalized linear hierarchical models.

Ideally, the use of reaction norms or CGM requires the identification of all the environmental covariates that affect the trait of interest and the availability of environmental data at the plant level. This refers to the concept of envirotyping (Xu 2016) and its extension to large scale across time and space and environments (Resende et al. 2021). To our knowledge, only two GS studies have considered multi-environment

trials in tropical perennial crops and plantation trees so far. Souza et al. (2019) made genomic predictions obtained with multi-environment data and modeling approaches including environmental effects and GEI applied to rubber trees grown in two environmental conditions. These authors showed that multi-environment models captured a larger proportion of the genetic variance than single-environment approaches. In *Coffea canephora*, Ferrão et al. (2019) used multiplicative models in which genetic and environmental effects were handled in a common random effect associated with a variance–covariance matrix obtained by the Kronecker product of genetic and environmental variance–covariance matrices. These authors showed that this approach resulted in more accurate GS than traditional GBLUP, as the latter did not account for environmental information. This area of GS needs further study in tropical perennial crops and plantation trees, and particular attention should be paid to the use of CGM, reactions norms, and enviromics. This could leverage tools and skills that are already

available in these species. Thus, crop growth models have already been developed, for example, in cocoa (Zuidema et al. 2005), oil palm (Huth et al. 2014), and eucalyptus (de Freitas et al. 2020), and reaction norms were constructed in arabica coffee (Bertrand et al. 2015) and used with random regression for GEI analysis in conventional eucalyptus breeding (Alves et al. 2020).

Beyond single-locus genotype data

Different types of molecular information can now be exploited by the GS model, which could lead to an increase in the accuracy of predictions by better modeling the genotype–phenotype relationship (Fig. 1).

The use of haploblocks made of two or more adjacent SNPs instead of single SNPs was investigated for genomic predictions, as it could increase GS accuracy by better capturing identity-by-descent between individuals, giving higher LD between QTLs and haploblock alleles, or capturing epistatic effects between

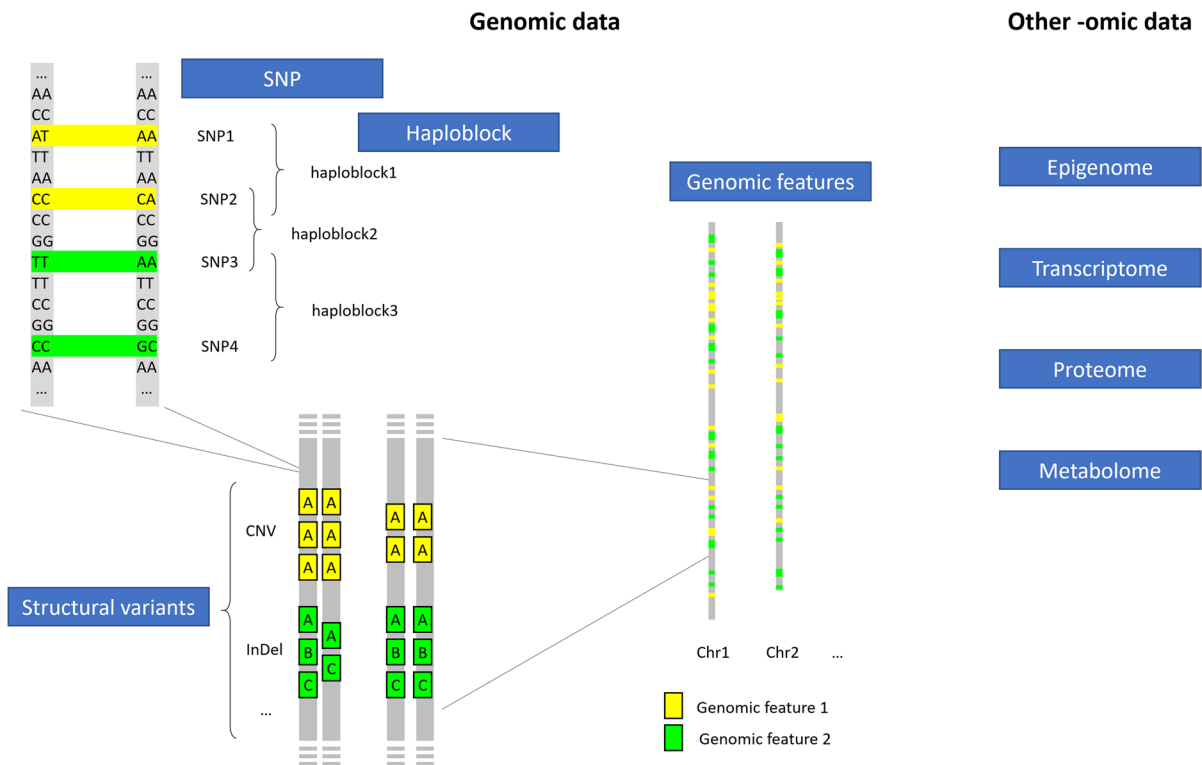


Fig. 1 Overview of possible molecular information for optimizing GS models. Genomic features can be defined in various ways: location in QTL, functional and structural annotations,

etc. (Sørensen et al. 2013). Two genomic features were considered here for illustration

SNPs in the same haploblock (Bhat et al. 2021; Goddard and Hayes 2007; Hess et al. 2017). Ballesta et al. (2019) explored the advantages of using haplotypic data for GS in *Eucalyptus globulus* and showed that prediction accuracy was significantly higher for low heritable traits when haploblocks were used instead of single SNPs. However, the relative efficiency of using haploblocks or single SNPs for genomic predictions is affected by many parameters, in particular the size of the training population, the level of LD, the method used to define the haploblocks, and the phasing accuracy (Bhat et al. 2021; Goddard and Hayes 2007; Hess et al. 2017). This aspect requires further investigation in tropical perennial crops and plantation trees.

The use of pangenomes is another possible avenue of GS research. Progress in sequencing techniques has enabled the comparison of individual genomes within species and shown that structural variations (SV) represent a significant proportion of polymorphism (Yuan et al. 2021). SVs consist of deletions, insertions, copy number variations, inversions, or translocations, with size > 50 bp. In particular, SVs include variations in gene presence/absence, with core genes that are found in all individuals and variable genes that are absent in some individuals. SVs cannot be represented by single reference genomes, and pangenomes are thus required to harness the whole genetic diversity of the breeding population (Bayer et al. 2021; Scossa et al. 2021). So far, very few studies have considered using structural variations for genomic predictions. In wheat, Würschum et al. (2017) obtained a slight increase in GS accuracy when markers specifically targeting a CNV contributing to the genetic control of the target trait were included in the model. Similarly, in maize and cattle, the use of CNV information in the GS model increased prediction accuracy in some cases (El Hamidi et al. 2018; Lyra et al. 2019). The use of SV information for genomic predictions deserves greater attention, and this will be greatly facilitated by pangenomes. Several reference genomes are already available for certain tropical perennial crops and plantation trees (e.g., cocoa and oil palm), and the next step should be the construction of pangenomes. The biggest impact could be on polyploid crops, such as bananas, as SV may represent an even higher proportion of polymorphisms in polyploids (Schissel et al. 2019).

Another way of improving GS accuracy is to incorporate existing information concerning polymorphisms, particularly that obtained from studies of QTL detection, in the prediction model (Xu et al. 2020). Different modeling approaches have been developed for this purpose, and their efficiency has been demonstrated in animal and plant studies, including temperate perennial fruit trees (Nsibi et al. 2020). However, very few studies have investigated this aspect in tropical perennial crops and plantation trees so far. In oil palm, Kwong et al. (2017a) applied RRBLUP using only SNPs with the highest GWAS association score, which made it possible to reduce marker density while achieving better or the same accuracy as using all the SNPs. A similar result was obtained in eucalyptus (Tan and Ingvarsson 2019). However, these approaches depend on a careful definition of the training and application populations. Thus, in cocoa, the inclusion of the SNPs detected by GWAS as fixed effects in the GS model did not improve prediction accuracies, which likely resulted from a too high genetic differentiation between the training and application populations, making the detected SNPs irrelevant (McElroy et al. 2018).

Incorporating endophenotypes, or intermediate phenotypes, in prediction models is another promising feature of GS research. Endophenotypes, and in particular transcriptomic and metabolomic data, have been used jointly with genomic data in a few crops (Scossa et al. 2021; Tong and Nikoloski 2021; Xu et al. 2020). These multi-omics prediction approaches are expected to better capture minor and non-additive effects and to better model the relationship between genotypes and phenotypes. Multi-omics prediction produced promising results in rice and maize, where they outperformed single-omic predictions. This requires specific statistical approaches, like machine learning (Montesinos-López et al. 2021; Tong and Nikoloski 2021). Investigating these aspects would be of interest to tropical perennial crops and plantation trees.

GS aided re-domestication and introgression breeding

Some perennial tropical crops have breeding populations with narrow genetic bases, and hence, only a fraction of the genetic diversity of the species is exploited, for instance, in *Coffea Arabica* (Tran

et al. 2016), cacao (Lanaud et al. 2001; Zhang and Motilal 2016), and rubber (Priyadarshan 2011). This usually resulted from choices and constraints dating back to the beginning of the breeding of these crops, or even before. In addition, the criteria originally used to select individuals might differ from the criteria that are of interest today, and current breeding populations may no longer correspond to current needs in terms of diversity. For example, in oil palm, the Deli breeding population, which today is used as one of the two heterotic populations mated to produce the vast majority of the oil palm cultivars, originated from four individuals collected in Africa and planted in Indonesia in 1848, decades before the establishment of the first commercial plantations (Corley and Tinker 2016). The other oil palm breeding populations derived from a small number of founders selected among individuals collected in restricted regions during prospecting, usually in the first half of the twentieth century. Although this led to reduced effective sizes (Cros et al. 2014), which is advantageous for GS accuracy, it constrains the long-term genetic gain. Also, for the La Mé oil palm breeding population, the founder individuals were selected in the 1920s, giving less importance to the proportion of pulp in the fruits than breeders do today (Cochard 2008). Although this has not prevented significant genetic progress (e.g., in oil palm, genetic progress is considered to be 1–1.5% per year (Rival and Levang 2014), and in rubber tree, yield increased from 500 kg ha⁻¹ in primary clones developed in the 1930–1960 period to 2500 kg ha⁻¹ in the best clones today (Priyadarshan 2011)), broader genetic diversity of the crops concerned would help maintain the rate of the genetic progress and likely increase it. This could be achieved through the re-domestication of existing crops (Tian et al. 2021), which consists in initiating breeding afresh from a renewed and broader diversity comprising ancestors and/or natural populations of existing crops. Introgression breeding could also play an important role in increasing genetic diversity by transferring exotic alleles from the related species of cultivated crops (Gramazio et al. 2021). GS is an attractive way of implementing these processes efficiently (Crossa et al. 2017). Indeed, re-domestication or introgression breeding of perennial tropical crops and plantation trees would normally require many decades of phenotypic selection, making GS a particularly attractive option. One example is already

available in a temperate perennial fruit tree, apple (Kumar et al. 2020), a study which suggested that, for the introgression of monogenic traits into a superior germplasm by backcrosses or pseudo-backcrosses, GS would be efficient for the background selection implemented among the individuals that inherited the trait of interest from the exotic donor germplasm, as it would accelerate the elimination of the unwanted alleles of the donor, compared to conventional phenotypic background selection. The use of GS for this purpose should be considered in perennial tropical crops and plantation trees where introgression breeding from wild species has already been shown to be of interest, including citrus, banana, and cacao (Scossa et al. 2016).

Combining profiles of predicted marker effects and targeted recombination

As mentioned above, one limiting factor in breeding perennial crops is the constrained size of the population of selection candidates, as the larger the population, the more exhaustive the search for elite individuals within the diversity generated by meiosis. GS makes it possible to increase the population of selection candidates by replacing phenotyping with genotyping. Controlling the gametes generated at meiosis could further increase the efficiency of the breeding scheme. This could be made possible by combining genome-wide profiles of marker effects estimated using GS models and targeted recombination (Bernardo 2017). The profiles of marker effects along the chromosomes of heterozygote individuals could be used to identify sites in the genome where recombinations would maximize the genetic value of their gametes by aggregating blocks of favorable alleles. Recombinations could be obtained at these sites through genome editing, and the progenies of the regenerated edited individuals were screened to identify the best ones. This approach has great potential to increase genetic progress (Bernardo 2017; Brandariz and Bernardo 2019b). Genome editing tools are under active development in perennial tropical crops and plantation trees, for example, in cacao (Fister et al. 2018) and oil palm (Yeap et al. 2021). However, further studies are required in these species to develop efficient, targeted recombination approaches and to evaluate the relative efficiency of breeding

schemes involving targeted recombinations and conventional schemes.

GS-based breeding consortia

Breeding for perennial crops is highly complex and very costly, and only limited resources are available for breeding many tropical perennials. Furthermore, as we have seen throughout this review, using GS requires expertise in a range of scientific and technical fields, including quantitative genetics, biostatistics, bioinformatics, genomics, computer programming, and, in particular, with the growing interest in machine learning, mathematics. GS also often requires a large training population which, in the context of climate change, will need to be evaluated in multiple environments. This puts tropical perennial crops in a completely different situation than many other crops including temperate cereals and legumes that can rely on a dynamic private sector to bring together the required human resources, phenotyping and genotyping capacities, etc. and to make rapid progress in innovative methods, resulting in the release of cultivars that have benefited from these methods. One possible solution for tropical perennial crops would be to strengthen international collaboration by sharing the efforts required for the practical implementation of GS, i.e., multi-environment phenotyping, high-throughput genotyping, and statistical analyses for genomic predictions. Sneller et al. (2021) called for the construction of GS-based breeding consortia, which would allow each member of a consortium to share the overall GS costs while predicting the genetic value of its selection candidates using a large training population comprising genetic material from all the consortium partners. Another advantage of such consortia would be the possibility to evaluate genetic material in different environments through the exchange of plant material among the consortium partners. Even so, there would have to be some relatedness between the plant material shared by the members of the consortium, and sufficient genotypes would have to be evaluated in different partners' environments (Sneller et al. 2021). Such a consortium is a possible solution for the implementation of GS for tropical perennial species on which, to our knowledge, no GS studies have been published so far, including coconut, papaya, avocado, mango, or teak, despite their major economic importance.

Projects in this sense are currently being set up for some perennial tropical crops and plantation trees, like coffee (World Coffee Research 2022), while others could emerge by building on existing networks, like MusaNet (<https://musanet.org/>) and CacaoNet (<https://www.cacaonet.org/>).

Conclusion

Genomic selection (GS) should revolutionize the breeding of perennial tropical crops and plantation trees as it has already produced promising results in terms of an increase in the rate of genetic progress. GS will (i) enable increased selection intensity and/or a shorter generation interval by replacing all or some phenotyping by genotyping in selected breeding cycles and (ii) increase accuracy for traits that are difficult to phenotype. Overall, the main factors that affect GS accuracy have been well studied in perennial tropical crops and plantation trees. However, the level of studies on GS varied in the following species: Some, like eucalyptus and oil palm, can be considered as models for GS including an in-depth assessment of its practical potential; in others, like banana and guava, GS studies were recently initiated, while in other species, like coconut, papaya, avocado, mango, and teak, despite their economic importance, no GS study has been conducted so far.

The results obtained in the plant and animal species where GS has been investigated to date suggest that optimal GS predictions could be achieved through joint analysis of all available information concerning genotype-to-phenotype relations, possibly including multiple omics and phenotypic data on multiple traits in several well-characterized environments, using prior information available on markers and all types of polymorphisms present in the populations concerned. For perennial crops, in which phenotyping is particularly complex and resource-consuming, there is an urgent need for increased international cooperation in the form of GS-based consortia to be able to gather such large datasets at a reasonable cost. The optimal implementation of GS will also require going beyond the standard GS technologies and methodologies used today. In particular, high-throughput phenotyping is a key approach to gathering the required amount of phenotypic data on such large populations at a reasonable rate and cost. Statistical methodologies able to handle

large multidimensional heterogeneous datasets are also required, and machine learning approaches are crucial, particularly artificial neural networks.

Future GS research in tropical perennial crops and plantation trees should systematically consider the use of single-step GBLUP when phenotypic data are available on ungenotyped individuals, the use of multivariate models when the traits of interest comprise correlated traits with contrasting levels of heritability, and random regression models for longitudinal traits. Training population optimization should also be undertaken. Targeted recombinations on sites identified based on the profiles of predicted marker effects should be investigated. Furthermore, GS has the potential to make re-domestication possible as well as to boost introgression breeding.

Acknowledgements The authors acknowledge the GENES program of the Intra-Africa Academic Mobility Scheme of the European Union for financial support (EU-GENES:2017-2552/001-001). The authors also thank Marie Denis, Gilles Trouche, André Clément-Demange, Angélique D'Hont, Dominique Dessauw, and Xavier Argout for discussions that improved the manuscript.

Author contribution EGS and DC carried out the literature review and wrote the manuscript, with help from WGA, NHB, NM, and JMB. All authors read and approved the final manuscript.

Funding This study was funded by the GENES Intra-Africa Academic Mobility Scheme of the European Union (EU-GENES:2017-2552/001-001) program, by CIRAD, and by a grant from PalmElit SAS.

Data availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent for publication Not applicable.

Competing interests We declare that all authors do not have any kind of financial or non-financial interests that are directly or indirectly related to this review article.

References

- Akdemir D, Isidro-Sánchez J (2019) Design of training populations for selective phenotyping in genomic prediction. *Sci Rep* 9:1–15
- Alkimim ER, Caixeta ET, Sousa TV, Resende MDV, da Silva FL, Sakiyama NS, Zambolim L (2020) Selective efficiency of genome-wide selection in *Coffea canephora* breeding. *Tree Genet Genomes* 16:1–11
- Álvarez C, Pérez E, Cros E, Lares M, Assemat S, Boulanger R, Davrieux F (2012) The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (–)-epicatechin contents in unfermented and sun-dried beans of Criollo cocoa. *J near Infrared Spectrosc* 20:307–315
- Alves RS, de Resende MDV, Azevedo CF, de Rocha JRAS-CDO, Nunes ACP, Carneiro APS, dos Santos GA (2020) Optimization of Eucalyptus breeding through random regression models allowing for reaction norms in response to environmental gradients. *Tree Gen Genomes* 16:1–8
- Aneani F, Ofori-Frimpong K (2013) An analysis of yield gap and some factors of cocoa (*Theobroma cacao*) yields in Ghana. *Sustainable Agriculture Research* 2: 2:526–2016–37857
- Asaari MSM, Mertens S, Dhondt S, Inzé D, Wuyts N, Scheunders P (2019) Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Comput Electron Agric* 162:749–758
- Ballesta P, Maldonado C, Pérez-Rodríguez P, Mora F (2019) SNP and haplotype-based genomic selection of quantitative traits in *Eucalyptus globulus*. *Plants (basel)* 8:331. <https://doi.org/10.3390/plants8090331>
- Bayer PE, Peteret J, Danilevicz MF, Anderson R, Batley J, Edwards D (2021) The application of pangenomics and machine learning in genomic selection in plants. *Plant Genome* 14:e20112
- Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci* 34:20–25
- Bernardo R (2020) Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. *Heredity* 125:375–385. <https://doi.org/10.1038/s41437-020-0312-1>
- Bernardo R (2017) Prospective targeted recombination and genetic gains for quantitative traits in maize. *The Plant Genome* 10(2):1–9. <https://doi.org/10.3835/plantgenome2016.11.0118>
- Bertrand B, Bardil A, Baraille H, Dussert S, Doullbeau S, Dubois E, Severac D, Dereeper A, Etienne H (2015) The greater phenotypic homeostasis of the allopolyploid *Coffea arabica* improved the transcriptional homeostasis over that of both diploid parents. *Plant Cell Physiol* 56:2035–2051
- Bhat JA, Yu D, Bohra A, Ganie SA, Varshney RK (2021) Features and applications of haplotypes in crop breeding. *Commun Biol* 4:1–12
- Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, Tyagi A, Mushtaq M, Jain N, Singh PK, Singh GP, Prabhu KV (2016) Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front Genet* 7:221. <https://doi.org/10.3389/fgene.2016.00221>
- Blancón J, Dutartre D, Tixier M-H, Weiss M, Comar A, Praud S, Baret F (2019) A high-throughput model-assisted method for phenotyping maize green leaf area index

- dynamics using unmanned aerial vehicle imagery. *Front Plant Sci* 10:685
- Boote KJ, Jones JW, White JW, Asseng S, Lizaso JI (2013) Putting mechanisms into crop production models. *Plant, Cell Environ* 36:1658–1672
- Borges MVV, de Oliveira Garcia J, Batista TS, Silva ANM, Baio FHR, da Silva Junior CA, de Azevedo GB, de Oliveira Sousa Azevedo GT, Teodoro LPR, Teodoro PE (2021) High-throughput phenotyping of two plant-size traits of Eucalyptus species using neural networks. *J Forest Res* 33(2):591–599. <https://doi.org/10.1007/s11676-021-01360-6>
- Borrelli GM, Orrù L, Vita PD, Barabaschi D, Mastrangelo AM, Cattivelli L (2015) Chapter 18 - Integrated views in plant breeding: from the perspective of biotechnology. In: Sadras VO, Calderini DF (eds), *Crop Physiology*, 2nd edn, vol 2. Academic Press, San Diego pp 467–486. <https://doi.org/10.1016/B978-0-12-417104-6.00018-2>
- Bouvet J-M, Makouanzi G, Cros D, Vigneron P (2016) Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. *Heredity* (edib) 116:146–157. <https://doi.org/10.1038/hdy.2015.78>
- Brandariz SP, Bernardo R (2019a) Small ad hoc versus large general training populations for genomewide selection in maize biparental crosses. *Theor Appl Genet* 132:347–353. <https://doi.org/10.1007/s00122-018-3222-3>
- Brandariz SP, Bernardo R (2019b) Predicted genetic gains from targeted recombination in elite biparental maize populations. *Plant Genome* 12:180062. <https://doi.org/10.3835/plantgenome2018.08.0062>
- Brault C, Lazerges J, Doligez A, Thomas M, Ecartot M, Roumet P, Bertrand Y, Berger G, Pons T, François P, Le Cunff L, This P, Segura V (2022) Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods* 18:108. <https://doi.org/10.1186/s13007-022-00940-9>
- Brauner PC, Müller D, Molenaar WS, Melchinger AE (2020) Genomic prediction with multiple biparental families. *Theor Appl Genet* 133:133–147
- Bustos-Korts D, Malosetti M, Chapman S, van Eeuwijk F (2016) Modelling of genotype by environment interaction and prediction of complex traits across multiple environments as a synthesis of crop growth modelling, genetics and statistics. *Crop Systems Biology* 55–82. https://doi.org/10.1007/978-3-319-20562-5_3
- Calleja-Rodríguez A, Pan J, Funda T, Chen Z, Baison J, Isik F, Abrahamsson S, Wu HX (2020) Evaluation of the efficiency of genomic versus pedigree predictions for growth and wood quality traits in Scots pine. *BMC Genomics* 21:796. <https://doi.org/10.1186/s12864-020-07188-4>
- Cappa EP, de Lima BM, da Silva-Junior OB, Garcia CC, Mansfield SD, Grattapaglia D (2019) Improving genomic prediction of growth and wood traits in Eucalyptus using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Sci* 284:9–15
- Cericola F, Lenk I, Fè D, Byrne S, Jensen CS, Pedersen MG, Asp T, Jensen J, Janss L (2018) Optimized use of low-depth genotyping-by-sequencing for genomic prediction among multi-parental family pools and single plants in perennial ryegrass (*Lolium perenne* L.) *Front Plant Sci* 9:369. <https://doi.org/10.3389/fpls.2018.00369>
- Chattopadhyay K, Behera L, Bagchi TB, Sardar SS, Moharana N, Patra NR, Chakraborti M, Das A, Marndi BC, Sarkar A (2019) Detection of stable QTLs for grain protein content in rice (*Oryza sativa* L.) employing high throughput phenotyping and genotyping platforms. *Sci Rep* 9:1–16
- Chia Wong JA, Clement DPL, Mournet P, dos Santos Nascimento A, Solís Bonilla JL, Lopes UV, Pires JL, Gramacho KP (2022) A high-density genetic map from a cacao F2 progeny and QTL detection for resistance to witches' broom disease. *Tree Genet Genomes* 18:1–14
- Clark SA, Hickey JM, Daetwyler HD, van der Werf JH (2012) The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet Sel Evol* 44:1–9. <https://doi.org/10.1186/1297-9686-44-4>
- Cochard B., 2008. Etude de la diversité génétique et du déséquilibre de liaison au sein de populations améliorées de palmier à huile (*Elaeis guineensis* Jacq.).
- Collins AR (ed) (2007) Linkage disequilibrium and association mapping: analysis and applications, methods in molecular biology. Humana Press, p 376. <https://doi.org/10.1007/978-1-59745-389-9>
- Combs E, Bernardo R (2013) Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *The Plant Genome* 6(1):7. <https://doi.org/10.3835/plantgenome2012.11.0030>
- Corley RHV, Tinker PB (2016) *The oil palm*, 5th edn. Wiley-Blackwell, Chichester. UK
- Costa-Neto G, Fritsche-Neto R (2021) Enviromics: bridging different sources of data, building one framework. *Crop Breed Appl Biotechnol* 21:1–14. <https://doi.org/10.1590/1984-70332021v21Sa25>
- Costa-Neto G, Galli G, Carvalho HF, Crossa J, Fritsche-Neto R (2021) EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. *G3* 11, jk040.
- Cros D, Sánchez L, Cochard B, Samper P, Denis M, Bouvet J-M, Fernández J (2014) Estimation of genealogical coancestry in plant species using a pedigree reconstruction algorithm and application to an oil palm breeding population. *Theor Appl Genet* 127:981–994. <https://doi.org/10.1007/s00122-014-2273-3>
- Cros D, Denis M, Sánchez L, Cochard B, Flori A, Durand-Gasselin T, Nouy B, Omoré A, Pomiès V, Riou V, Suryana E, Bouvet J-M (2015) Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). *Theor Appl Genet* 128:397–410. <https://doi.org/10.1007/s00122-014-2439-z>
- Cros D, Bocs S, Riou V, Ortega-Abboud E, Tisné S, Argout X, Pomiès V, Nodichao L, Lubis Z, Cochard B (2017) Genomic preselection with genotyping-by-sequencing increases performance of commercial oil palm hybrid crosses. *BMC Genomics* 18:1–17
- Cros D, Mbo-Nkoulou L, Bell JM, Oum J, Masson A, Soumahoro M, Tran DM, Achour Z, Le Guen V, Clement-Demange A (2019) Within-family genomic selection in rubber tree (*Hevea brasiliensis*) increases genetic gain for rubber production. *Ind Crops Prod* 138:111464
- Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Campos los G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker

- S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney RK (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci* 22:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Crossa J, Fritsche-Neto R, Montesinos-Lopez OA, Costa-Neto G, Dreisigacker S, Montesinos-Lopez A, Bentley AR (2021) The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. *Front Plant Sci* 12:651480. <https://doi.org/10.3389/fpls.2021.651480>
- Daetwyler HD, Calus MPL, Pong-Wong R, Campos G, de Hickey los JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193:347–365. <https://doi.org/10.1534/genetics.112.147983>
- Daval A, Pomiès V, Le Squin S, Denis M, Riou V, Breton F, Bink M, Cochard B, Jacob F, Billotte N (2021) In silico mapping in an oil palm breeding program reveals a quantitative and complex genetic resistance to *Ganoderma boninense*. *Mol Breed* 41(9):1–18. <https://doi.org/10.1007/s11032-021-01246-9>
- de Campos los G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327–345. <https://doi.org/10.1534/genetics.112.143313>
- de Freitas ECS, de Paiva HN, Neves JCL, Marcatti GE, Leite HG (2020) Modeling of eucalyptus productivity with artificial neural networks. *Ind Crops Prod* 146:112149
- de Moraes BFX, dos Santos RF, de Lima BM, Aguiar AM, Missiaggia AA, da Costa Dias D, Rezende GDPS, Gonçalves FMA, Acosta JJ, Kirst M (2018) Genomic selection prediction models comparing sequence capture and SNP array genotyping methods. *Mol Breeding* 38:1–14
- de Peixoto LA, Laviola BG, Alves AA, Rosado TB, Bhering LL (2017) Breeding *Jatropha curcas* by genomic selection: a pilot assessment of the accuracy of predictive models. *PLOS One* 12:e0173368. <https://doi.org/10.1371/journal.pone.0173368>
- de Souza LM, dos Santos LHB, Rosa JRBF, da Silva CC, Mantello CC, Conson ARO, Scaloppi EJJ, Fialho J de F, de Moraes MLT, Gonçalves P de S, Margarido GRA, Garcia AAF, Le Guen V, de Souza AP (2018) Linkage disequilibrium and population structure in wild and cultivated populations of rubber tree (*Hevea brasiliensis*). *Front Plant Sci* 9:815. <https://doi.org/10.3389/fpls.2018.00815>
- Denis M, Bouvet J-M (2013) Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genet Genomes* 9(1):37–51. <https://doi.org/10.1007/s11295-012-0528-1>
- Denis M, Cros D, Cochard B, Camus-Kulandaivelu L, Durand-Gasselín T, Bouvet JM (2012) Potential of genomic selection in perennial crops: preliminary results in the context of Eucalyptus and oil palm breeding : P-180 [WWW Document]. Programme and book of abstracts of the 4th International Conference of Quantitative Genetics: Understanding Variation in Complex Traits, Edinburgh, UK. 17–22. <http://agritrop.cirad.fr/568293/> (Accessed 6 Jun 2019)
- dos Santos Scholz MB, Kitzberger CSG, Pereira LFP, Davrieux F, Pot D, Charmetant P, Leroy T (2014) Application of near infrared spectroscopy for green coffee biochemical phenotyping. *J Near Infrared Spectrosc* 22(6):411–421. <https://opg.optica.org/jnirs/abstract.cfm?URI=jnirs-22-6-411>
- Durán R, Isik F, Zapata-Valenzuela J, Balocchi C, Valenzuela S (2017) Genomic predictions of breeding values in a cloned Eucalyptus globulus population in Chile. *Tree Genet Genomes* 13:74. <https://doi.org/10.1007/s11295-017-1158-4>
- Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. *Theor Appl Genet* 126:1–11
- El Hamidi AH, Utsunomiya YT, Xu L, Zhou Y, Neves HH, Carneiro R, Bickhart DM, Ma L, Garcia JF, Liu GE (2018) Genomic predictions combining SNP markers and copy number variations in Nellore cattle. *BMC Genomics* 19:1–8
- Elli E, Sentelhas P, Freitas C, Carneiro R, Alcarde Alvares C (2019) Assessing the growth gaps of Eucalyptus plantations in Brazil – magnitudes, causes and possible mitigation strategies. *For Ecol Manag* 451:117464. <https://doi.org/10.1016/j.foreco.2019.117464>
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Falconer D, Mackay T (1996) Introduction to quantitative genetics. Essex, UK: Longman Group
- Fanelli Carvalho H, Galli G, Ventrone Ferrão LF, Vieira Almeida Nonato J, Padilha L, Perez Maluf M, Ribeiro Resende de Jr MF, Guerreiro Filho O, Fritsche-Neto R (2020) The effect of bienniality on genomic prediction of yield in arabica coffee. *Euphytica* 216:1–16
- FAO, 2015. World programme for the census of agriculture 2020: volume 1-Programme, concepts and definitions.
- Ferrão LFV, Ferrão RG, Ferrão MAG, Fonseca A, Carbonetto P, Stephens M, Garcia AAF (2019) Accurate genomic prediction of Coffea canephora in multiple environments using whole-genome statistical models. *Heredity (edinb)* 122:261–275. <https://doi.org/10.1038/s41437-018-0105-y>
- Fister AS, Landherr L, Maximova SN, Guiltinan MJ (2018) Transient expression of CRISPR/Cas9 machinery targeting TcNPR3 enhances defense response in Theobroma cacao. *Frontiers Plant Sci* 9:268. <https://doi.org/10.3389/fpls.2018.0026>
- Flint-Garcia SA, Thornsberry JM, Buckler ES IV (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374
- Fugerey-Scarbel A, Bastien C, Dupont-Nivet M, Lemarié S (2021) R2D2 Consortium. Why and how to switch to genomic selection: lessons from plant and animal breeding experience. *Front Genet* 12:1185
- Gianola D, Van Kaam JB (2008) Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178:2289–2303
- Goddard ME, Hayes BJ (2007) Genomic selection. *J Anim Breed Genet* 124:323–330. <https://doi.org/10.1111/j.1439-0388.2007.00702.x>
- Gois IB, Borém A, Cristofani-Yaly M, Resende MDV, Azevedo C, Bastianel M, Novelli V, Machado M (2016)

- Genome wide selection in citrus breeding. *Genet Mol Res* 15(4). <https://doi.org/10.4238/gmr15048863>
- Gramazio P, Prohens J, Toppino L, Plazas M (2021) Introgression breeding in cultivated plants. *Frontiers Plant Sci* 12:764533. <https://doi.org/10.3389/fpls.2021.764533>
- Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, Isik F, Ratcliffe B, El-Kassaby YA (2018) Quantitative genetics and genomics converge to accelerate forest tree breeding. *Front Plant Sci* 9:1693. <https://doi.org/10.3389/fpls.2018.01693>
- Grattapaglia D, Resende MD (2011) Genomic selection in forest tree breeding. *Tree Genet Genomes* 7:241–255
- Grattapaglia D (2014) Breeding forest trees by genomic selection: current progress and the way forward. In: Tuberosa R, Graner A, Frison E (eds) *Genomics of plant genetic resources*, vol 1. Managing, Sequencing and mining genetic resources. Springer Netherlands, Dordrecht, pp 651–682. https://doi.org/10.1007/978-94-007-7572-5_26
- Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol Biol* 57:461–485. <https://doi.org/10.1007/s11103-005-0257-z>
- Gupta PK, Kumar J, Mir RR, Kumar A (2010) Marker-assisted selection as a component of conventional plant breeding. *Plant Breed Rev* 33:145–217. <https://doi.org/10.1002/9780470535486.ch4>
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92:433–443. <https://doi.org/10.3168/jds.2008-1646>
- Hazel LN, Lush JL (1942) The efficiency of three methods of selection. *J Hered* 33:393–399. <https://doi.org/10.1093/oxfordjournals.jhered.a105102>
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. *Crop Sci* 49(1):1–12. <https://doi.org/10.2135/cropsci2008.08.0512>
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447. <https://doi.org/10.2307/2529430>
- Heslot N, Jannink J-L, Sorrells M (2015) Perspectives for genomic selection applications and research in plants. *Crop Sci* 55:1–12. <https://doi.org/10.2135/cropsci2014.03.0249>
- Hess M, Druet T, Hess A, Garrick D (2017) Fixed-length haplotypes can improve genomic prediction accuracy in an admixed dairy cattle population. *Genet Sel Evol* 49:54. <https://doi.org/10.1186/s12711-017-0329-y>
- Hickey LT, Hafeez AN, Robinson H, Jackson SA, Leal-Bertioli SCM, Tester M, Gao C, Godwin ID, Hayes BJ, Wulff BBH (2019) Breeding crops to feed 10 billion. *Nat Biotechnol* 37:744–754. <https://doi.org/10.1038/s41587-019-0152-9>
- Hu T, Chitnis N, Monos D, Dinh A (2021) Next-generation sequencing technologies: an overview. *Hum Immunol* 82:801–811
- Huth NI, Banabas M, Nelson PN, Webb M (2014) Development of an oil palm cropping systems model: lessons learned and future directions. *Environ Model Softw* 62:411–419
- Imai A, Kuniga T, Yoshioka T, Nonaka K, Mitani N, Fukamachi H, Hiehata N, Yamamoto M, Hayashi T (2019) Single-step genomic prediction of fruit-quality traits using phenotypic records of non-genotyped relatives in citrus. *PLoS One* 14:e0221880. <https://doi.org/10.1371/journal.pone.0221880>
- Isidro J, Jannink J-L, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. *Theor Appl Genet* 128:145–158. <https://doi.org/10.1007/s00122-014-2418-4>
- Isidro y Sánchez J, Akdemir D, (2021) Training set optimization for sparse phenotyping in genomic selection: a conceptual overview. *Front Plant Sci* 12:715910. <https://doi.org/10.3389/fpls.2021.715910>
- Isik F (2014) Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forest* 45:379–401. <https://doi.org/10.1007/s11056-014-9422-z>
- Ithnin M, Xu Y, Marjuni M, Serdari NM, Amiruddin MD, Low E-TL, Tan Y-C, Yap S-J, Ooi LCL, Nookiah R (2017) Multiple locus genome-wide association studies for important economic traits of oil palm. *Tree Genet Genomes* 13:1–14
- Jamnadas R, McMullin S, Iiyama M, Dawson IK, Powell B, Termote C, Ickowitz A, Kehlenbeck K, Vinceti B, van Vliet N, Keding G, Stadlmayr B, Van Damme P, Carsan S, Sunderland T, Njenga M, Gyau A, Cerutti P, Schure J, Kouame C, Obiri BD, Ofori D, Agarwal B, Neufeldt H, Degrande A, Serban A (2016) 2. Understanding the Roles of Forests and Tree-based Systems in Food Provision. In: Mansourian S, Vira B, Wildburger C (eds) *Forests and food: addressing hunger and nutrition across sustainable landscapes*, OBP Collection. Open Book Publishers, Cambridge, pp 29–72
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177. <https://doi.org/10.1093/bfpg/elq001>
- Juliana P, Montesinos-López OA, Crossa J, Mondal S, González Pérez L, Poland J, Huerta-Espino J, Crespo-Herrera L, Govindan V, Dreisigacker S (2019) Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theor Appl Genet* 132:177–194
- Kitony JK, Sunohara H, Tasaki M, Mori J-I, Shimazu A, Reyes VP, Yasui H, Yamagata Y, Yoshimura A, Yamasaki M (2021) Development of an Aus-derived nested association mapping (Aus-NAM) population in rice. *Plants* 10:1255
- Kopitke PM, Menzies NW, Wang P, McKenna BA, Lombi E (2019) Soil and the intensification of agriculture for global food security. *Environ Int* 132:105078. <https://doi.org/10.1016/j.envint.2019.105078>
- Kumar S, Hilario E, Deng CH, Molloy C (2020) Turbocharging introgression breeding of perennial fruit crops: a case study on apple. *Hortic Res* 7:1–7
- Kwong QB, Teh CK, Ong AL, Heng HY, Lee HL, Mohamed M, Low JZ-B, Apparow S, Chew FT, Mayes S, Kulaveerasingham H, Tammi M, Appleton DR (2016) Development and validation of a high-density SNP genotyping array for African oil palm. *Mol Plant* 9:1132–1141. <https://doi.org/10.1016/j.molp.2016.04.010>
- Kwong QB, Ong AL, Teh CK, Chew FT, Tammi M, Mayes S, Kulaveerasingham H, Yeoh SH, Harikrishna JA, Appleton DR (2017) Genomic selection in commercial perennial crops: applicability and improvement in oil palm (*Elaeis*

- guineensis Jacq). *Scientific Reports* 7:2872. <https://doi.org/10.1038/s41598-017-02602-6>
- Kwong QB, Teh CK, Ong AL, Chew FT, Mayes S, Kulaveerasingam H, Tammi M, Yeoh SH, Appleton DR, Harikrishna JA (2017) Evaluation of methods and marker systems in genomic selection of oil palm (*Elaeis guineensis* Jacq). *BMC Genetics* 18:107. <https://doi.org/10.1186/s12863-017-0576-5>
- LaFramboise T (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* 37:4181–4193. <https://doi.org/10.1093/nar/gkp552>
- Lanaud C, Motamayor JC, Risterucci A-M (2001) Implications of new insight into the genetic structure of *Theobroma cacao* L. for breeding strategies. In: Bekele F, End M, Eskes AB (eds) *Proceeding of the international workshop on new technologies and cacao breeding*, pp 89–107. Kota Kinabalu, Sabah 2001. <https://agritrop.cirad.fr/476853>
- Lebedev VG, Lebedeva TN, Chernodubov AI, Shestibratov KA (2020) Genomic selection for forest tree improvement: methods, achievements and perspectives. *Forests* 11:1190. <https://doi.org/10.3390/f11111190>
- Li Y, Suontama M, Burdon RD, Dungey HS (2017) Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application. *Tree Genet Genomes* 13:1–18
- Lin Z, Hayes B, Daetwyler H (2014) Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci* 65:1177–1191
- Liu X, Wang H, Wang H, Guo Z, Xu X, Liu J, Wang S, Li W-X, Zou C, Prasanna BM, Olsen MS, Huang C, Xu Y (2018) Factors affecting genomic selection revealed by empirical evidence in maize. *Crop J* 6:341–352. <https://doi.org/10.1016/j.cj.2018.03.005>
- Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME, Jannink JL (2011) Genomic selection in plant breeding. *Knowl Prospects ADVANCES IN AGRONOMY* 110:77–123. <https://doi.org/10.1016/B978-0-12-385531-2.00002-5>
- Lourenco D, Legarra A, Tsuruta S, Masuda Y, Aguilar I, Misztal I (2020) Single-step genomic evaluations from theory to practice: using SNP chips and sequence data in BLUPF90. *Genes* 11:790
- Lyra DH, Galli G, Alves FC, Granato ÍSC, Vidotti MS, e Sousa MB, Morosini JS, Crossa J, Fritsche-Neto R (2019) Modeling copy number variation in the genomic prediction of maize hybrids. *Theor Appl Genet* 132:273–288
- Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci* 12:57–63
- Maldonado C, Mora F, Contreras-Soto R, Ahmar S, Chen J-T, do Amaral Júnior AT, Scapim CA (2020) Genome-wide prediction of complex traits in two outcrossing plant species through deep learning and Bayesian regularized neural network. *Frontiers Plant Sci* 11:1734
- Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA (2016) Predicting responses in multiple environments: issues in relation to genotype × environment interactions. *Crop Sci* 56:2210–2222
- Marchal A, Legarra A, Tisne S, Carasco-Lacombe C, Manez A, Suryana E, Omoré A, Nouy B, Durand-Gasselin T, Sánchez L (2016) Multivariate genomic model improves analysis of oil palm (*Elaeis guineensis* Jacq) progeny tests. *Molecular Breeding* 36:2
- Marchal A, Schlichting CD, Gobin R, Balandier P, Millier F, Muñoz F, Pâques LE, Sánchez L (2019) Deciphering hybrid larch reaction norms using random regression. *G3: Genes, Genomes, Genetics* 9:21–32
- McElroy MS, Navarro AJR, Mustiga G, Stack C, Gezan S, Peña G, Sarabia W, Saquicela D, Sotomayor I, Douglas GM, Migicovsky Z, Amores F, Tarqui O, Myles S, Motamayor JC (2018) Prediction of cacao (*Theobroma cacao*) resistance to *Moniliophthora* spp. diseases via genome-wide association analysis and genomic selection. *Front Plant Sci* 9:343. <https://doi.org/10.3389/fpls.2018.00343>
- Merrick LF, Herr AW, Sandhu KS, Lozada DN, Carter AH (2022) Optimizing plant breeding programs for genomic selection. *Agronomy* 12:714
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genet* 157:1819–1829
- Minamikawa MF, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, Yoshioka T, Imai A, Hamada H, Hayashi T (2017) Genome-wide association study and genomic prediction in citrus: potential of genomics-assisted breeding for fruit quality traits. *Sci Rep* 7:1–13
- Momen M, Mehrgardi AA, Sheikhi A, Kranis A, Tusell L, Morota G, Rosa GJM, Gianola D (2018) Predictive ability of genome-assisted statistical models under various forms of gene action. *Sci Rep* 8:12309. <https://doi.org/10.1038/s41598-018-30089-2>
- Montesinos-López OA, Montesinos-López A, Pérez-Rodríguez P, Barrón-López JA, Martini JW, Fajardo-Flores SB, Gaytan-Lugo LS, Santana-Mancilla PC, Crossa J (2021) A review of deep learning applications for genomic selection. *BMC Genomics* 22:1–23
- Morota G, Gianola D (2014) Kernel-based whole-genome prediction of complex traits: a review. *Front Genet* 5:363
- Mphahlele MM, Isik F, Hodge GR, Myburg AA (2021) Genomic breeding for diameter growth and tolerance to leptocybe gall wasp and botryosphaeria/teratosphaeria fungal disease complex in *Eucalyptus grandis*. *Front Plant Sci* 12:228. <https://doi.org/10.3389/fpls.2021.638969>
- Mrode R, Ojango JMK, Okeyo AM, Mwacharo JM (2019) Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: current status and future prospects. *Front Genet* 9:694. <https://doi.org/10.3389/fgene.2018.00694>
- Mrode RA (2014) *Linear models for the prediction of animal breeding values*, 2nd edn. CABI International, Wallingford, Oxon, pp 235–245
- Müller BSF, Neves LG, de Almeida Filho JE, Resende MFR, Muñoz PR, dos Santos PET, Filho EP, Kirst M, Gratapaglia D (2017) Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genomics* 18:524. <https://doi.org/10.1186/s12864-017-3920-2>
- Munyengwa N, Le Guen V, Bille HN, Souza LM, Clément-Demange A, Mournet P, Masson A, Soumahoro M, Kouassi D, Cros D (2021) Optimizing imputation of

- marker data from genotyping-by-sequencing (GBS) for genomic selection in non-model species: rubber tree (*Hevea brasiliensis*) as a case study. *Genomics* 113:655–668. <https://doi.org/10.1016/j.ygeno.2021.01.012>
- Ni G, Cavero D, Fangmann A, Erbe M, Simianer H (2017) Whole-genome sequence-based genomic prediction in laying chickens with different genomic relationship matrices to account for genetic architecture. *Genet Sel Evol* 49(1):1–14. <https://doi.org/10.1186/s12711-016-0277-y>
- Nielsen NH, Jahoor A, Jensen JD, Orabi J, Cericola F, Edriss V, Jensen J (2016) Genomic prediction of seed quality traits using advanced barley breeding lines. *PLoS One* 11(10). <https://doi.org/10.1371/journal.pone.0164494>
- Nsibi M, Gouble B, Bureau S, Flutre T, Sauvage C, Audergon J-M, Regnard J-L (2020) Adoption and optimization of genomic selection to sustain breeding for apricot fruit quality. *G3: Genes, Genomes Genet* 10:4513–4529
- Nyine M, Uwimana B, Blavet N, Hřibová E, Vanrespaille H, Batte M, Akech V, Brown A, Lorenzen J, Swennen R, Doležel J (2018) Genomic prediction in a multiploid crop: genotype by environment interaction and allele dosage effects on predictive ability in banana. *Plant Genome* 11:170090. <https://doi.org/10.3835/plantgenome2017.10.0090>
- Oliveira H, Brito L, Lourenco D, Silva F, Jamrozik J, Schaeffer L, Schenkel F (2019) Invited review: advances and applications of random regression models: from quantitative genetics to genomics. *J Dairy Sci* 102:7664–7683
- Paludeto JGZ, Grattapaglia D, Estopa RA, Tambarussi EV (2021) Genomic relationship-based genetic parameters and prospects of genomic selection for growth and wood quality traits in *Eucalyptus benthamii*. *Tree Genet Genomes* 17:1–20
- Peixoto MA, Alves RS, Coelho IF, Evangelista JSPC, de Resende MDV, de Rocha JRDOASC, e Silva FF, Laviola BG, Bhering LL (2020) Random regression for modeling yield genetic trajectories in *Jatropha curcas* breeding. *Plos one* 15:e0244021
- Persa R, de Oliveira Ribeiro PC, Jarquin D (2021) The use of high-throughput phenotyping in genomic selection context. *Crop Breed App Biotechnol* 21:1–11. <https://doi.org/10.1590/1984-70332021v21Sa19>
- Pirker J, Mosnier A, Kraxner F, Havlík P, Obersteiner M (2016) What are the limits to oil palm expansion? *Glob Environ Chang* 40:73–81. <https://doi.org/10.1016/j.gloenvcha.2016.06.007>
- Pootakham W, Jomchai N, Ruang-areerate P, Shearman JR, Sonthirod C, Sangsrakru D, Tragoonrung S, Tangphat-sornruang S (2015) Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics* 105:288–295. <https://doi.org/10.1016/j.ygeno.2015.02.002>
- Priyadarshan P (2011) *Biology of Hevea rubber*. Springer
- Pszczola M, Strabel T, Mulder H, Calus M (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci* 95:389–400
- Rambolarimanana T, Ramamonjisoa L, Verhaegen D, Tsy J-MLP, Jacquin L, Cao-Hamadou T-V, Makouanzi G, Bouvet J-M (2018) Performance of multi-trait genomic selection for *Eucalyptus robusta* breeding program. *Tree Genet Genomes* 14:1–13
- Resende MDV, Resende MFR, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ, Kilian A, Grattapaglia D (2012) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol* 194:116–128. <https://doi.org/10.1111/j.1469-8137.2011.04038.x>
- Resende RT, Resende MDV, Silva FF, Azevedo CF, Takahashi EK, Silva-Junior OB, Grattapaglia D (2017) Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity (edinb)* 119:245–255. <https://doi.org/10.1038/hdy.2017.37>
- Resende RT, Piepho H-P, Rosa GJ, Silva-Junior OB, de Resende MDV, Grattapaglia D (2021) Environments in breeding: applications and perspectives on envirotypic-assisted selection. *Theor Appl Genet* 134:95–112
- Reyes VP, Angeles-Shim RB, Mendiore MS, Manuel M, Carmina C, Lapis RS, Shim J, Sunohara H, Nishiuchi S, Kikuta M (2021) Marker-assisted introgression and stacking of major QTLs controlling grain number (Gn1a) and number of primary branching (WFP) to NERICA cultivars. *Plants* 10:844
- Rincent R, Kuhn E, Monod H, Oury F-X, Rousset M, Allard V, Le Gouis J (2017) Optimization of multi-environment trials for genomic selection based on crop models. *Theor Appl Genet* 130:1735–1752
- Rincent R, Charpentier J-P, Faivre-Rampant P, Paux E, Le Gouis J, Bastien C, Segura V (2018) Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar. *G3: Genes, Genomes Genet* 8:3961–3972
- Rival A, Levang P (2014) Palms of controversies: oil palm and development challenges. Bogor: Center for International Forestry Research. <https://doi.org/10.17528/cifor/004860>
- Robertsen CD, Hjortshøj RL, Janss LL (2019) Genomic Selection in Cereal Breeding. *Agronomy* 9:95. <https://doi.org/10.3390/agronomy9020095>
- Romero Navarro JA, Phillips-Mora W, Arciniegas-Leal A, Mata-Quirós A, Haiminen N, Mustiga G, Livingstone Iii D, van Bakel H, Kuhn DN, Parida L, Kasarskis A, Motamayor JC (2017) Application of genome wide association and genomic prediction for improvement of cacao productivity and resistance to black and frosty pod diseases. *Front Plant Sci* 8:1905. <https://doi.org/10.3389/fpls.2017.01905>
- Röös E, Bajželj B, Smith P, Patel M, Little D, Garnett T (2017) Greedy or needy? Land use and climate impacts of food in 2050 under different livestock futures. *Glob Environ Chang* 47:1–12. <https://doi.org/10.1016/j.gloenvcha.2017.09.001>
- Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94(3):441–448
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* 74:5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>

- Schiessl S-V, Katche E, Ihien E, Chawla HS, Mason AS (2019) The role of genomic structural variation in the genetic improvement of polyploid crops. *The Crop Journal* 7:127–140
- Scossa F, Brotman Y, de e Lima FA, Willmitzer L, Nikoloski Z, Tohge T, Fernie AR (2016) Genomics-based strategies for the use of natural variation in the improvement of crop metabolism. *Plant Sci* 242:47–64
- Scossa F, Alseekh S, Fernie AR (2021) Integrating multi-omics data for crop improvement. *J Plant Physiol* 257:153352
- Silva FA, Viana AP, Corrêa CCG, Santos EA, Oliveira JAVS, Andrade JDG, Ribeiro RM, Glória LS (2021) Bayesian ridge regression shows the best fit for Ssr markers in *Psidium guajava* among Bayesian models. *Sci Rep* 11(1): 1–11. <https://doi.org/10.1038/s41598-021-93120-z>
- Silva-Junior OB, Faria DA, Grattapaglia D (2015) A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol* 206:1527–1540. <https://doi.org/10.1111/nph.13322>
- Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9:477–485
- Sneller C, Ignacio C, Ward B, Rutkoski J, Mohammadi M (2021) Using Genomic selection to leverage resources among breeding programs: consortium-based breeding. *Agronomy* 11:1555
- Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008) Genomic selection using different marker types and densities. *J Anim Sci* 86:2447–2454. <https://doi.org/10.2527/jas.2007-0010>
- Sørensen P, Edwards SM, Madsen P, Jensen P, Sørensen IF, de los Campos G, Sorensen D (2013) Genomic feature models: conference on genomics of common diseases. *Book of Abstracts. Conference on Genomics of Common Diseases*, Oxford, United Kingdom, 07/09/2013, p 68
- Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, Pereira AA, Sakiyama NS, Zambolim L, Resende MDV (2019) Early selection enabled by the implementation of genomic selection in *Coffea arabica* breeding. *Front Plant Sci* 9:1934. <https://doi.org/10.3389/fpls.2018.01934>
- de Sousa IC, Nascimento M, Silva GN, Nascimento ACC, Cruz CD, de Almeida DP, Pestana KN, Azevedo CF, Zambolim L, Caixeta ET (2020) Genomic prediction of leaf rust resistance to *Arabica* coffee using machine learning algorithms. *Sci Agrár* 78(4):1–8. <https://doi.org/10.1590/1678-992X-2020-0021>
- Souza LM, Francisco FR, Gonçalves PS, Scaloppi Junior EJ, Le Guen V, Fritsche-Neto R, Souza AP (2019) Genomic selection in rubber tree breeding: a comparison of models and methods for managing G×E interactions. *Front Plant Sci* 10:1353. <https://doi.org/10.3389/fpls.2019.01353>
- Steiger JH (1980) Tests for comparing elements of a correlation matrix. *Psychol Bull* 87:245
- Sun D, Cen H, Weng H, Wan L, Abdalla A, El-Manawy AI, Zhu Y, Zhao N, Fu H, Tang J (2019) Using hyperspectral analysis as a potential high throughput phenotyping tool in GWAS for protein content of rice quality. *Plant Methods* 15:1–16
- Tan B, Grattapaglia D, Martins GS, Ferreira KZ, Sundberg B, Ingvarsson PK (2017) Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Plant Biol* 17:110. <https://doi.org/10.1186/s12870-017-1059-6>
- Tan B, Grattapaglia D, Wu HX, Ingvarsson PK (2018) Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Sci* 267:84–93. <https://doi.org/10.1016/j.plantsci.2017.11.011>
- Tan B, Ingvarsson PK (2019) Integrating genome-wide association mapping of additive and dominance genetic effects to improve genomic prediction accuracy in *Eucalyptus*. *bioRxiv* 15(2). <https://doi.org/10.1002/tpg2.20208>
- Thistlethwaite FR, El-Dien OG, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU, Ingvarsson PK, El-Kassaby YA (2020) Linkage disequilibrium vs pedigree: genomic selection prediction accuracy in conifer species. *PLOS ONE* 15:e0232201. <https://doi.org/10.1371/journal.pone.0232201>
- Tian Z, Wang J, Li J, Han B (2021) Designing future crops: challenges and strategies for sustainable agriculture. *Plant J* 105:1165–1178
- Tong H, Nikoloski Z (2021) Machine learning approaches for crop improvement: leveraging phenotypic and genotypic big data. *J Plant Physiol* 257:153354
- Tran HT, Lee LS, Furtado A, Smyth H, Henry RJ (2016) Advances in genomics for the improvement of quality in coffee. *J Sci Food Agric* 96:3300–3312
- Tyczewska A, Woźniak E, Gracz J, Kuczyński J, Twardowski T (2018) Towards food security: current state and future prospects of agrobiotechnology. *Trends Biotechnol* 36:1219–1229. <https://doi.org/10.1016/j.tibtech.2018.07.008>
- Uitdewilligen JG, Wolters A-MA, D’hoop BB, Borm TJ, Visser RG, Van Eck HJ (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS one* 8:e62355
- Van Eeuwijk FA, Bustos-Korts D, Millet EJ, Boer MP, Kruijer W, Thompson A, Malosetti M, Iwata H, Quiroz R, Kuppe C (2019) Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci* 282:23–39
- VanRaden PM (2007) Genomic measures of relationship and inbreeding. *Interbull Bulletin* 37:33–33
- Varshney RK, Roorkiwal M, Sorrells ME (2017) Genomic selection for crop improvement: an introduction. In: Varshney RK, Roorkiwal M, Sorrells ME (eds) *Genomic selection for crop improvement: new molecular breeding strategies for crop improvement*. Springer International Publishing, pp 1–6. https://doi.org/10.1007/978-3-319-63170-7_1
- Voss-Fels KP, Cooper M, Hayes BJ (2019) Accelerating crop genetic gains with genomic selection. *Theor Appl Genet* 132:669–686. <https://doi.org/10.1007/s00122-018-3270-8>
- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping, and genotyping of

- single-nucleotide polymorphisms in the human genome. *Science* 280:1077–1082. <https://doi.org/10.1126/science.280.5366.1077>
- Wang N, Jassogne L, van Asten PJA, Mukasa D, Wanyama I, Kagezi G, Giller KE (2015) Evaluating coffee yield gaps and important biotic, abiotic, and management factors limiting coffee production in Uganda. *Eur J Agron* 63:1–11. <https://doi.org/10.1016/j.eja.2014.11.003>
- Wang X, Xu Y, Hu Z, Xu C (2018) Genomic selection methods for crop improvement: current status and prospects. *Crop J* 6(4):330–340. <https://doi.org/10.1016/j.cj.2018.03.001>
- Wartha CA, Lorenz AJ (2021) Implementation of genomic selection in public-sector plant breeding programs: current status and opportunities. *Crop Breeding Appl Biotechnol* 21:1–19. <https://doi.org/10.1590/1984-70332021v21Sa28>
- Weir BS (1979) Inferences about linkage disequilibrium. *Biometrics* 35:235–254. <https://doi.org/10.2307/2529947>
- Wientjes YCJ, Veerkamp RF, Calus MPL (2013) The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* 193:621–631. <https://doi.org/10.1534/genetics.112.146290>
- Wiggans GR, Cole JB, Hubbard SM, Sonstegard TS (2017) Genomic selection in dairy cattle: the USDA experience. *Annu Rev Anim Biosci* 5:309–327. <https://doi.org/10.1146/annurev-animal-021815-111422>
- Woittiez LS, van Wijk MT, Slingerland M, van Noordwijk M, Giller KE (2017) Yield gaps in oil palm: a quantitative review of contributing factors. *Eur J Agron* 83:57–77. <https://doi.org/10.1016/j.eja.2016.11.002>
- Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor Appl Genet* 116:815–824. <https://doi.org/10.1007/s00122-008-0715-5>
- World Coffee Research (2022) Innovea Global Coffee Breeding Network. <https://worldcoffeeresearch.org/programs/global-breeding-network>
- Wu D, Guo Z, Ye J, Feng H, Liu J, Chen G, Zheng J, Yan D, Yang X, Xiong X (2019) Combining high-throughput micro-CT-RGB phenotyping and genome-wide association study to dissect the genetic architecture of tiller growth in rice. *J Exp Bot* 70:545–561
- Würschum T, Longin CFH, Hahn V, Tucker MR, Leiser WL (2017) Copy number variations of CBF genes at the Fr-A2 locus are essential components of winter hardiness in wheat. *Plant J* 89:764–773
- Xu Y (2016) Envirotyping for deciphering environmental impacts on crop plants. *Theor Appl Genet* 129:653–673. <https://doi.org/10.1007/s00122-016-2691-5>
- Xu Y, Liu X, Fu J, Wang H, Wang J, Huang C, Prasanna BM, Olsen MS, Wang G, Zhang A (2020) Enhancing genetic gain through genomic selection: from livestock to plants. *Plant Communications* 1:100005. <https://doi.org/10.1016/j.xplc.2019.100005>
- Yeap W-C, Norkhairunnisa Che Mohd Khan, Norfadzilah Jamalludin, Muad MR, Appleton DR, Harikrishna Kulaveerasingam (2021) An efficient clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 mutagenesis system for oil palm (*Elaeis guineensis*). *Frontiers in Plant Science* 12
- Yuan Y, Bayer PE, Batley J, Edwards D (2021) Current status of structural variation studies in plants. *Plant Biotechnol J* 19:2153–2163
- Zhang D, Motilal L (2016) Origin, dispersal, and current global distribution of cacao genetic diversity. In: *Cacao Diseases*. Springer, pp 3–31. https://doi.org/10.1007/978-3-319-24789-2_1
- Zhou L, Holliday JA (2012) Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC Genomics* 13:1–12
- Zuidema PA, Leffelaar PA, Gerritsma W, Mommer L, Anten NP (2005) A physiological production model for cocoa (*Theobroma cacao*): model presentation, validation and application. *Agric Syst* 84:195–225

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.