



OPEN ACCESS

**Edited by:**

Fernando H. Toledo,  
International Maize and Wheat  
Improvement Center, Mexico

**Reviewed by:**

Russell Jessup,  
Texas A & M University, United States  
Filipe Matias,  
University of Wisconsin-Madison,  
United States  
Jose Crossa,  
International Maize and Wheat  
Improvement Center, Mexico

**\*Correspondence:**

Juarez C. Machado  
juarez.machado@embrapa.br

**†ORCID:**

Rosângela M. Simeão  
orcid.org/0000-0001-8351-846X  
Marcos D. V. Resende  
orcid.org/0000-0002-3087-3588  
Rodrigo S. Alves  
orcid.org/0000-0002-3038-6210  
Marco Pessoa-Filho  
orcid.org/0000-0002-7333-4429  
Ana Luisa S. Azevedo  
orcid.org/0000-0003-1939-0339  
Chris S. Jones  
orcid.org/0000-0001-9096-9728  
Jorge F. Pereira  
orcid.org/0000-0001-9340-065X  
Juarez C. Machado  
orcid.org/0000-0003-2551-2080

**Specialty section:**

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 07 February 2021

**Accepted:** 06 April 2021

**Published:** 30 April 2021

**Citation:**

Simeão RM, Resende MDV,  
Alves RS, Pessoa-Filho M,  
Azevedo ALS, Jones CS, Pereira JF  
and Machado JC (2021) Genomic  
Selection in Tropical Forage Grasses:  
Current Status and Future  
Applications.  
Front. Plant Sci. 12:665195.  
doi: 10.3389/fpls.2021.665195

# Genomic Selection in Tropical Forage Grasses: Current Status and Future Applications

Rosângela M. Simeão<sup>1†</sup>, Marcos D. V. Resende<sup>2†</sup>, Rodrigo S. Alves<sup>3†</sup>,  
Marco Pessoa-Filho<sup>4†</sup>, Ana Luisa S. Azevedo<sup>5†</sup>, Chris S. Jones<sup>6†</sup>, Jorge F. Pereira<sup>5†</sup> and  
Juarez C. Machado<sup>5\*†</sup>

<sup>1</sup> Embrapa Gado de Corte, Campo Grande, Brazil, <sup>2</sup> Embrapa Café, Universidade Federal de Viçosa, Viçosa, Brazil, <sup>3</sup> Instituto Nacional de Ciência e Tecnologia do Café, Universidade Federal de Viçosa, Viçosa, Brazil, <sup>4</sup> Embrapa Cerrados, Brasília, Brazil, <sup>5</sup> Embrapa Gado de Leite, Juiz de Fora, Brazil, <sup>6</sup> International Livestock Research Institute, Nairobi, Kenya

The world population is expected to be larger and wealthier over the next few decades and will require more animal products, such as milk and beef. Tropical regions have great potential to meet this growing global demand, where pasturelands play a major role in supporting increased animal production. Better forage is required in consonance with improved sustainability as the planted area should not increase and larger areas cultivated with one or a few forage species should be avoided. Although, conventional tropical forage breeding has successfully released well-adapted and high-yielding cultivars over the last few decades, genetic gains from these programs have been low in view of the growing food demand worldwide. To guarantee their future impact on livestock production, breeding programs should leverage genotyping, phenotyping, and envirotyping strategies to increase genetic gains. Genomic selection (GS) and genome-wide association studies play a primary role in this process, with the advantage of increasing genetic gain due to greater selection accuracy, reduced cycle time, and increased number of individuals that can be evaluated. This strategy provides solutions to bottlenecks faced by conventional breeding methods, including long breeding cycles and difficulties to evaluate complex traits. Initial results from implementing GS in tropical forage grasses (TFGs) are promising with notable improvements over phenotypic selection alone. However, the practical impact of GS in TFG breeding programs remains unclear. The development of appropriately sized training populations is essential for the evaluation and validation of selection markers based on estimated breeding values. Large panels of single-nucleotide polymorphism markers in different tropical forage species are required for multiple application targets at a reduced cost. In this context, this review highlights the current challenges, achievements, availability, and development of genomic resources and statistical methods for the implementation of GS in TFGs. Additionally, the prediction accuracies from recent experiments and the potential to harness diversity from genebanks are discussed. Although, GS in TFGs is still incipient, the advances in genomic tools and statistical models will speed up its implementation in the foreseeable future. All TFG breeding programs should be prepared for these changes.

**Keywords:** apomixis, brachiaria, elephant grass, forage breeding, Guinea grass, marker-assisted selection, polyploidy

## INTRODUCTION

The tropics are home to about a third of the world's population, accounting for 36% of the Earth's landmass, and where most of the global demographic increase takes place (Morales, 2009). The tropical region is the center of origin and domestication of many of the world's most important food crops and is responsible for 50% beef production and 40% milk production worldwide (Morales, 2009; Alexandratos and Bruinsma, 2012). Despite the huge importance of the tropical region, there is an evident gap in technological development between the tropical nations and industrialized temperate countries. The tropical region has great potential to meet the growing global demand for food requirements with intensification through improved management and technologies. One important step toward intensification is the acceleration of breeding programs.

Plant breeding has evolved from a rudimentary process in its early stages to a modern and sophisticated system in the past few decades. The rediscovery of Mendel's laws of genetics in 1900 is one of the pillars of modern plant breeding (Hallauer, 2011), which can also count on modern techniques such as high-throughput sequencing, bioinformatics, and automated phenotyping (Barabaschi et al., 2016). Breeding pipelines can apply these techniques to increase the rates of genetic gain taking into account the parameters found in the "breeder's equation" (Lush, 1937; Eberhart, 1970), which states that the genetic gain is directly proportional to the accuracy of the observed phenotype in relation to the true phenotype and genotype, selection intensity, and genetic variation, but inversely proportional to the time of the breeding cycle. Manipulating variables in the "breeder's equation" can increase genetic gain as well as reduce the timeframe to develop new cultivars (Pereira et al., 2018).

Pastures are the main food source for animal feeding in the tropics, as in Brazil, where approximately 90% of the livestock are solely grass-fed (Silva et al., 2016). An increase in productivity and quality of tropical forage grasses (TFGs) will have a significant impact on livestock production. Although, cattle and buffaloes already contribute to the largest proportion of global animal protein supply, increased quantities of milk and beef are necessary due to growing demands. Production will have to increase by 57% for beef and 48% for milk by 2050 compared to that in 2005, as projected by the FAO (Alexandratos and Bruinsma, 2012), while other estimates indicate that the global demand for livestock products will double by 2050 (Bajželj et al., 2014; Rao et al., 2015). This higher production needs

to take into account scenarios where the land destined for pastures may have to be reduced, as has been happening in Brazil (IBGE, 2016). Efforts to breed TFGs focus on increasing productivity and quality while also reducing losses due to biotic and abiotic stresses. However, breeding efforts have been hindered by many features of tropical forages that render the implementation of more dynamic breeding programs difficult. TFGs encompass perennial monocotyledonous plants from the family Poaceae, mostly polyploid, with a C4 photosynthetic pathway and showing both sexual and apomictic reproductive systems. Breeding programs of TFGs face challenges such as different ploidy levels and reproductive modes, evaluation of perennial plants over different cuts, distribution of efforts among different species, the evaluation of traits being laborious and expensive, and most breeding programs being held by public institutions. The development and release of a new cultivar can take up to 10 years (Jank et al., 2014).

Genomic selection (GS) offers the opportunity to increase agricultural production and reduce the breeding interval cycle to at least half of the conventional time (Crossa et al., 2017). Reduction of the breeding cycle is the main advantage of GS in forage breeding (Simeão-Resende et al., 2014). GS and genome-wide association studies (GWAS) have enormous potential for use in the selection of complex traits such as yield, disease, and insect resistance, facilitating the rapid selection of new cultivars to meet the future demand for food and fodder (Heffner et al., 2010; Talukder and Saha, 2017). However, breeding programs of TFGs are still behind those of grain and fiber crops, and even those of temperate/sub-tropical forages, regarding the application of genomic tools as a strategy to accelerate cultivar development. Challenges in applying GS in tropical forages include designing and obtaining adequately sized training populations; developing high-quality, low-cost, and reproducible marker panels; dealing with polyploidy; and gaining knowledge of the genetic architecture of target traits.

This article provides an overview of GS in TFGs focusing on the current scenario, recent advances, and prospects for the effective application of tools and strategies to accelerate TFG breeding. We have focused on elephant grass (*Cenchrus purpureus* syn. *Pennisetum purpureum*), Guinea grass (*Megathyrus maximus* syn. *Panicum maximum*), and brachiaria (*Urochloa brizantha* syn. *Brachiaria brizantha*, *U. decumbens* syn. *B. decumbens*, and *U. ruziziensis* syn. *B. ruziziensis*), which account for most of the pastures in many parts of the world, including Africa, Asia, Australia, and Latin America. Specific features of breeding programs, availability of genomic resources, statistical methods for GS, and gaps in the application of GS in TFG breeding are discussed here. This discussion is essential for the initiation and practical implementation of GS in TFG breeding programs.

## TFG BREEDING

TFG breeding began relatively recently (Valle et al., 2009). For example, EMBRAPA, the Brazilian Agricultural Research Corporation, started its breeding programs for *Urochloa* and

**Abbreviations:** BLUP, best linear unbiased prediction; CIAT, International Center for Tropical Agriculture; EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária; GBS, genotyping-by-sequencing; GEBV, genomic breeding values; GETV, total (genotypic) genomic values; GLS, generalized least squares; GS, genomic selection; GV, genomic value; GWAS, genome-wide association studies; IBERS, Institute of Biology, Environmental and Rural Sciences at Aberystwyth University; ICARDA, International Center for Agricultural Research in the Dry Areas; ILRI, International Livestock Research Institute; LD, linkage disequilibrium; NIRS, near-infrared spectroscopy; MAE, minor allele frequency; MAS, marker-assisted selection; QTL, quantitative trait loci; QTN, quantitative trait nucleotide; REML, restricted maximum likelihood; SARDI, South Australian Research and Development Institute; SNP, single-nucleotide polymorphism; TFG, tropical forage grass; USDA, United States Department of Agriculture.

*Megathyrus* in the 1980s (Jank et al., 2014), while the elephant grass breeding program only began in 1991 (Pereira et al., 2017). One of the first steps toward an effective breeding program is the compilation of a germplasm collection. Approximately 17,000 accessions of TFGs have been preserved in the primary global germplasm banks, such as CIAT, EMBRAPA, IBERS, ICARDA, ILRI, SARDI, and USDA, where *Urochloa* spp., *Cenchrus* spp., and *Megathyrus* spp. correspond to most of the accessions alongside *Digitaria* spp. and *Paspalum* spp. The genetic variability maintained in these germplasm banks is invaluable, and these banks offer genetic resources adapted to varied edaphoclimatic conditions and diverse purposes. However, there is no corresponding use of this variability during crossings in practical breeding programs. This indicates that these collections are not being used to their full potential, although, initiatives of germplasm exchange between institutions have been taken to increase genetic variability that can be used in breeding programs (Negawo et al., 2018; Habte et al., 2020). Better characterization of germplasm collections will enable quicker utilization to meet the dynamic demands of the production sector with the emergence of diverse limitations due to climate, pests, or changes in production systems.

The limited use of accessions preserved in germplasm banks for crossing is one of the limitations of TFG breeding (Valle et al., 2009; Pereira et al., 2018). Other limitations include the use of a large number of candidate genera and species, insufficient information on the biology of the species, low genetic variability for important traits, polyploidy, a complex mode of reproduction (apomixis), wild characteristics of important species (dehiscence and malformation of seeds, anti-quality factors, and sensitivity to photoperiod), lack of information on the genetic control and heritability of agronomic traits, and little participation of the private sector in the development of cultivars (Valle et al., 2009; Sandhu et al., 2015; Pereira et al., 2018). It is worth noting that brachiaria, Guinea grass, and elephant grass are perennial species, which implies that most traits are evaluated in the field over long periods, including several cuts. It is common for a specific trait to show variation among different cuts (Rocha et al., 2019), which is very different compared to that of annual species. The impact of these limitations, along with the differences in market demands and the ability of producers to absorb new releases, can be seen in the low number of cultivars released through the years when compared to the release of grain and fiber crops (Figure 1).

Despite these limitations, the improvement of forage grasses has revolutionized the pastoral systems. In a study conducted in sub-Saharan Africa, legume, and grass cultivars released for different animal production systems increased forage production by 2.65 times when compared to that of traditional cultivars. Production was even higher when only forage grass was used (Paul et al., 2020). Until recently, germplasm introduction was the key method used for forage grass breeding, which involved the evaluation and selection of germplasm accessions as a strategy to obtain cultivars. This method was used to release *U. brizantha* cv. Marandu in 1984 by EMBRAPA (Nunes et al., 1984), which is currently cultivated on fifty-mega hectares of land in Brazil alone (Jank et al., 2014). The germplasm introduction method, albeit simple, rapid, and cost-effective, tends to be prone to exhaustion

as it requires the use of accessions collected from nature or accessions obtained from a germplasm bank in other breeding programs (Jank et al., 2011). In addition, natural habitats of species are being increasingly degraded, with loss of variability as well as restrictions in free access to germplasm across different countries and breeding programs, notably due to recent laws for access to genetic diversity and protection of cultivars (Pengelly and Maass, 2019). Since 2000, the use of recombination as a key strategy for cultivar development has intensified. For example, among the new cultivars released by various breeding programs, intra- or interspecific hybrids, especially of *Urochloa*, *Megathyrus*, and *Cenchrus*, have been highlighted, in which the favorable traits of their progenitors are gathered. Of note, in the last decade, long-term recurrent selection programs have been established for major species, and promising results have been achieved (Miles et al., 2006; Reis et al., 2008; Barrios et al., 2013). The main objectives of TFG breeding are to identify and develop improved genotypes that contribute to increased animal productivity and reduced environmental impact (Figure 2). Thus, not only a better agronomic behavior of the plant, but also a more productive performance of the animal is sought, while ensuring minimal environmental impact (Valle, 2001).

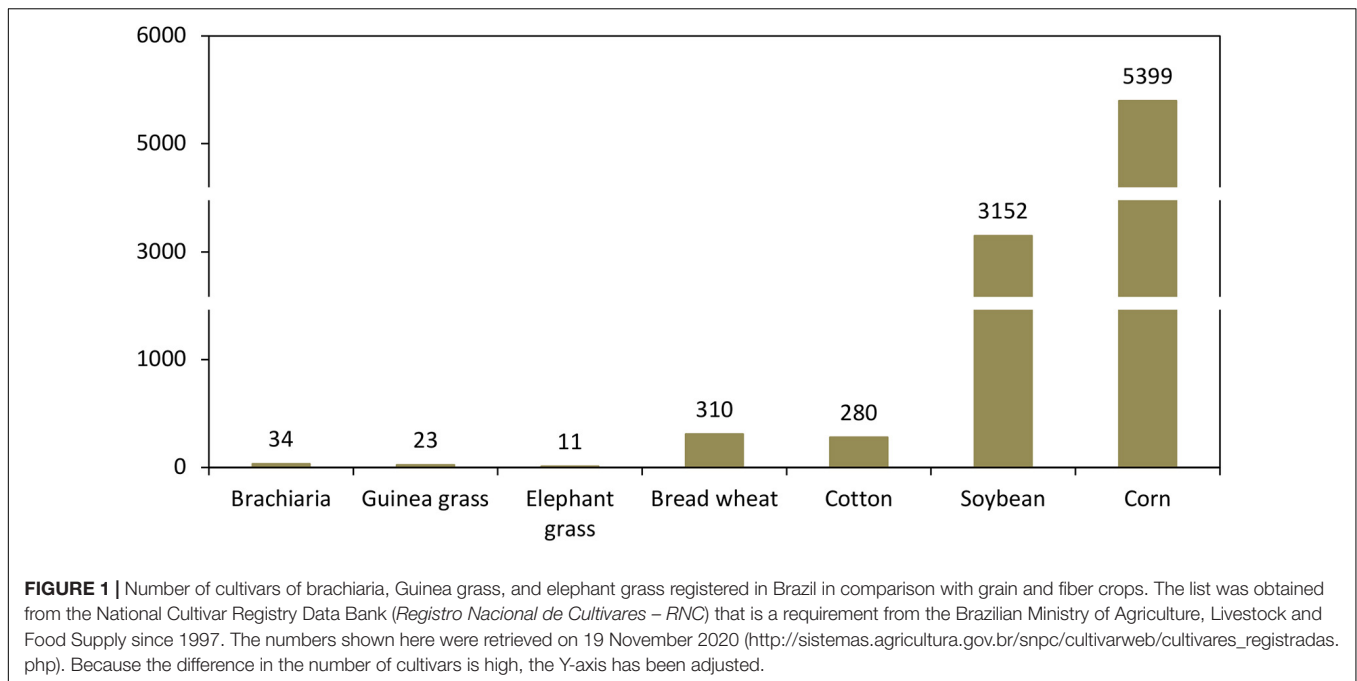
Although, TFG breeding programs have been successful in releasing new and important cultivars over the years, there are certain challenges to overcome. In Brazil, these challenges include reducing losses due to biotic stresses (especially spittlebug attacks), increasing adaptation based on expected climate changes, and improving nutritional value to enhance animal performance, resulting in more beef and milk per kilogram of pasture. To address these challenges, research priorities have focused on the development of new capabilities such as the availability of genome sequences, high-throughput genotyping, and germplasm characterization of tropical forage grasses; identification of genes associated with important traits; development and use of large-scale phenotyping tools; and implementation of GS (Pereira et al., 2018).

Therefore, the prospects of applying genomic tools in TFG breeding programs are promising, and these tools, coupled with adequate pasture management, can continue to promote substantial advances in livestock productivity. For each forage species, the objectives of the breeding programs should be well-defined, as highlighted in Figure 2. In addition to clear objectives, it is important to use the latest technologies available to accelerate the development of cultivars. In this regard, the use of genomic tools for TFG breeding is fundamental.

## GS: AN APPRAISAL IN TFG BREEDING

### Peculiarities in Breeding Perennial TFGs: Polyploidy and Apomixis

Perennial forages require selection methods that consider the effects of both between families and within family individuals for higher selection gains, mostly in lower magnitudes of narrow-sense heritability (NSH) (Simeão-Resende et al., 2013, 2014). In forage breeding, a combination of GS methods is expected to be useful for predicting genomic breeding values (GEBVs)



and total (genotypic) genomic values (GETVs), for clonally propagated cultivars. The estimation of marker effects and genomic values should enable an increase in selection accuracy and may reduce the time required for completing a breeding cycle and the evaluation cost per genotype. Forage breeding methods associated with GS show differential accuracy and gains, as demonstrated by Simeão-Resende et al. (2014).




Considering all these premises, and the methods of the current breeding programs, the uses, and advances of GS in tropical forages will be presented. Firstly, facts about perennial TFGs must be pointed out, as most are polyploids and reproduce apomictically. Apomixis is asexual reproduction by seed (Barcaccia et al., 2020) producing genetically identical progeny (Hand and Koltunow, 2014). In both important genera of TFG, *Megathyrus* and *Urochloa*, gametophytic apomixis subtype apospory occurs (Ozias-Akins and van Dijk, 2007) which is a mode of reproduction in which the embryo originates from a polyploid nucellar cell as a maternal clone by the seed (Valle and Savidan, 1996). Therefore, the commercial cultivars of these species are generally both polyploid and apomictic.

Autotetraploid individuals have been developed in *U. ruziziensis* and *M. maximus* by artificially duplicated chromosomes from diploid sexual individuals (Jank et al., 2014). These sexual individuals are essential for hybridization with apomictic ones in breeding programs of both genera to increase genetic variability and enhance selection. Therefore, cytogenetic analysis is constantly performed in parents and hybrids and should be evaluated by considering the importance of the species targeted by GS.

According to Bourke et al. (2018), the knowledge of the meiotic behavior of a species is sometimes required to analyze polyploid data using dosage calling software that uses the expected segregation ratios in the  $F_1$  autotetraploid population.

Unlike allotetraploids, autotetraploids do not behave like diploids during meiosis and require specialized methods and tools for genetic studies and mapping (Gallais, 2003). Autotetraploid plants exhibit polysomic inheritance, which can be detected during a cytogenetic analysis by visualization of tetravalent formation and segmental pairing among “partially homologous” chromosomes (Stebbins, 1947) as well as by molecular inference (Worthington et al., 2016). The consequence of chromosome pairing in a tetravalent is the generation of unbalanced gametes and individuals with non-Mendelian inheritance. Even in recent autotetraploids induced by colchicine, chromosome pairing may not show tetravalent formation or other meiotic abnormalities (Pagliarini et al., 2008); however, the four alleles per locus are always present. This may generate errors in genetic mapping, haplotype designation, and the estimation of marker effects, which are important factors in genomic prediction.

Diploid sexual individuals of *M. maximus* were collected in Korogwe, Tanzania, and artificially duplicated (Jank et al., 2014). The cytogenetic evaluation of autotetraploid ( $2n = 4x = 32$ ) sexual and supposedly segmental allopolyploid apomictic plants revealed a low-to-moderate rate of meiotic abnormalities among sexual (5%–31%) and apomictic (7%–11%) parents (Pessim et al., 2010, 2015). Hybrids originating from a single cross showed abnormal cells at a rate ranging from 16% to 52% (Pessim et al., 2010, 2015). The frequency of meiotic abnormalities found in *M. maximus* is lower than that reported in the tetraploid *Urochloa* ( $2n = 4x = 36$ ) interspecific hybrids, which ranged from 18% to 82% (Risso-Pascotto et al., 2005; Mendes-Bonato et al., 2006, 2007; Fuzinato et al., 2007). Pagliarini et al. (2008) found that the mean occurrence of meiotic abnormalities in five induced autotetraploid *U. ruziziensis* accessions ranged from 5% to 10% and only one accession reached 55% abnormalities. However, contrary to expectation, in all the autotetraploidized

	Advantages	Specific breeding goals	General breeding goals
 Brachiaria	Good adaptability to acid and low fertility soils, easy multiplication by seeds, high capacity to compete with invasive plants and good animal performance when compared to native pastures	Resistance to spittlebugs, nutritional value, adaptation to acidic soils, seed production, seasonality, productivity, seed production, quality, vigor, persistence, and tolerance of shade	Total dry matter productivity, leaf dry matter productivity, good distribution of production throughout the year, nutritional value and forage quality (consumption, palatability, chemical composition, digestibility, and absence of anti-nutritional factors), adaptation to the environment of interest (soil, climate, and management), persistence and vigor (survival and income over time), resistance to pests and diseases, resistance/tolerance to abiotic factors (drought, flooding, and temperature), highly efficient and/or sustainable use of nutrients (e.g., biological nitrogen fixation), ease of propagation, speed of establishment, cycle, seed production, compatibility of association with other species
 Elephant grass	High potential of forage production, adaptation to different ecosystems, good acceptance by animals, high leaf/stem ratio, high nutritional value of pasture, high animal performance, resistance to water deficit	Productivity, nutritional value, seed propagation and quality, adaptation to mechanical harvesting, adaptation to grazing, resistance to spittlebugs, tolerance of low temperature, silage, and energy	
 Guinea grass	High potential of forage production, high nutritional value and acceptance by animals of different categories, easily propagated by seeds	Leaf productivity, resistance to <i>Bipolaris</i> , and seed production	

**FIGURE 2 |** Characteristics of brachiaria, Guinea grass, and elephant grass and breeding goals to improve their use as tropical forage grasses. The advantages and breeding goals are based on Machado et al. (2019). Source of the pictures: Embrapa.

accessions, chromosome pairing was preferentially bivalent (Pagliarini et al., 2008).

Regardless of the low rate of tetraivalent formation in tetraploid and sexual *M. maximus* and *U. ruziziensis*, the issues of allele dosage and compatibility between apomictic and sexual genomes still remains unresolved, as explained by Gallais (2003) for recently doubled genotypes. Therefore, efficient cytogenetic identification of the best crosses at early stages would allow for the identification of the best and most cytogenetically stable parents and progenies. Consequently, all subsequent stages of breeding programs will certainly benefit from genomic prediction and the unbiased estimation of marker effects.

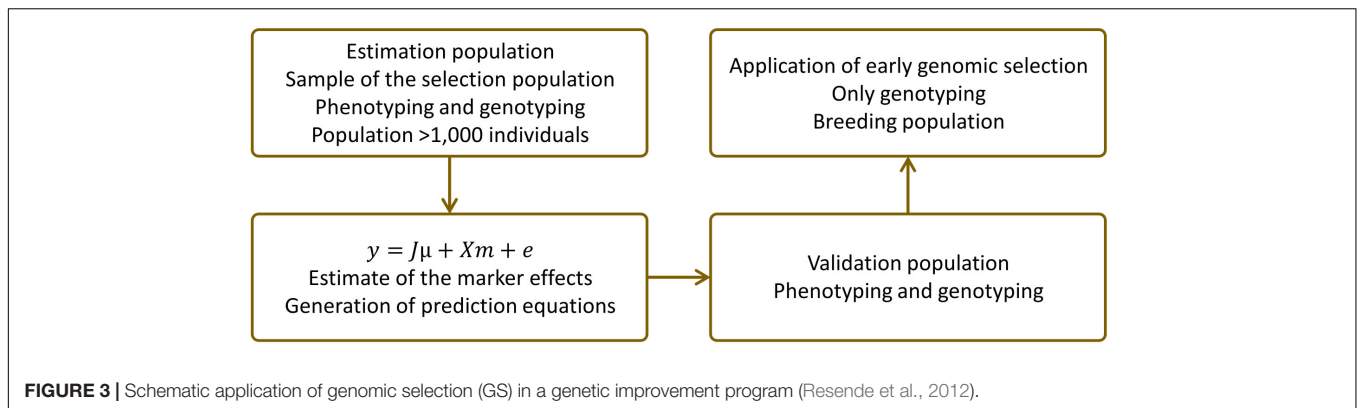
## Availability of TFG Breeding Populations for GS

In practice, three populations must be defined for GS: estimation, validation, and breeding populations (Goddard and Hayes, 2007; Meuwissen, 2007). These populations may be as follows: i) physically distinct (three different populations), ii) with two simultaneous functions (only one population used for estimation

and validation), or iii) with three simultaneous functions (only one population used for estimation, validation, and selection). **Figure 3** illustrates strategy ii.

## Estimation Population

The estimation population is also called the discovery, training, or reference population. This dataset includes a large number of markers assessed in a moderate number of individuals (1,000 to 2,000 depending on the desired accuracy), which should have their phenotypes assessed for various traits of interest. Equations for predicting genomic values (random multiple regression) are obtained for each trait. These equations associate each marker or interval with its effect (predicted by RR-BLUP) on the trait of interest. The markers that explain the loci regulating the traits are identified in this population, and their effects are estimated. Recently, Lara et al. (2019) and Matias et al. (2019a) used estimation populations with 530 individuals of *M. maximus* and 272 individuals of *Urochloa* hybrids, respectively. Predictive abilities were lower than 0.4 for the evaluated traits. This indicates that factors affecting GS efficiency in TFG breeding, such as adequately sized training populations, still need to be improved.



### Validation Population

When physically separated from the estimation population, this dataset is smaller than the discovery population and includes individuals that are assessed for SNP markers and various traits of interest. The equations for predicting genomic values are tested to verify their accuracy for this independent sample. To calculate the accuracy, genomic values are predicted, using the estimated effects from the estimation population and subjected to correlation analysis with the observed phenotypic values. As the validation sample is not involved in predicting the marker effects, errors from genomic values and phenotypic values are independent. Correlations between these values are predominantly genetic in nature and equivalent to the predictive ability ( $r_{\hat{y}y}$ ) of GS in estimating phenotypes, which is given by the accuracy of the selection itself ( $r_{\hat{q}q}$ ) multiplied by the square root of the heritability ( $h$ ), or  $r_{\hat{y}y} = r_{\hat{q}q}h$ . Thus, to estimate the accuracy, one should obtain  $r_{\hat{q}q} = r_{\hat{y}y}/h$ . This method is valid when raw phenotypic values are used to calculate the correlations. When using genotypic values predicted based on phenotypes instead of raw phenotypic values, heritability should be replaced by the reliability of the prediction. In general, strategy ii is adopted according to a k-fold scheme for cross-validation. According to Meuwissen (2007), when dozens to hundreds of thousands of haplotypes are estimated, there is a risk of over-parameterization; in other words, errors in the data are explained by the marker effects. Cross-validation is therefore extremely important to address this problem.

### Breeding Population

This dataset only contains the markers assessed in the candidates for selection, and the phenotypes do not need to be assessed in this population. Therefore, the prediction equations derived from the estimation population are used to predict the GVs or future phenotypes of the candidates for selection. The associated selection accuracy is calculated for the validation population.

In most TFG breeding programs, the training population is the same as, or part of, the breeding population, and this population may have experienced directional selection for many generations (Simeão-Resende et al., 2014). It is likely that validation may never exist for most breeding programs, firstly because cross-validation (Kohavi, 1995) has been the commonly used method

and secondly because it is difficult to find out more than one ongoing breeding program per species and country.

Elephant grass (*Cenchrus purpureus*) is a tetraploid and allogamous species in which open pollinated divergent populations are easy to establish. The two main breeding strategies are: (i) recurrent selection (Reis et al., 2008); and (ii) clonal selection (Pereira et al., 2017; Machado et al., 2019). The recommended number of individuals in estimation populations for GS can be easily reached, in which the effective population size can be previously assumed and effectively estimated after genotyping. Validation can be performed by cross-validation and this may be extended to different environments or even to related populations of breeding programs in other countries.

Guinea grass (*M. maximus*) and brachiaria (*Urochloa*) breeding programs have recently adopted the full-sib reciprocal recurrent selection as a method, in which thousands of hybrids are generated annually (Barrios et al., 2013; Worthington and Miles, 2015). As the intrapopulation recurrent selection is feasible only on sexual populations, because there is no possible crossing between apomictic accessions, selection is performed on sexual individuals as a function of heterosis expressed in crossings with apomictic accessions. The schematic drawing of this procedure was presented for *M. maximus* (Simeão-Resende et al., 2004) and *Urochloa* (Jank et al., 2014). Based on this information, and the fact that these methods have been recently implemented, some limitations for the establishment of estimation populations in these genera must be discussed. Firstly, the  $N_e$  of the sexual population is extremely low ( $N_e < 7$ ) in *M. maximus* (Simeão-Resende et al., 2004; Lara et al., 2019), in tetraploid *U. ruziziensis* used as female parents in crossings with apomictic *U. brizantha* (Simeão et al., 2015), and in tetraploid sexual *U. decumbens* (Barrios et al., 2013). Secondly, albeit the genetic diversity within populations of apomictic accessions of *Urochloa* species is high (Vigna et al., 2011), the number of accessions used in crosses is low, because the crosses are performed based only on the adapted and agronomically selected apomictic individuals (Jones et al., 2021). Therefore, the  $N_e$  is likewise low ( $< 20$ ). Thirdly, as a result of the intra or interspecific crosses,  $F_1$  progeny segregates for mode of reproduction in which individuals in the progeny may vary from 99% apomictic to 99% sexual. While this procedure is efficient to explore the panmictic heterosis (Lamkey and Edwards, 1999) and can readily generate apomictic

individuals that are potential cultivar candidates, these hybrid swarms do not constitute a breeding population *per se*. Therefore, the  $N_e$  among and within hybrid families must be adjusted and considered for efficient GS use in both genera. Using the equation  $N_e = \frac{4N_f n}{n+1}$ , in which  $N_f$  is the number of full-sib families and  $n$  the number of individuals per family, we may suggest the evaluation of 42 families and 10 individuals per family, a total of 420 genotyped individuals, to achieve  $N_e$  of approximately 152. For phenotyping, the total number may be approximately 1000 individuals (Resende, 2015). Finally, in practice, the TFG estimation population for GS in the genera *Megathyrus* and *Urochloa* could be composed of sexual hybrids in an open pollinated population, followed by validation in apomictic hybrids.

## Genetic Markers in Selection and Gene Discovery

The use of molecular genetic markers for selection and genetic improvement is based on the genetic linkage between these markers and a quantitative trait locus (QTL) of interest (Resende et al., 2013). Thus, linkage disequilibrium (LD) between markers and QTLs is essential for genomic selection from genomic information (Bourke et al., 2018). It must be made clear that a QTL refers only to the statistical association between a genomic region and a trait.

Recently, molecular genetic markers that consist of SNPs (based on the detection of polymorphisms that arise from a single nucleotide change in the genome) have been widely used in many species (Elshire et al., 2011). Generally, for a SNP to be considered genetically derived, the polymorphism must occur in at least 1% of the population (Resende, 2015). SNPs are the most common type of genomic variation and preferred over other genetic markers because of their abundance, ease of obtainment, and low genotyping cost. Thousands of SNPs can be used to cover the entire genome of an organism with markers not more than 1 cM apart from each other.

LD analysis is based on LD between a marker and a QTL in the whole population and not only within a family, as performed in linkage analysis (Würschum, 2012). For this to occur, the marker and QTL must be closely linked. When this occurs, the association between them is a property of the entire population and persists for many generations.

Association analysis is used for fine mapping and is based on population-level LD (Resende et al., 2013). Linkage can occur when the gene directly affects a trait, and when there is an LD between the marker and the gene controlling the trait. In the first case, the effect of the gene is directly evaluated, and the marker is classified as functional. The functional mutations are known as quantitative trait nucleotides (QTNs). In the second case, the linkage test requires LD between the marker and QTL. When a mutation occurs on a given chromosome, it creates a haplotype with adjacent loci on the chromosome. In the subsequent generations, this mutation tends to occur within the same haplotype unless there is recombination, which creates the LD used for association mapping (Resende et al., 2013).

In recent times, efforts to use molecular markers in genetic improvement research have evolved into two approaches: (1) GWAS for QTL identification and mapping; and (2) genome-wide selection (GWS) or GS (Resende, 2015). GS was proposed by Meuwissen et al. (2001) to increase breeding efficiency and accelerate genetic improvement. GS emphasizes the simultaneous prediction (without the use of significance tests for individual markers) of the genetic effects of thousands of SNP markers dispersed throughout the genome of an organism to capture the effects of all loci (both small and large effects) and identify the overall genetic variation of a quantitative trait (Resende and Alves, 2020). In this case, the sum of the estimated genetic effects of the markers present in an individual provides the genetic value of the individual for selection purposes. Meuwissen et al. (2001) obtained a complete array of estimates of haplotype effects using the ridge regression best linear unbiased prediction (RR-BLUP), BayesA, and BayesB methods. RR had already been used by Whittaker et al. (2000) for marker selection. Haley and Visscher (1998) suggested the name GS for selection on a whole genome scale (Resende and Alves, 2020).

The conceptual development of GS coincides with the technology associated with SNPs, which is accurate and relatively affordable. GS uses the associations between many SNP markers throughout the genome along with phenotypes and takes advantage of LD between markers and QTLs in close linkage (Resende et al., 2013). The predictions derived from phenotypes and SNP genotypes with high density in a generation are thus used to obtain genomic values (GVs) for individuals in any subsequent generations based on their genotypic markers, in which the genetic effects have been estimated.

When LD between markers is incomplete, the joint allele frequencies for the two loci can change markedly across generations, thereby leading to changes in haplotypes. In this case, the marker effects would need to be estimated again to maintain the accuracy of GS for various generations (Dekkers, 2007). In the case of a complete or close LD, the estimated effects remain constant across different families and generations within the same environment.

The number of markers used are directly associated to the genome size, extent of LD and population structure (Ballesta et al., 2020). Larger genomes, rapid breakdown of LD and greater effective population size imply that a higher density of SNP loci would be needed.

## Factors Affecting the Efficiency of GS

The accuracy of GS depends on five factors (Resende et al., 2014): i) heritability of the trait; ii) number of loci regulating the trait (also given by  $2N_e L$ ) and the distribution of their effects; iii) number of individuals in the discovery population; iv) effective population size ( $N_e$ ); and v) marker density, which depends on the number and genome size ( $L$ , in Morgans). The first two factors are beyond the breeder's control, and the latter three factors can be modified by the breeder to increase the accuracy of GS.

An increase in the selection efficiency using GS can be achieved by changing the four components of the expression for genetic progress, given by  $SG = (kr_{g\hat{g}}\sigma_g)/T$ , where  $k$

is the standardized selection differential (dependent on the selection intensity),  $r_{\hat{g}g}$  is the accuracy of selection,  $\sigma_g$  is the genetic standard deviation (genetic variability) of the trait in the population, and  $T$  is the time required to complete a selection cycle.

In perennial and vegetatively propagated plant species, the benefit of GS is from an increase in  $r_{\hat{g}g}$  and a reduction in  $T$ . The increase in  $r_{\hat{g}g}$  is due to the use of an actual kinship matrix (Resende, 2007). This increase depends on the size of the estimation population and marker density. Factor  $T$  is greatly reduced by GS because genomic prediction and selection can be performed at the seedling stage. Thus, even if  $r_{\hat{g}g}$  shows the same magnitude as obtained by phenotypic selection, GS is still better than selection based on phenotypes due to the reduction of  $T$ .

### Inferring the Quality and Efficiency of GS

The quality of GS is inferred by correlation and regression among the predicted genetic values and phenotypes in the validation population, as well as by the accuracy of the prediction. The correlation and regression coefficients involving observed and predicted values are practical measures of the ability of the methods to make predictions that are accurate and unbiased, respectively. Correlation provides predictive ability, which is equivalent to the product of accuracy and the square root of heritability (Resende et al., 2014). The regression coefficient is algebraically equal to 1. Regression coefficients of less than 1 indicate that the genetic values are overestimated and exhibit greater variability than expected; coefficients greater than 1 indicate that the estimated genetic values exhibit variability lower than expected. A lack of bias is important when selection involves individuals from many generations using the estimated marker effects from a single generation. Regression coefficients near 1 indicate that the assessments are unbiased and effectively predict the actual magnitudes of differences among the individuals assessed.

The expected value of the regression coefficient is 1, which indicates an unbiased prediction. Thus, the regression coefficient can also be used to estimate the heritability of the markers (Resende et al., 2013). Various heritability values are assessed, and those that provide a regression equal to 1 should be selected as the best estimate. If the regression yields a result less than 1, the magnitude of the assessed heritability value is too high and should be reduced until the regression coefficient is converged to 1. If the regression yields a result greater than 1, the magnitude of the assessed heritability value is too low and should be increased until it converges to 1.

## CURRENT GENOMIC RESOURCES FOR TFG BREEDING

The availability of a large number of high-quality single-nucleotide polymorphisms (SNPs) that can be genotyped at a reasonable cost is a prerequisite for implementing GS in a crop of choice (Hayes et al., 2013). Gold-standard methods for SNP discovery rely on resequencing individual samples at minimum coverage and mapping reads to a reference genome

(McKenna et al., 2010); this is also true for low-coverage, genotyping-by-sequencing (GBS) methods (Elshire et al., 2011). Although, reference-free pipelines are available for GBS, such as UNEAK (Lu et al., 2013), reference-based methods allow for more informed decisions during the selection of high-quality SNPs (McCormick et al., 2015). Alternative methods have been recently tested and reported in *Urochloa*, for which the GBS-SNP-CROP pipeline (Melo et al., 2016) was used to generate a “mock” reference from GBS data (Matias et al., 2019b). This led to the discovery of a larger number of biallelic SNPs, when compared to mapping reads to the available genome of the closest related species (*Setaria viridis* and *Setaria italica*).

Assembly quality, which is measured by its accuracy, haplotype phasing, and contiguity, is an important factor influencing marker discovery, GWAS, and GS. In tetraploid blueberries, the selection of probes for targeted SNP genotyping, investigation of the genetic architecture of fruit traits, identification of candidate genes, and genomic prediction benefited from a more complete, chromosome-scale, haplotype-phased genome assembly (Benevenuto et al., 2019). A huge reduction in sequencing costs and increased throughput brought about by second-generation sequencing have allowed unprecedented access to crop genomic information (Shamshad and Sharma, 2018). However, the genomic complexity of most TFG species is still challenging and has kept them recalcitrant to sequencing efforts targeting reference-grade assemblies using mainly second-generation short reads. Forage grass species with the largest breeding programs in Latin America are mostly polyploid and highly heterozygous and have genomes with a high repeat content (Table 1). Third-generation long reads can circumvent some of these problems, and, combining with a technology such as optical mapping or chromosome conformation capture can potentially allow chromosome-scale reference-grade assemblies (Belser et al., 2018; Shi et al., 2019). The use of these tools to accelerate genomic research on TFGs is promising and can thus advance the use of GS in breeding programs. We argue that the availability of high quality genome assemblies is the starting point for the development of genotyping systems that will be useful for successfully deploying GS in TFG breeding programs.

**TABLE 1** | Reproductive system and genomic information of economically important tropical forage grasses used in livestock production.

Scientific name	Predominant reproductive system	Genome size (Gpb)	Chromosome number and ploidy level	WGS <sup>#</sup>
<i>Urochloa brizantha</i>	Apomictic	1.4	2n = 4x = 36	No
<i>Urochloa humidicola</i>	Apomictic	1.9	2n = 6x, 9x = 36 to 54	No
<i>Urochloa decumbens</i>	Apomictic	1.6	2n = 4x = 36	No
<i>Urochloa ruziziensis</i>	Sexual	0.6	2n = 2x = 18	Yes
<i>Megathyrsus maximus</i>	Apomictic	1.0	2n = 4x = 32	No
<i>Cenchrus purpureus</i>	Sexual	2.1	2n = 4x = 28	Yes

<sup>#</sup>WGS, availability of whole-genome sequence.



## Genome Assembly in *Urochloa* spp.

*Urochloa* P. Beauv. Grasses, many of which were previously included in *Brachiaria* (Trin.) Griseb., are the most widely used forage in Latin America. The main species are *U. ruziziensis*, *U. brizantha*, *U. decumbens*, and *U. humidicola*. The first genome assembly of a forage grass species in the genus *Urochloa* has recently been reported for *U. ruziziensis* (Worthington et al., 2020). The diploid genotype CIAT26162 was sequenced using short reads (approximately 100 × coverage). Assembly and scaffolding were performed and PacBio RSII reads were used for gap filling. This is an invaluable resource for an orphan species with no previously published genome information available. However, it also highlights the huge challenges in the assembly of highly heterozygous, repetitive genomes with short-read technologies. The publicly available assembly was fragmented into 102,577 scaffolds, with an N50 of 27.8 kbp. Completeness metrics based on benchmarking universal single-copy orthologs (BUSCOs) indicate that 86.7% are complete (1,248 out of 1,440 in the Embryophyta\_odb9 dataset), suggesting that there is still room for improvement. Because of the lack of linkage maps for an intraspecific diploid cross in *U. ruziziensis*, which could be used to cluster and order scaffolds based on linkage groups, the study relied on anchoring scaffolds based on synteny with the *Setaria italica* genome to obtain pseudo-molecules. This potentially affected the anchored assembly due to undetected chromosomal rearrangements that might be present between *S. italica* and *U. ruziziensis*, as phylogenetic information and evolutionary relationships were not taken into account in the anchoring process.

## Genomic Resources for Guinea Grass (*M. maximus*)

The placement of Guinea grass in the phylogeny of Paniceae has changed over the years, however, plant breeders, seed producers, and farmers in Latin America mostly refer to it as *Panicum maximum*, or “Panicum” as a common name for the species. There is no reference genome assembly available for *M. maximus*. The fact that Guinea grass was once included in *Panicum* and is still mostly regarded as a *Panicum* species in Latin America may lead to the assumption that the available genome assemblies for panic grasses such as *P. hallii* (two publicly available chromosome-scale assemblies; Lovell et al., 2018) and *P. milliaceum* (two chromosome-scale assemblies; Shi et al., 2019; Zou et al., 2019) would provide suitable shortcuts for the development of genomic resources for *M. maximus*. Lara et al. (2019) illustrated how this approach might limit the genomic distribution and number of SNPs available when genotyping *M. maximus* breeding populations using *Panicum* genome assemblies as references. In this study, a multi-parental population of *M. maximus* half-sib progenies was genotyped using GBS, and six different assemblies were tested as references for read mapping. Two of them comprised the genome assemblies for *P. hallii* and *P. virgatum*, while the remaining included genome assemblies for *Setaria* (*S. italica* and *S. viridis*) and transcriptome assemblies for *M. maximus*. The alignment rates ranged between 19.05% for *P. hallii* and 24.24% for *M. maximus*

transcriptomes, showing that a very large proportion of reads were not used for SNP discovery and genotyping. The sets of allele-dosed SNPs containing up to 5% of missing data from each of the reference assemblies ranged between 5,032 for one of the *M. maximus* transcriptomes and 8,112 for *S. viridis*. Although, this was the first report on the development and assessment of prediction models, considering the allele dosage, for GS in *M. maximus*, increased marker density and prediction accuracies may be expected when a high-quality genome assembly for the species is available for SNP discovery and genotyping.

## Genomic Resources and Genome Assemblies for Elephant Grass (*C. purpureus*)

Elephant grass (*C. purpureus* Schumach. Morrone, syn. *Pennisetum purpureum* Schumach.), also called napier grass, merker grass, or Uganda grass, is a tropical grass native to Eastern and Central Africa. It is an allotetraploid species with a chromosome constitution of  $2n = 4x = 28 A'ABB$  and an average amount of DNA per G1 nucleus of 4.58 pg (Hanna, 1981; Taylor and Vasil, 1987). Elephant grass is used as animal fodder and is a promising lignocellulosic biofuel feedstock due to its high growth rate, high biomass yield, and persistence (Morais et al., 2009; Singh et al., 2013; Daud et al., 2014; Rocha et al., 2019).

Wang et al. (2018) conducted a genome survey of elephant grass and estimated its genome size to be 2.01 Gb with 71.36% of repetitive elements and a heterozygosity of 1.02%. A total of 114.36 Gb of raw data, (approximately 57-fold coverage) was generated using the Illumina HiSeq sequencing platform for the Zise genotype (purple elephant grass). A partial draft assembly was obtained using SOAPdenovo. As expected for such a complex genome, this effort allowed a preliminary investigation of the repetitive content of elephant grass and the identification of thousands of genomic SSR markers, 30 of which were tested for genotyping a set of 28 elephant grass accessions. Another genome survey of Merkeron and UF1 cultivars was conducted by Paudel et al. (2018), and they also developed a high-density linkage map using GBS.

More recently, two chromosome-scale assemblies of elephant grass were reported (Yan et al., 2020; Zhang et al., 2020). The initial assembly of the cv. Purple (Yan et al., 2020) was obtained using Nanopore long reads and was then polished with Illumina short reads and scaffolded with Hi-C data. Approximately 2,000 contigs were grouped and oriented into 14 chromosome-scale scaffolds, with a total size of 1.9 Gbp, 66.3% of which were annotated as repetitive elements. The assembly showed high contiguity at the contig level (N50 1.8 Mbp) and 97.8% completeness using BUSCOs. The predicted protein-coding gene set also showed high BUSCO completeness (97.1%).

The second assembly (Zhang et al., 2020, available as a preprint), for the CIAT6263 accession, was obtained with Nanopore reads and ultra-long reads, which were assembled to a total size of 2.07 Gbp with a contig N50 of approximately 2.9 Mbp. The authors used a combination of BioNano optical maps and Hi-C to obtain a final chromosome-scale assembly of approximately 2 Gbp in 14 pseudomolecules. This assembly was

also 97.8% BUSCO complete, and repetitive elements accounted for 60.7% of the genome. The difference in repetitive content between the two assemblies might be explained by the different methods applied for *de novo* identification of repeats.

These two assemblies place elephant grass in a unique position among the TFGs. Current research trends indicate the benefits of having not only one but multiple genome assemblies for a species of interest (Della Coletta et al., 2021). Large sequencing projects now target pan-genomes, instead of a single reference that does not capture the full diversity of a species (Zhao et al., 2018). Future efforts of variant discovery and association of phenotypes with genomic locations will be possible using these assemblies as anchors for read mapping, opening up the possibility of revisiting previous datasets that were generated without a reference genome. However, while both efforts resulted in chromosome-scale scaffolds, the fact that the SMARTDENOV0 assembler does not generate haplotype-resolved contigs indicates that the high heterozygosity levels of elephant grass were not represented in the assemblies.

## STATISTICAL METHODS IN GS

An ideal method for estimation of SNP effects in GS should accommodate the genetic architecture of the trait in terms of genes of small and large effects and their distributions, regularize the estimation process in the presence of multicollinearity and a larger number of markers than individuals, using shrinkage estimators, and perform the selection of covariables (markers) that affect the trait under analysis. The main problem with GS is the estimation of a large number of effects from a limited number of observations and the collinearities arising from LD between the markers. Shrinkage estimators deal with this appropriately by treating the effects of markers as random variables and estimating them simultaneously (Resende, 2007; Resende et al., 2008; Azevedo et al., 2015; Resende and Alves, 2020).

If the effects of markers are taken as fixed, it is not possible to consider the covariance between the effects of the markers. With a high density of markers, more than one marker will be in LD with a segregating QTL, which will result in covariance between the marker effects. Most markers will have no effect on a trait, and the estimated effects of these empty markers will be false. This problem is greater when the markers are considered to have fixed effects, because in that case, these pseudo effects will not shrink toward zero (Resende and Alves, 2020).

In the context of marker-assisted selection (MAS) and genomic prediction, the method of least squares (LS) has serious drawbacks. According to Gianola et al. (2003), the selection index (calculated as the regression involving molecular scores) presented by Lande and Thompson (1990) for MAS fails when formulated vectorially. This failure occurs because the covariance matrix for the molecular scores is singular, as the distribution of fitted regression values is defined only in the  $p$ -dimensional space (number of covariables) and not in the  $n$ -dimensional space (number of individuals with molecular scores). Therefore, the selection index leads to an infinite number of solutions.

Another difficulty arises when the number of markers is equal to or greater than the number of genotyped individuals. In this case, the collinearity of the predictor variables causes parametric identification problems, and thus, some type of dimensional reduction, such as singular value decomposition, should be used. Another problem is the inadmissibility (unable to provide the minimum mean square error) of LS estimators, a result that collapses estimates by LS and generalized LS (GLS). Thus, the LS method is not recommended for the MAS and GS analyses. In summary, the LS method is inefficient because it is impossible to simultaneously estimate all effects when the number of effects to be estimated is greater than the number of data points; thus, estimating one effect at a time and testing its significance leads to an overestimation of significant effects, and the accuracy of the method becomes low. In addition, only QTLs with large effects will be detected and used, and consequently, not all genetic variations will be captured by the markers. The LS method assumes *a priori* QTL distribution, with an infinitely large variance that disagrees with the known total genetic variance.

Because the number of markers in GS is greater than the number of individuals, there is a lack of degrees of freedom to estimate the effects of all markers. A solution to this problem is to use the RR method (Whittaker et al., 2000) or to consider the marker effects as random instead of fixed. Fitting random effects does not expend degrees of freedom, and the effects of all markers can be estimated simultaneously. This method leads to RR-BLUP, which considers the effects of QTLs with normal distributions and equal variance through chromosomal segments.

The main problem for GS is estimating a large number of effects from a limited number of observations, in addition to collinearities resulting from LD between markers. The shrinkage estimators adequately address this issue by treating the marker effects as random variables and estimating them simultaneously (Resende et al., 2008).

The main methods for GS are based on Random Regression and can be divided into three major classes: explicit, implicit, and dimensionally reduced regression. In the first class, the RR-BLUP, Lasso, BayesA, and BayesB methods stand out among others. In the class of implicit regression, the Reproducing Kernel Hilbert Spaces (RKHS) method, which is semiparametric, is the most popular. The Independent Components, Partial Least Squares, and Principal Components stand out among the regression methods with dimensional reduction. Two new non-parametric approaches for GS proposed by Resende (2015) and Lima et al. (2019a,b) have proven to be efficient (Resende and Alves, 2020) and are called triple categorical regression (TCR) and Delta-p, respectively.

The explicit regression methods are divided into two groups: (i) penalized estimation methods (RR-BLUP, Lasso) and (ii) Bayesian estimation methods (including BayesA, BayesB, fast BayesB, BayesC $\pi$ , BayesD $\pi$ , Bayesian regression, BayesRR, BayesRS, BLasso, and IBLasso). Among these, the best and most effective in practice are RR-BLUP and BayesB (Visscher et al., 2006, 2008, 2010; Mrode et al., 2010; Mrode, 2014). Each method without covariate selection has a similar method with covariate selection. Thus, the following are the pairs without -

with covariate selection: BayesA - BayesB; BayesRR - BayesC $\pi$ ; BLasso - IBLasso (Resende and Alves, 2020).

The RR-BLUP is a model equivalent to genomic best linear unbiased prediction (G-BLUP), which is the BLUP method at an individual level with the genealogical relationship matrix  $A$  changed to a genomic relationship matrix  $G$ . The equivalence between these two methods was given by Habier et al. (2007) and Van Raden (2008). The G-BLUP and RR-BLUP are equivalent when the number of QTLs is large, and no major QTL is present. The use of matrix  $G$  based on markers had already been established by Bernardo (1994); Nejati-Javaremi et al. (1997), and Fernando (1998). A single-step BLUP simultaneously using phenotypic, genotypic, and genealogical information, called H-BLUP single-step, was proposed by Misztal et al. (2009), using an  $H$  matrix composed of the  $A$  and  $G$  matrices (Resende and Alves, 2020). The idea of H-BLUP was given by Fernando (1998).

The traditional quantitative genetics rely on random mating populations. Nowadays, with the availability of SNP markers, random mating does not need to be assumed, because breeders can track the transmission of chromosomal segments. Another assumption is linkage equilibrium in the breeding population. Once linkage among markers is accounted for in the  $G$  coefficient matrix in RR-BLUP, this circumvents the need to assume linkage equilibrium (Resende and Alves, 2020).

A refinement of GS can be achieved by using QTNs instead of SNPs. The evolution of genomic technology is predictable and the causal mutation of a genetic variation at the nucleotide level (QTN) can be accessed soon. Thus, GS can be improved by the direct use of QTNs instead of SNPs. The use of QTNs will bring the following advantages (Weller, 2016): GS will not depend on the LD as the QTN will be accessed directly and not via markers and, this will increase the robustness of the genomic prediction, which will also be useful in the long run; the genomic prediction may have transferability across different populations and species in the same genus; genomic prediction will use specific QTNs for each trait, unlike G-BLUP by means of SNPs, which uses the same  $G$  relationship matrix for all traits; the multiple-trait selection indices will directly weigh the QTNs and not the phenotypic traits; GS may use a smaller number of generations (only the last ones) for the composition of the  $G$  matrix, which will bring greater genetic gain and lesser mass of data to be processed; the allele frequencies of the QTNs will be accessed directly and not through LD with SNPs (Resende and Alves, 2020).

## Single-Environment RR-BLUP and G-BLUP Models

The parametric regression model for a single environment  $j^{\text{th}}$  ( $j = 1, \dots, m$ ) is defined as  $y_j = 1_{nj}\mu_j + X_j\beta_j + \varepsilon_j$ , where the vector  $y_j$  represents  $n_j$  independent centered observations of the response variable in the  $j^{\text{th}}$  environment;  $1_{nj}$  is a vector of ones of order  $n_j$ ;  $\mu_j$  is the overall mean of the  $j^{\text{th}}$  environment;  $X_j$  is the matrix for the  $p$  centered and standardized molecular markers in the  $j^{\text{th}}$  environment; vector  $\beta_j$  represents the effect of each of the  $p$  markers in the  $j^{\text{th}}$  environment, and  $\varepsilon_j$  is the vector of random errors in the  $j^{\text{th}}$  environment with normal distribution

and common variance  $\sigma_{\varepsilon_j}^2$ . The RR-BLUP assumes that the effects of the markers have a multivariate normal distribution  $\beta_j \sim N(0, I\sigma_{\beta_j}^2)$ .

Assuming that the effects of the markers  $\beta_j$  and  $\varepsilon_j$  are independent, and that  $u_j = X_j\beta_j$ , then the above model for the  $j^{\text{th}}$  environment can be written as  $y_j = 1_{nj}\mu_j + u_j + \varepsilon_j$ , where  $u_j$ , and  $\varepsilon_j$  are independent random variables with  $u_j \sim N(0, \sigma_{u_j}^2 K_j)$ , and  $\varepsilon_j \sim N(0, \sigma_{\varepsilon}^2 I)$ , respectively;  $\sigma_{u_j}^2$  is the variance of  $u_j$  (to be estimated), and  $K_j$  is a symmetric matrix representing the covariance of the genetic values. Thus, for a single-environment where the  $K_j$  is of the linear form  $K_j = G_j = X_j X_j' / p$  the G-BLUP is equivalent to RR-BLUP (Van Raden, 2008).

## Genetic Parameterization of Additive, Dominance, and Total Genotype Effects Additive Model

The following linear mixed model can be fitted to estimate the marker effects  $y = J\mu + Xm + e$ , where  $y$  is the vector of phenotypic observations,  $\mu$  is the vector of the fixed effect of the general mean,  $m$  is the vector of random marker effects and  $e$  is the vector of random residuals.  $J$  and  $X$  are the incidence matrices for  $\mu$  and  $m$ , respectively. The incidence matrix  $X$  contains functions of the values 0, 1, and 2 for the number of alleles for the marker (or the supposed QTL) in a diploid individual. A similar coding method uses the values of -1, 0, and 1. The genomic mixed-model equations for predicting  $m$  using the RR-BLUP method are

equivalent to 
$$\begin{bmatrix} J'J & J'X \\ X'J & X'X + I\frac{\sigma_a^2}{(\sigma_a^2/n_Q)} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} J'y \\ X'y \end{bmatrix}$$
. The

total additive genomic value for an individual  $j$  is given by  $AGV_j = \hat{y}_j = \sum_i X_i \hat{m}_i$ , where  $X_i$  is equal to 0, 1, or 2 for the genotypes mm, Mm, and MM, respectively, for biallelic and co-dominant markers, such as SNPs.

These prediction equations assume *a priori* that all loci explain equal amounts of genetic variation. Thus, the genetic variation explained by each locus is given by  $\sigma_a^2/n_Q$ , where  $\sigma_a^2$  is the total genetic variation and  $n_Q$  is the number of loci (when each locus is perfectly marked by a single marker), which can be given by  $n_Q = 2 \sum_i^n p_i (1 - p_i)$ , where  $p_i$  is the frequency of the allele of the type M in locus  $i$ . The genetic variation  $\sigma_a^2$  can be estimated by restricted maximum likelihood (REML) on the phenotypic data in a traditional manner or by the variation among markers or QTL chromosomal segments.

There is no need to use the kinship matrix with the RR-BLUP method. The pedigree-based kinship matrix used for traditional BLUP was replaced by a kinship matrix estimated by the markers. This kinship matrix is a function of  $X'X$  present in the equations of the mixed model described above. This procedure is more efficient because it effectively captures the kinship produced for each individual and not an average kinship matrix associated with the pedigree.

The parameterization of the incidence matrix  $X$  uses the values 0, 1, and 2 for the number of alleles of a marker (or supposed QTL) in a diploid individual and  $2p$  for individuals with missing

marker data. These values should be centered around 0 so that the effects of co-dominant markers are effects of allelic substitution with a mean of 0 in the population. In this case, assuming Hardy-Weinberg equilibrium, the additive genetic variation of the trait in the population is equal to  $\sigma_a^2 = 2 \sum_i^n p_i (1 - p_i) \sigma_m^2$ . Thus, the values of  $X_i$  should be replaced by  $0 - 2p$ ,  $1 - 2p$ , and  $2 - 2p$ , to obtain a variable with a mean of 0. Thus, with centralization,  $n_Q = 2 \sum_i^n p_i (1 - p_i)$ , should be used for the RR-BLUP method, and the additive genetic effects of individuals are given by  $\hat{a} = X\hat{m}$ .

Additionally, the data for markers in matrix  $X$  can be standardized as follows for each matrix element  $X_i$  corresponding to locus  $i$ :

- $X_i = (0 - 2p_i)/(\text{Var}(X_i))^{1/2}$  if the individual is homozygous for the first allele (mm).
- $X_i = (1 - 2p_i)/(\text{Var}(X_i))^{1/2}$  if the individual is heterozygous (Mm).
- $X_i = (2 - 2p_i)/(\text{Var}(X_i))^{1/2}$  if the individual is homozygous for the second allele (MM).
- $X_i = 0$  if the individual has missing marker data. The quantity  $p_i$  is the frequency of the second marker allele.

The cut-off point for including a marker in the analysis can be determined by the minor allele frequency (MAF), which is a measure related to the variation of alleles in the population, given by  $MAF = (1/2N)^{1/2}$  which comes from the standard deviation of a proportion, given by  $(pq)^{1/2}/(2N)^{1/2}$ , where  $N$  is the number of genotyped individuals, meaning that the lower the  $N$  value, the greater the MAF needs to be for accurate estimation of the marker effect (Resende, 2015; Resende and Alves, 2020).

### Coding and Additive Kinship Matrix in Polyploids

The incidence matrix  $X$  contains the values 0, 1, 2, 3, and 4 for the number of alleles for the marker (or the supposed QTL) in a tetraploid individual. Analysis by G-BLUP uses the kinship matrix given by  $G = \frac{(X^*X^{*'})}{[2 \sum_i^n p_i(1-p_i)]^{1/2}}$ , where  $X^*$  is the  $X$  matrix after centralization.

### Additive-Dominance Model

According to the marker model  $y = J\mu + W\alpha + S\delta + e$ , (where coefficients of  $\alpha$  and  $\delta$  are the additive and dominance effects, respectively), the most appropriate parameterization to estimate the effects on the additive-dominance model (Vitezica et al., 2013; Azevedo et al., 2015) is:

Additive effects (W):

$$W = \begin{cases} \text{If } MM; 2 \rightarrow 2 - 2p = 2q \\ \text{If } Mm; 1 \rightarrow 1 - 2p = q - p \\ \text{If } mm; 0 \rightarrow 0 - 2p = -2p \end{cases}$$

The values of  $W$  must be centered at zero so that the effects of the codominant markers are effects of allelic substitution ( $\alpha$ ) with a mean of 0 in the population.

Dominance effects (S):

$$S = \begin{cases} \text{If } MM; 0 \rightarrow -2q^2 \\ \text{If } Mm; 1 \rightarrow 2pq \\ \text{If } mm; 0 \rightarrow -2p^2 \end{cases}$$

### G-BLUP for the Additive-Dominance Model

The individual mixed model is given by  $y = J\mu + Za + Zd + e$ , where  $a$  is the additive genetic vector of the individuals, and  $d$  is the dominance genetic vector of the individuals;  $a \sim N(0, G_a\sigma_a^2)$ ,  $d \sim N(0, G_d\sigma_d^2)$ , and  $e \sim N(0, I\sigma_e^2)$ .

The mixed-model equations for the additive-dominance model are equivalent to

$$\begin{bmatrix} J'J & J'Z & J'Z \\ Z'J & Z'Z + G_a^{-1}\frac{\sigma_e^2}{\sigma_a^2} & Z'Z \\ Z'J & Z'Z & Z'Z + G_d^{-1}\frac{\sigma_e^2}{\sigma_d^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \\ \hat{d} \end{bmatrix} = \begin{bmatrix} J'y \\ Z'y \\ Z'y \end{bmatrix}, \quad \text{where}$$

$G_a = \frac{WW'}{\sum_{i=1}^n (2p_iq_i)}$ , and  $G_d = \frac{SS'}{\sum_{i=1}^n (2p_iq_i)^2}$ ;  $p_i$  and  $q_i$  are the allelic frequencies;  $\sigma_a^2 = \sum_{i=1}^n [2p_i(1-p_i)]\sigma_m^2$ , and  $\sigma_d^2 = \sum_{i=1}^n [2p_i(1-p_i)]^2\sigma_m^2$ ; and  $\sigma_a^2$  and  $\sigma_d^2$  are the additive and dominance genetic variances, respectively.

Adjusting an individual genomic model is equivalent to adjusting an individual traditional model but with the pedigree-based matrices  $A$  and  $D$  replaced by the genomic kinship matrices  $G_a$  and  $G_d$  for additive and dominance effects, respectively.

### H-BLUP and Single-Step BLUP

In a simultaneous analysis of genotyped and non-genotyped individuals via G-BLUP, for a global evaluation of the three classes of individuals in a single step, the same additive model  $y = J\mu + Za + e$  can be fitted with one alteration (replacing matrix  $G$  with matrix  $H$ ) to the mixed-model equations,

$$\text{according to Misztal et al. (2009)} \begin{bmatrix} J'J & J'Z \\ Z'J & Z'Z + H^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} J'y \\ Z'y \end{bmatrix}.$$

Matrix  $H$  includes both the relationships, based on pedigree ( $A$ ) and differences between those and the genomic relationships ( $A_\delta$ ), such that  $H = A + A_\delta$ . Thus,  $H$  is given by

$$H = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & G \end{bmatrix} = A + \begin{bmatrix} 0 & 0 \\ 0 & G - A_{22} \end{bmatrix}, \quad \text{where the}$$

subscripts 1 and 2 represent non-genotyped and genotyped individuals, respectively.

The inverse of  $H$ , which allows simpler calculations, is given by

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & G^{-1} + A^{22} - A_{22}^{-1} \end{bmatrix},$$

where  $A_{22}^{-1}$  is the inverse of the kinship matrix based on pedigree for only genotyped individuals.

From the estimation of genetic values ( $\hat{a}$ ) by G-BLUP, the estimated marker effects ( $\hat{m}$ ) can be obtained by:  $\hat{m} = (X'X)^{-1}X'\hat{a}$ . Models with dominance effects ( $d$ ) can also be fitted.

Another important application of this analysis is the estimation of total heritability explained by all the markers simultaneously. With the kinship matrix given by

$G = (XX')/[2 \sum_i^n p_i(1 - p_i)]$ , total heritability can be estimated by REML using the mixed-model equations to estimate the variance components  $\sigma_a^2$  and  $\sigma_e^2$ . The elements of matrix  $G$  represent the average multilocus kinship and are given by  $G_{jk} = (\frac{1}{n}) \sum_{i=1}^n \frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}$ . Another favorable feature of G-BLUP is the possibility of directly estimating (by prediction error variance (PEV)) the accuracy of GS. For individuals with known phenotypes, this accuracy is valid for the estimation population without cross-validation. In G-BLUP, the phenotypes of the validation population are replaced by missing data. Therefore, individuals from this validation population will have a validated accuracy estimate.

Models at the level of individuals, including genotype  $\times$  environment (*ae*) interactions, can also be fitted if there are related individuals within the same environment and across environments. In this case, the model is equal to  $y = Wb + Za + Zae + e$ , where *ae* is the vector of effects from the interaction between additive genetic effects and environmental effects (random), and  $Z$  is the incidence matrix for *a* and *ae*. The mixed-model equations for predicting *a* and *ae* using the BLUP method

$$\text{are } \begin{bmatrix} W'W & W'Z & W'Z \\ Z'W & Z'Z + G_a^{-1} \frac{\sigma_e^2}{\sigma_a^2} & Z'Z \\ Z'W & Z'Z & Z'Z + G_d^{-1} \frac{\sigma_e^2}{\sigma_{ae}^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{ae} \end{bmatrix} = \begin{bmatrix} W'y \\ Z'y \\ Z'y \end{bmatrix},$$

where  $G_{ae} = G$  for pairs of individuals in the same environment, and  $G_{ae} = 0$  for pairs of individuals in different environments. The variance of the interaction between the additive genetic and environmental effects is denoted by  $\sigma_{ae}^2$ .

### Additive, Dominance, and Total Genotype Effects in Polyploids

As SNP markers are biallelic, the inference of dosage allelic effect is dependent on the genetic effects of interactions (Gallais, 2003; Mackay et al., 2019). Exclusive additivity may create more classes of genotypic values than any other first-degree interaction among alleles and must be studied using allele dosage in GS. In autotetraploids or populations derived from their "allotetraploids," such as those evidenced in *M. maximus* and *U. ruziziensis*, the additive genetic variance and NSH cannot be estimated based solely on testing half-sib progenies or regression of offspring on the progenitors, because half-sib families may have fractions of the dominant genetic variance (Gallais, 1989). G-BLUP analysis of half-sib polyploid data allows the estimation of broad-sense heritability (BSH) using the information of all genetic relationships available in the kinship marker matrix since some "identity-by-state" dominance relationships allow the estimation of dominance effects, which along with estimated additive genetic effects, provide the estimation of the total genotypic value and then the BSH. This procedure was used in *M. maximus* hybrids (Lara et al., 2019) using a low number of parents and in *Urochloa* interspecific hybrids (Matias et al., 2019a).

Experimental crossing that provides simultaneous half-sib and full-sib progenies should be preferentially designed to estimate

additive and dominance genetic effects simultaneously (Simeão-Resende et al., 2004), aiming at the total genotypic-genomic value prediction. In this case, there is more information about dominance relationships, thus generating better estimates. The BSH is estimated by the additive-dominance model, in which  $g = a + d$  and  $\text{var}(g) = \text{var}(a) + \text{var}(d)$ . Making estimations of GEBV and GETV simultaneously in full-sibs and half-sibs with some progenitors in common in a training/validation population will allow the summation of the family effect in both predictions as well as the prediction of crosses that have not been performed.

In the case of similar magnitudes of NSH and BSH, there is no need for dominance adjustment, and only half-sibs can be used. In *M. maximus*, the estimated NSH and BSH for important traits showed a remarkably low and high magnitude, respectively, based on the use of phenotypic data (Simeão-Resende et al., 2004) or genomic data (Lara et al., 2019). In this species, GS based on additive and dominant effects needs to be performed to obtain the highest levels of genetic gain. In this way, it is important to work with tetrasomic inheritance more than disomic inheritance (Lara et al., 2019) and more genetically diverse synthetic populations to elevate the heritability and accuracy of GEBV prediction. The higher dominance effect evidenced in tetraploid *M. maximus* cannot be simply extended to other species unless the effect is previously known, or simply tested by different models of GS. de Bem Oliveira et al. (2019) predicted GEBV in blueberries by comparing diploid (data coded as 0, 1, and 2), tetraploid (data coded as 0, 1, 2, 3, and 4), and continuous (data coded as continuous parameterization assuming values between 0 and 1 and a cumulative additive effect) data models at the individual level. The researchers concluded that the use of continuous data generated estimated genetic gain values that were not significantly different from the best models of all traits. As diploid and tetraploid inferences of data did not affect the predictive ability, we can infer that simplified models can perform adequately.

### Ridge, Bayes, and Lasso Methods

Bayesian methods are associated with systems of nonlinear equations, and non-linear predictions can be more efficient when the QTL effects are not normally distributed owing to the presence of genes with major effects. The linear predictions associated with RR-BLUP assume that all markers with the same allele frequency contribute equally to genetic variation (lack of genes with major effects). In Bayesian estimation, the shrinkage of effect estimates for the model is controlled by the *a priori* distribution assumed for these effects. Different distributions produce different shrinkages. Methods for penalized and Bayesian estimation may include (BayesB, Fast BayesB, BayesC $\pi$ , BayesD $\pi$ , Lasso, BLasso, and IBLasso) or lack (RR-BLUP, EN, RR-BLUP-Het, and BayesA) direct covariable selection. Bayesian methods are more efficient when the distribution of QTL effects is leptokurtic (positive kurtosis) because of the presence of genes with large effects. The RR-BLUP method is equally efficient when the QTL effects are normally distributed.

Comparisons among the methods for predicting genomic breeding values have been performed. Meuwissen et al. (2001) concluded that the BayesB method is theoretically best because it

is slightly superior to RR-BLUP. However, the author simulated genotypic data with the same *a priori* distribution used for the estimation. This approach yielded greater accuracy for this method, although, such accuracy is unattainable in practice if the actual distribution associated with genetic effects differs from the *a priori* distribution assumed for analysis. In general, there is no method that is best under any circumstances because each method may yield significantly different results depending on the population structure and nature of the trait. However, the results obtained by Guo et al. (2012) indicate that the RR-BLUP method is easier to apply and equal to or better than the others for most applications in plants.

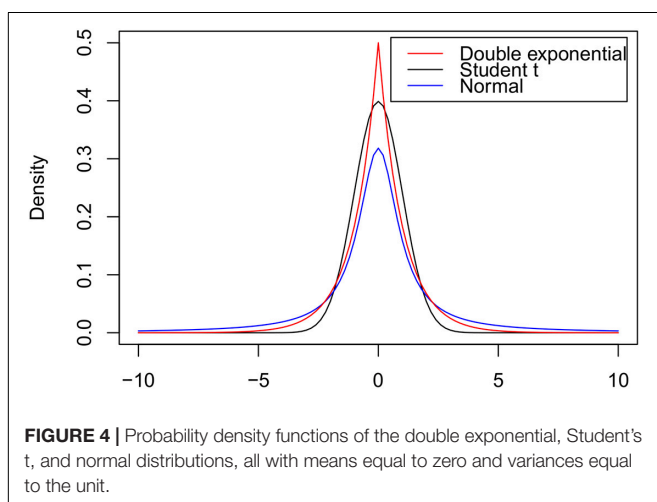
The assumed distributions for the genetic effects of markers in the different GS methods are Gaussian normal with common variance for RR-BLUP, Student's t-distribution given chi-square priori for variances for Bayesian methods, and Double Laplace exponential for Lasso. **Figure 4** illustrates the forms of the normal (RR-BLUP), t (BayesA), and double exponential (Lasso) distributions.

It is observed that, in relation to RR-BLUP, the prior density used in Bayesian Lasso shows a greater density mass at zero point and more robust tails providing greater shrinkage on regression coefficients close to zero and lower shrinkage on regression coefficients away from zero. The prior density used in BayesA also has a higher density mass at the zero point and more robust tails than the normally used RR-BLUP. Bayesian Lasso has greater shrinkage on regression coefficients close to zero than BayesA. However, the distribution tails are similar between the two methods (**Figure 4**).

The BayesA method implies a large number of markers with small effects or a few markers with moderate to large effects. B-Lasso implies a large number of markers with effects close to zero or a few markers with moderate to large effects. RR-BLUP implies a large number of markers with small effects.

## Deep Learning

Machine-learning algorithms (random forest, bagging, support vector machine, and others) have been successful in recognizing complex patterns and making correct decisions based on data.



Machine learning is a science of creating and studying algorithms that improve their own behavior in an iterative manner by design (Beysolow, 2017). Recent developments in machine learning enable the implementation of high-dimensional regression using nonlinear methods (Bellot et al., 2018). Another class of models, indeed a subfield of machine learning that became more used to prediction in recent times is deep learning. This theory is devoted to building algorithms that explain and learn a high and low level of abstractions of data that traditional machine learning algorithms often cannot (Beysolow, 2017).

Bellot et al. (2018) present an application of deep learning for the prediction of complex traits comparing the Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN) with commonly used linear regression methods (BayesB and BayesRR). The deep learning under the linear regression, in some cases was very competitive. The MLP and CNN are very heterogeneous classes of predictor that depend on the number of layers, number of neurons per layer, and the activation function. However, the predictive accuracy of Bayesian linear methods is highly dependent on the heritability and this is not a main factor in MLP and CNN.

## GWAS via the BayesC $\pi$ and BayesD $\pi$ Methods

Gene discovery or GWAS, which can be accomplished by the BayesC $\pi$  and BayesD $\pi$  methods (described by Habier et al., 2011), are advantageous because they provide information on the genetic architecture of the quantitative trait and identify the QTL positions by modeling the frequencies of SNPs with nonzero effects. They are advantageous over the regression analysis of single markers because they simultaneously account for all markers. However, care needs to be taken whenever the number of markers is larger than the number of individuals genotyped and phenotyped. Gianola (2013) showed that in such cases, the prior in Bayesian approaches such as BayesC and BayesD, is always influential, which could affect the inference of whether a marker is associated with the trait.

In the BayesC method, a common variance is specified for all loci. The BayesD method maintains the specific variances for each locus. Additionally,  $\pi$  is treated as an unknown with a uniform *a priori* distribution (0,1), thus producing the BayesC $\pi$  and BayesD $\pi$  methods. The modeling of  $\pi$  is interesting in the association analysis. The majority of the markers are not in LD with the genes; therefore, a set of markers associated with a trait must be identified. In contrast, the BayesB method determines  $\pi$  subjectively. Using the indicator variable  $\delta_i$ , the BayesC $\pi$  and BayesD $\pi$  methods model the additive genetic effect of individual  $j$  as  $a_j = \sum_{i=1}^n \beta_i x_{ij} \delta_i$ , where  $\delta_i = (0, 1)$ . The distribution of  $\delta = (\delta_1, \delta_n)$  is binomial with a probability of  $\pi$ . This mixed model is more parsimonious than the BayesB method. According to the model hierarchy, a distribution must be postulated for  $\pi$ , and there must be a beta distribution, which when appropriately specified, becomes a uniform distribution (0,1) (Legarra et al., 2011).

The quantities for  $x_{ij}$  are elements of the codominant marker genotype vector and are generally coded as 0, 1 or 2, depending on the number of copies of one of the alleles at the marker locus  $i$ , and  $\beta_i$  is defined as the element of the vector of the regression

coefficients, which includes the marker effects on a phenotypic trait  $y$  by means of the LD with the genes that control the trait (Resende et al., 2013).

### Sample Size for GS and GWAS

Genomic data are especially useful for GS, which allows selection at the seedling stage to increase genetic gain in the adult stage. With a high density of markers, the expected squared accuracy of GS is given by Daetwyler et al. (2008); Resende et al. (2008); Goddard et al. (2011); Grattapaglia and Resende (2011)

$$r_{gg}^2 = \frac{Nh^2}{(Nh^2 + n_{QTL})} = \frac{Nh^2}{(Nh^2 + m_e)} = \frac{Nh^2}{(Nh^2 + 2N_e L)} = \frac{Nh^2}{(Nh^2 + \frac{L}{F})}$$

where  $N$  is the number of genotyped and phenotyped individuals,  $L$  is the genome size (in Morgans) of the species,  $m_e$  is the number of independent chromosomal segments,  $N_e$  is the effective population size, and  $F$  is the inbreeding coefficient of the population. For a desired  $r_{gg}^2$ ,  $h^2$ , and  $n_{QTL}$ ,  $N$  can be determined.

The reliability of GS is given by the expression  $r_{gg}^2 = \frac{Nh^2}{Nh^2 + N_{QTL}}$ , where  $r_{gg}$  equals GS accuracy,  $N$  is the number of individuals in the population,  $N_{QTL}$  is the number of QTLs that control each trait, and  $h^2$  is the individual heritability. The estimate of the number of individuals that must be evaluated to obtain the desired accuracy can be obtained by the following expression, derived from the previous one,  $N = \frac{r_{gg}^2 N_{QTL}}{(1 - r_{gg}^2) h^2}$  (Resende et al., 2014).

**Figure 5** shows the curve graphs with  $N$  in various scenarios (functions of  $h^2$ ,  $N_{QTL}$ , and  $r_{gg}$ ). Based on these graphs and the genetic information of the traits, breeders can adequately size their studies on inheritance and maximize genetic gain with the improvement made by selection.

Various kinds of information can be obtained from **Figure 5**. For example, considering scenario 3, it appears that for a trait with individual heritability equal to 0.30 and that is controlled by 100 QTLs, an accuracy of 90% can be obtained if the sample size is equal to 1,500 genotyped and phenotyped individuals.

From the first equation, the estimate of the number of QTLs that control each trait can be calculated based on the expression  $N_{QTL} = \frac{(1 - r_{gg}^2) N h^2}{r_{gg}^2}$ . Once the selective accuracy and heritability are estimated, given the  $N$  practiced in a study, the  $N_{QTL}$  can be estimated for several traits.

A possible exercise is the theoretical determination of  $N_{QTL}$ , given the  $N$  and the estimated  $h^2$  while varying  $r_{gg}^2$ . For a case of  $h^2$  equal to 0.30, and  $N$  equal to 1,500, the  $N_{QTL}$  values can be inferred according to **Figure 6**. The same figure shows the case of  $h^2$  equal to 0.20, and  $N$  equal to 1,500.

Based on **Figure 6**, it can be seen that for  $N = 1,500$  genotyped individuals, the QTL numbers vary from 49 to 468 (Scenario 1,  $h^2 = 0.30$ ) and from 32 to 312 (Scenario 2,  $h^2 = 0.20$ ), when the accuracy varies from 0.95 to 0.70, respectively.

### Sample Size for Gene Detection

The sample size ( $N$ ), with power of detection at a significance level of  $10^{-5}$  according to the  $h_{mi}^2$  magnitude of the QTL (considered as random effect), is given by Resende (2015)  $N \approx$

$\frac{(Z_{(1-\frac{\alpha}{2})} + Z_{(1-\beta)})^2 (1-h^2)}{h_{mi}^2}$ , where  $Z_{(1-\frac{\alpha}{2})}$  and  $Z_{(1-\beta)}$  are the values of the cumulative distribution function of the standard normal distribution, associated with the probabilities of error type I ( $\alpha$ ) and error type II ( $\beta$ ) for bilateral hypothesis tests.

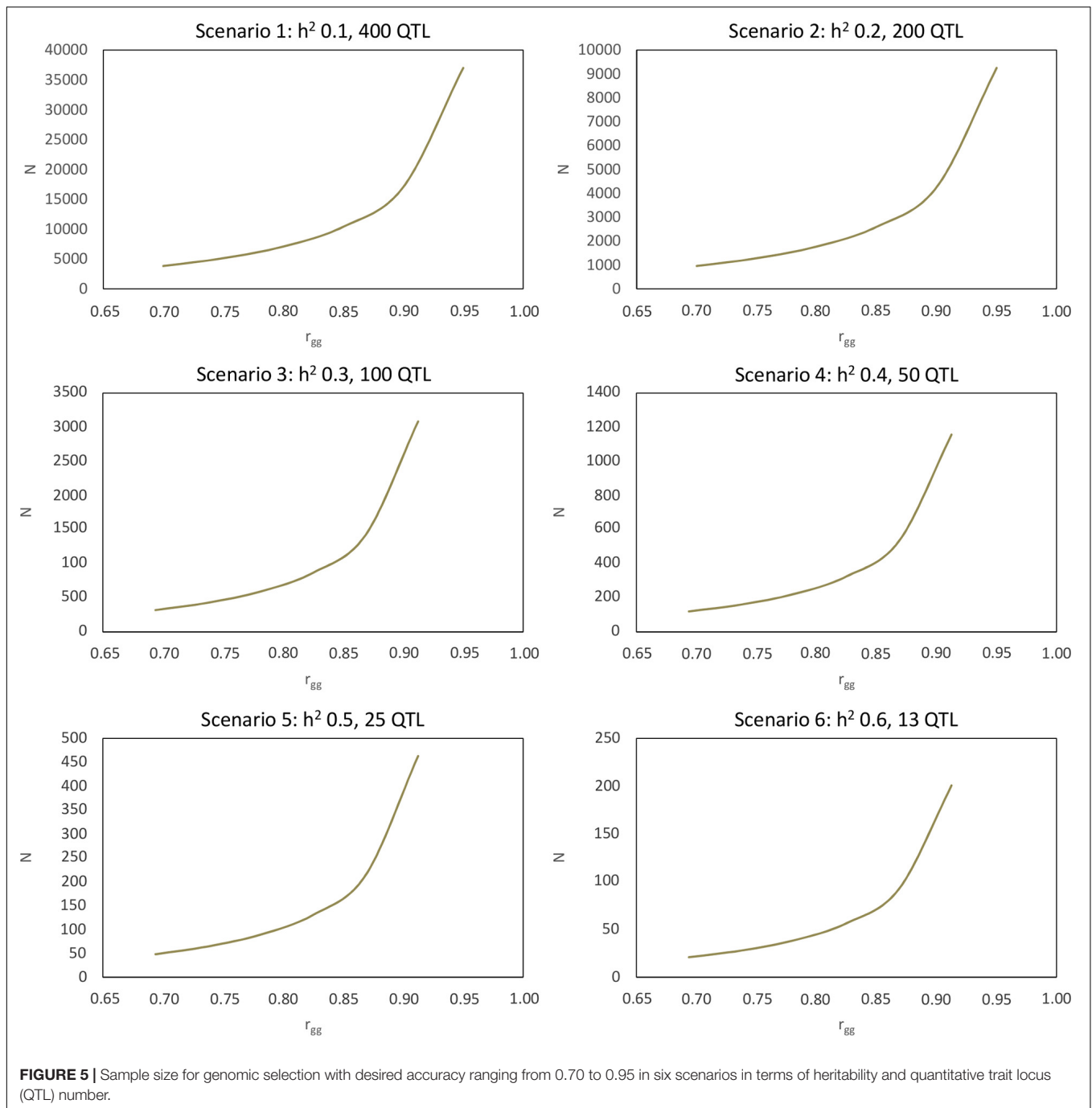
The quantity  $(1 - \beta)$  is the probability that the experiment will exhibit a statistically significant difference between the treatment averages. Values of 0.80 and 0.90 are common and appropriate in practice.

**Table 2** and **Figure 7** show that the sample sizes ( $< 1,000$ ) commonly used in plant breeding only detect QTL when the QTL explains 5% or more of the phenotypic variation, a fact that is unlikely under polygenic inheritance (total trait  $h^2 < 0.50$ ). The power of 0.90 is more appropriate because it leads to an  $81\% = 0.90^2$  probability that two independent studies will detect the same QTL.

## GS APPLICATIONS IN BRACHIARIA, GUINEA GRASS AND ELEPHANT GRASS BREEDING

In TFG breeding, combining conventional breeding efforts and GS has not been a simple task. As opposed to advances in animal breeding and crop commodities, they have been slow and challenging in tropical forages. The number of candidate TFG species is high, and decisions about investments need to be made considering the effective benefits of GS, the potential profit that can be achieved by the new cultivars, and the real impact of new forage on livestock production. The three tropical genus/species brachiaria (*Urochloa* spp.), Guinea grass (*M. maximus*) and elephant grass (*C. purpureus*) are very important and extensively used as pastures in tropical America, Asia, and sub-Saharan Africa.

Marker number and density are important factors influencing the efficient use of GS in TFG breeding. One of the reasons for the low accuracy of GS is the exceptionally low number of effective markers, which may result from a non-adequate reference genome. Before the *U. ruziziensis* genome assembly was publicly available, the genomes of *S. viridis* and *P. virgatum* were frequently used as reference genomes for SNP calling and linkage map construction in *Urochloa* species with 1,000 SNPs (Ferreira et al., 2019) and *M. maximus* with 1,322 SNPs (Deo et al., 2020). For *C. purpureus*, 20,144 SilicoDArT and 28,610 SNP markers have been mapped onto the pearl millet (*C. americanus*) reference (Muktar et al., 2019). Compared with other agronomically important *Poaceae* species, such as maize, for which the 50 K Illumina MaizeSNP50 BeadChip (Ganal et al., 2011) and the 600 K Affymetrix Axiom Maize Genotyping Array (Unterseer et al., 2014) are available, the number of markers in TFG needs to be significantly enhanced. Genome calling using reference genomes of other grasses improves the number of SNPs, as shown by Matias et al. (2019a), who reported  $> 26k$  SNPs in *Urochloa* hybrids. However, the minimum allele depth used was  $\leq 2$  reads considering allele dosage and resulted in a low predictive ability ( $< 0.31$ ) in GS for agronomic traits. Similar

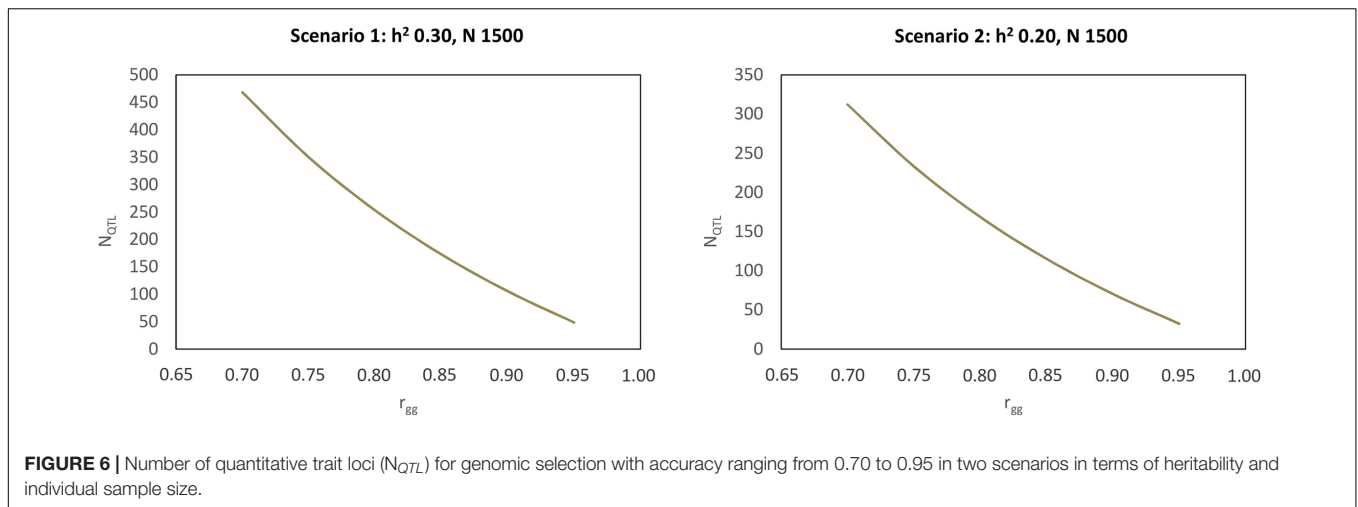


results were obtained by Lara et al. (2019) in *M. maximus* in which >32k SNPs were classified as unique and used in GS, although, the maximum value of predictive ability using tetraploid dosage of 0.3955 was achieved for the trait organic matter that displayed secondary importance in forage breeding.

Current TFG breeding programs lack the important information that could improve and allow the efficient use of GS. Firstly, a major impact will be obtained by increasing the number of markers per genome size by sequencing and generating reference genomes for the target species or more closely related

species. Secondly, we need to improve our knowledge about the inheritance of target traits in tropical forages including the genetic effects of biallelism in (auto) tetraploids. Thirdly, we need to work with training populations connected with validation and breeding populations and testing environments that must be correlated with the environment of the target population (Burgueño et al., 2012; Jarquín et al., 2014; Santantonio et al., 2020). Finally, we need to work with half-sib and full-sib progenies to improve the predictive ability for traits in which the dominance effects are significant, aiming to predict crosses



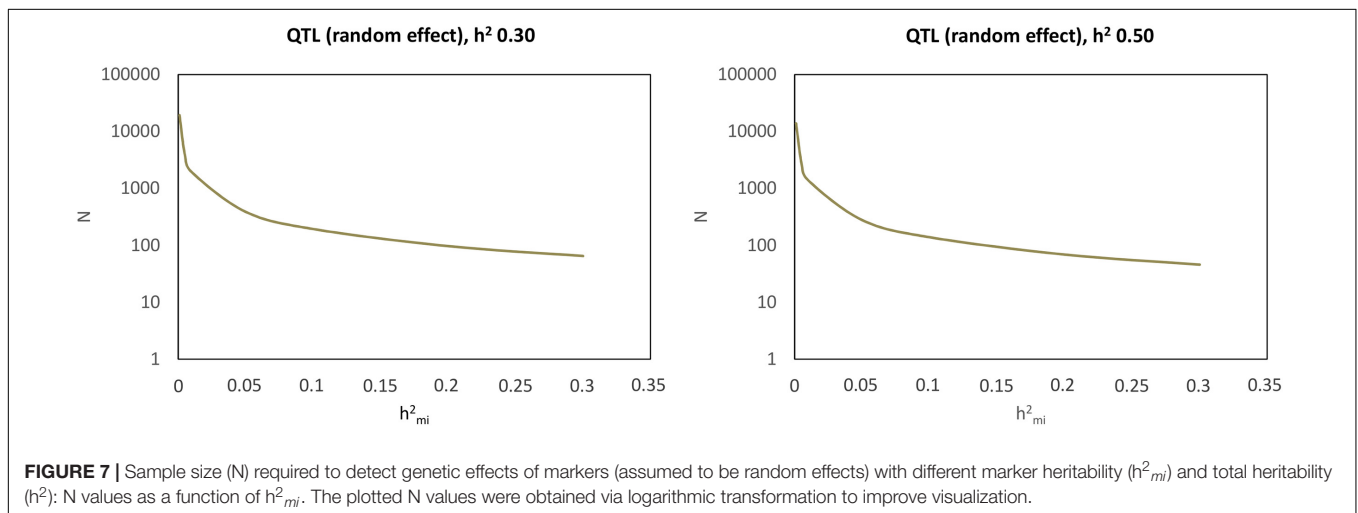


**FIGURE 6 |** Number of quantitative trait loci ( $N_{QTL}$ ) for genomic selection with accuracy ranging from 0.70 to 0.95 in two scenarios in terms of heritability and individual sample size.

**TABLE 2 |** Sample size ( $N$ ) and power for detection of significance level  $10^{-5}$  according to the  $h^2_{mi}$  magnitude of the quantitative trait locus, considered as having a

random effect: 
$$N \approx \frac{\left( Z_{(1-\frac{\alpha}{2})} + Z_{(1-\beta)} \right)^2 (1-h^2)}{h^2_{mi}}$$

$h^2 = 0.30$					$h^2 = 0.50$				
Z for $\beta = 0.90$	Z for $\alpha = 10^{-5}$	$\left( Z_{(1-\frac{\alpha}{2})} + Z_{(1-\beta)} \right)^2$	$h^2_{mi}$	N	Z for $\beta = 0.90$	Z for $\alpha = 10^{-5}$	$\left( Z_{(1-\frac{\alpha}{2})} + Z_{(1-\beta)} \right)^2$	$h^2_{mi}$	N
1.28	3.99	27.7729	0.001	19441	1.28	3.99	27.7729	0.001	13886
1.28	3.99	27.7729	0.005	3888	1.28	3.99	27.7729	0.005	2777
1.28	3.99	27.7729	0.01	1944	1.28	3.99	27.7729	0.01	1389
1.28	3.99	27.7729	0.05	389	1.28	3.99	27.7729	0.05	278
1.28	3.99	27.7729	0.1	194	1.28	3.99	27.7729	0.1	139
1.28	3.99	27.7729	0.2	97	1.28	3.99	27.7729	0.2	69
1.28	3.99	27.7729	0.3	65	1.28	3.99	27.7729	0.3	46



**FIGURE 7 |** Sample size ( $N$ ) required to detect genetic effects of markers (assumed to be random effects) with different marker heritability ( $h^2_{mi}$ ) and total heritability ( $h^2$ ):  $N$  values as a function of  $h^2_{mi}$ . The plotted  $N$  values were obtained via logarithmic transformation to improve visualization.

that are not performed in tetraploid *Urochloa* and *M. maximus* hybrid development programs.

The current *M. maximus* and *Urochloa* breeding programs generate thousands of hybrids annually, and those hybrids must pass through several steps of selection until they finally achieve the status for evaluation under animal feeding pressure to

prove their value in animal production and be released as new cultivars. However, it is mandatory for hybrids to show apomixis and resistance/tolerance to spittlebugs (mostly in *Urochloa*) and diseases (mostly in *M. maximus* and *C. purpureus*). These traits are a great bottleneck slowing the subsequent evaluation steps in the breeding program since phenotyping of individuals

demands significant labor, time (two or more years), and the dedication of trained technicians. As a result, only a small number of individuals can be evaluated annually reducing the rate of genetic gain. Methods such as GWAS and MAS may help to speed up the identification of individuals showing these important traits and increased rates of genetic gain. Recently published data on mapping genomic regions associated with apospory emphasize the routine application of markers for selection in *Urochloa* and *Megathyrus* (Worthington et al., 2016; Deo et al., 2020).

Finally, TFG breeding programs can be sped up by the application of GS. However, this demands greater investments; collaboration among breeders, molecular biologists, and bioinformaticians; integration of research teams from different institutions and countries; and most importantly, continuity associated with critical course corrections.

## ADDITIONAL METHODS APPLIED TO TFG BREEDING

The use of new efficient high-throughput methodologies in addition to GS should be discussed according to their accuracies and potential use in higher numbers of individuals at initial stages of selection. The first obvious application is its use when no genotyping tool is available at a reasonable cost. A second application would be to use phenomics to screen nearly fixed genetic materials which is likely to capture non-additive genetic effects. Nonetheless, phenomics could deliver breeding innovations, and the challenge represented by the breeding target scenario (Reynolds et al., 2020).

Phenomic selection (PS) using high-throughput phenotyping methods less expensive than genotyping by sequencing is an opportunity for tropical forage breeding. Rincent et al. (2018) proposed using near-infrared spectroscopy (NIRS) variables generated as regressors or to estimate kinship in the same statistical models used in GS to perform PS. The results were promising and cost affordable for wheat and poplar when compared to GS. TFGs are a probable candidate for this method because phenotyping using NIRS to obtain bromatological data is routine in research programs and may be studied and amplified to other spectra to be performed as a routine method of PS. Biomass measuring in TFG is a laborious, time-consuming, and biased task, because of the necessity of several annual evaluations (4 to 7) during selection. It also limits the number of individuals in experiments (300 to 2,000). Sensor-based images enabled high-throughput non-invasive phenotyping throughout the growing cycles of forage grasses, and the models established a high correlation between images and the biomass yield in *M. maximus* (Castro et al., 2020) as well as for crude protein percentage and chlorophyll concentration in *Urochloa* (Jiménez et al., 2020). Deep learning-based neural network studies demonstrated that accuracies must be increased by pre-trained models and data augmentation (Castro et al., 2020). Nevertheless, deep learning progress is accelerating and will be able to perform better predictions than ever

(Montesinos-López et al., 2021). Although it has been the subject of debate in the past, extra investment in phenotyping technologies is becoming more accepted to capitalize on recent developments in crop genomics and prediction models. In this context the different strategies for phenotyping can be built from phenomic selection (Rincent et al., 2018), high-throughput phenotyping, and detailed characterization or 'precision' phenotyping (Reynolds et al., 2020).

## THE FUTURE OF GS IN FORAGE BREEDING

The availability of genome-wide, high-throughput, and cost-effective flexible markers, across the genome, suitable for large populations with or without a reference genome sequence, is the most important factor for the effective and efficient implementation of GS. Recent advances in long read quality and sequence throughput, in addition to other technologies such as Hi-C or optical maps, make it possible for virtually any research group with reasonable funding to obtain reference-grade genome assemblies for their crop of choice. While not necessarily easy, the generation of high-quality genome assemblies should be considered as a starting point for any orphan species that would benefit from the use of genomic tools for crop improvement. These assemblies could be extremely useful for resequencing and variant discovery, which can lead to genotyping platforms for association studies and GS. When coupled with well-designed and thoroughly phenotyped training populations, these genomic resources could serve as the basis for implementing GS steps in the breeding of TFGs.

As discussed by Lin et al. (2014) and Bhat et al. (2016), the cost of identifying and genotyping a large number of SNPs is still a barrier for TFGs, although, second-generation sequencing technology has provided new SNP genotyping platforms, particularly GBS. In addition, phenotyping large representative reference populations is expensive. Reduction of phenotype assessment costs per individual and new phenomic approaches are essential to take advantage of the true benefits of GS. Marker technologies must be combined with high-throughput phenotyping to achieve significant genetic gains for complex traits.

Furthermore, the considerations stated by Simeão-Resende et al. (2014) are still valid. GS will allow an increase in the early-generation of number of individuals evaluated considering the large number of targeted traits. However, when we deal with GS in the improvement of tropical forages, we realize that there is still a long way to go. Theoretically, by models and methods already developed and successfully applied in commodity species, the procedures could be easily incorporated into the routine of breeding programs. Nevertheless, in orphan species, all knowledge needs to be built on solid molecular bases. In principle, the evaluation of a large number of individuals for selection purposes increases the probability of the best allelic combinations for traits of economic importance without narrowing the genetic basis for selection. This should be considered in the improvement of polyploid and

apomictic *Urochloa* spp. and *M. maximus*. Performing inter- and intraspecific crosses with sexual plants in these genera increases the variability available for selection and allows the generation of genetic combinations not found in apomictic accessions. The spectrum of possibilities for GS expands considerably for these species; however, the identification of markers is narrowed to large-scale phenotyping and genotyping. The conformation of the discovery population should be carefully considered in terms of the number of hybrid families to be evaluated, the number of individuals per half-sib and full-sib families, and the distribution of markers on the chromosomes of the paternal and maternal genomes. The mother plants to be used in crosses should be exclusively sexual, so that they do not generate, in addition to hybrids, their own clones (by apomixis) in the progeny, which would cause an incorrect bias in the determination of GEBV and the identification of markers and their effects.

Finally, designing forage breeding programs, mainly for polyploid and apomictic grasses, and proposing breeding schemes that make optimum use of GS is a significant task for plant breeders. Although, this is a challenge, it is also a great opportunity to accelerate genetic gain in TFG breeding.

## REFERENCES

- Alexandratos, N., and Bruinsma, J. (2012). *World Agriculture Towards 2030/2050: The 2012 Revision. ESA Working Paper 12-03*. Rome: FAO, doi: 10.22004/ag.econ.288998
- Azevedo, C. F., Resende, M. D. V., Silva, F. F., Viana, J. M. S., Valente, M. S. F., Resende, M. F. R. Jr., et al. (2015). Ridge, Lasso and Bayesian additive dominance genomic models. *BMC Genet.* 16:105. doi: 10.1186/s12863-015-0264-2
- Bajželj, B., Richards, K. S., Allwood, J. M., Smith, P., Dennis, J. S., Curmi, E., et al. (2014). Importance of food-demand management for climate mitigation. *Nat. Clim. Change* 4, 924–929. doi: 10.1038/nclimate2353
- Ballesta, P., Bush, D., Silva, F. F., and Mora, F. (2020). Genomic predictions using low-density SNP markers, pedigree and GWAS information: a case study with the non-model species *Eucalyptus cladocalyx*. *Plants* 9:99. doi: 10.3390/plants9010099
- Barabaschi, D., Tondelli, A., Desiderio, F., Volante, A., Vaccino, P., Valè, G., et al. (2016). Next generation breeding. *Plant Sci.* 242, 3–13. doi: 10.1016/j.plantsci.2015.07.010
- Barcaccia, G., Palumbo, F., Sgorbati, S., Albertini, E., and Pupilli, F. (2020). A reappraisal of the evolutionary and developmental pathway of apomixis and its genetic control in angiosperms. *Genes* 11:859. doi: 10.3390/genes11080859
- Barrios, S. C., Valle, C. B., Alves, G. F., Simeão, R. M., and Jank, L. (2013). Reciprocal recurrent selection in the breeding of *Brachiaria decumbens*. *Trop. Grassl Forrajes Trop.* 1, 52–54.
- Bellot, P., De Los Campos, G., and Pérez-Enciso, M. (2018). Can deep learning improve genomic prediction of complex human traits? *Genetics* 210, 809–819. doi: 10.1534/genetics.118.301298
- Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4
- Benevenuto, J., Ferrão, L. F. V., Amadeu, R. R., and Munoz, P. (2019). How can a high-quality genome assembly help plant breeders? *GigaScience* 8:giz068. doi: 10.1093/gigascience/giz068
- Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci.* 34, 20–25. doi: 10.2135/cropsci1994.0011183X003400010003x

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

We acknowledge the financial support from Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Instituto Nacional de Ciência e Tecnologia do Café (INCT Café), and Associação para o Fomento à Pesquisa de Melhoramento de Forrageiras (UNIPASTO).

## ACKNOWLEDGMENTS

We are thankful to João Romero do Amaral Santos de Carvalho Rocha and Marco Antonio Machado for valuable comments and discussions.

- Beysolow, T. II (2017). “Introduction to deep learning,” in *Introduction to Deep Learning Using R*, (Berkeley, CA: Apress), 1–9.
- Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., et al. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.* 7:221. doi: 10.3389/fgene.2016.00221
- Bourke, P. M., Voorrips, R. E., Visser, R. G. F., and Maliepaard, C. (2018). Tools for genetics studies in experimental populations of polyploids. *Front. Plant Sci.* 9:153. doi: 10.3389/fpls.2018.00513
- Burgueño, J., de los Campos, G., Weigel, K., and Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52, 707–719. doi: 10.2135/cropsci2011.06.0299
- Castro, W., Marcato-Junior, J., Polidoro, C., Osco, L. P., Gonçalves, W., Rodrigues, L., et al. (2020). Deep learning applied to phenotyping of biomass in forages with UAV-based RGB imagery. *Sensors* 20:4802. doi: 10.3390/s20174802
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011
- Daetwyler, H. D., Villanueva, B., Bijma, P., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3:e3395. doi: 10.1371/journal.pone.0003395
- Daud, Z., Hatta, M., Kassim, A., Mohd Aripin, A., and Awang, H. (2014). Analysis of Napier grass (*Pennisetum purpureum*) as a potential alternative fiber in paper industry. *Mater. Res. Innov.* 18, 18–20. doi: 10.1179/1432891714Z.000000000925
- de Bem Oliveira, I., Resende, M. F. R., Ferrão, L. F. V., Amadeu, R. R., Endelman, J. B., Kirst, M., et al. (2019). Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. *G3* 9, 1189–1198. doi: 10.1534/g3.119.400259
- Dekkers, J. C. M. (2007). Prediction of response to marker assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* 124, 331–341. doi: 10.1111/j.1439-0388.2007.00701.x
- Della Coletta, R., Qiu, Y., Ou, S., Hufford, M. B., and Hirsch, C. N. (2021). How the pan-genome is changing crop genomics and improvement. *Genome Biol.* 22:3. doi: 10.1186/s13059-020-02224-8

- Deo, T. G., Ferreira, R. C. U., Lara, L. A. C., Moraes, A. C. L., Alves-Pereira, A., de Oliveira, F. A., et al. (2020). High-resolution linkage map with allele dosage allows the identification of regions governing complex traits and apospory in Guinea Grass (*Megathyrsus maximus*). *Front. Plant Sci.* 11:15. doi: 10.3389/fpls.2020.00015
- Eberhart, S. A. (1970). Factors affecting efficiencies of breeding methods. *Afr. Soils* 15, 655–680.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Fernando, R. L. (1998). “Genetic evaluation and selection using genotypic, phenotypic and pedigree information,” in *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, Armidale, NSW, 329–336.
- Ferreira, R. C. U., Lara, L. A., de, C., Chiari, L., Barrios, S. C. L., do Valle, C. B., et al. (2019). Genetic mapping with allele dosage information in tetraploid *Urochloa decumbens* (Stapf) R. D. Webster reveals insights into spittlebug (*Nototuzilia enterriana* Berg) resistance. *Front. Plant Sci.* 10:92. doi: 10.3389/fpls.2019.00092
- Fuzinato, V. A., Pagliarini, M. S., and Valle, C. B. (2007). Microsporogenesis in sexual *Brachiaria* hybrids (Poaceae). *Genet. Mol. Res.* 6, 1107–1117.
- Gallais, A. (1989). Concepts of varietal value and of test value in autotetraploids: application to genetic advance in population improvement. *Genome* 32, 420–424. doi: 10.1139/g89-465
- Gallais, A. (2003). *Quantitative Genetics and Breeding Methods in Autopolyploid Plants*. Paris: Institut National de la Recherche Agronomique.
- Ganal, M. W., Durstewitz, G., Polley, A., Berard, A., Buckler, E. S., Charcosset, A., et al. (2011). A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS One* 6:e28334. doi: 10.1371/journal.pone.0028334
- Gianola, D. (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 194, 573–596. doi: 10.1534/genetics.113.151753
- Gianola, D., Perez-Enciso, M., and Toro, M. A. (2003). On marker-assisted prediction of genetic value: beyond the ridge. *Genetics* 163, 347–365.
- Goddard, M. E., and Hayes, B. J. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330. doi: 10.1111/j.1439-0388.2007.00702.x
- Goddard, M. E., Hayes, B. J., and Meuwissen, T. H. E. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* 128, 409–421. doi: 10.1111/j.1439-0388.2011.00964.x
- Grattapaglia, D., and Resende, M. D. V. (2011). Genomic selection in forest tree breeding. *Tree Genet. Genomes* 7, 241–255. doi: 10.1007/s11295-010-0328-4
- Guo, Z., Tucker, D. M., Lu, J., Kishore, V., and Gay, G. (2012). Evaluation of genome-wide selection efficiency in maize nested association mapping populations. *Theor. Appl. Genet.* 124, 261–275. doi: 10.1007/s00122-011-1702-9
- Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi: 10.1534/genetics.107.081190
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186
- Habte, E., Muktar, M. S., Abdena, A., Hanson, J., Sartie, A. M., Negawo, A. T., et al. (2020). Forage performance and detection of marker trait associations with potential for napier grass (*Cenchrus purpureus*) improvement. *Agronomy* 10:542. doi: 10.3390/agronomy10040542
- Haley, C. S., and Visscher, P. M. (1998). Strategies to utilize marker-quantitative trait loci associations. *J. Dairy Sci.* 81, 85–97. doi: 10.3168/jds.S0022-0302(98)70157-2
- Hallauer, A. R. (2011). Evolution of plant breeding. *Crop Breed. Appl. Biotechnol.* 11, 197–206. doi: 10.1590/S1984-70332011000300001
- Hand, M. L., and Koltunow, A. M. G. (2014). The genetic control of apomixis: asexual seed formation. *Genetics* 197, 441–450. doi: 10.1534/genetics.114.163105
- Hanna, W. W. (1981). Method of reproduction in napiergrass and in the 3X and 6X allopolyploid hybrids with pearl millet. *Crop Sci.* 21, 123–126. doi: 10.2135/cropsci1981.0011183X002100010033x
- Hayes, B. J., Cogan, N. O. I., Pembleton, L. W., Goddard, M. E., Wang, J., Spangenberg, G., et al. (2013). Prospects for genomic selection in forage plant species. *Plant Breed.* 132, 133–143. doi: 10.1111/pbr.12037
- Heffner, E. L., Lorenz, A. J., Jannink, J. L., and Sorrells, M. E. (2010). Plant breeding with Genomic selection: gain per unit time and cost. *Crop Sci.* 50, 1681–1690. doi: 10.2135/cropsci2009.11.0662
- IBGE (2016). *Mudanças na Cobertura e Uso da Terra do Brasil 2000 – 2010 – 2012 – 2014*. Rio de Janeiro: IBGE, 29.
- Jank, L., Barrios, S. C., do Valle, C. B., Simeão, R. M., and Alves, G. F. (2014). The value of improved pastures to Brazilian beef production. *Crop Pasture Sci.* 65, 1132–1137. doi: 10.1071/CP13319
- Jank, L., Valle, C. D., and Resende, R. M. S. (2011). Breeding tropical forages. *Crop Breed. Appl. Biotechnol.* 11, 27–34. doi: 10.1590/S1984-70332011000500005
- Jarquín, D., Crossa, J., Lacaze, X., Cheyron, P. D., Daucourt, J., Lorgeou, J., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127, 595–607. doi: 10.1007/s00122-013-2243-1
- Jiménez, J. C., Leiva, L., Cardoso, J. A., French, A. N., and Thorp, K. R. (2020). Proximal sensing of *Urochloa* grasses increases selection accuracy. *Crop Pasture Sci.* 71, 401–409. doi: 10.1071/CP19324
- Jones, C., Vega, J., Worthington, M., Thomas, A., Gasior, D., Harper, J., et al. (2021). A comparison of differential gene expression in response to the onset of water stress between three hybrid *Brachiaria* genotypes. *Front. Plant Sci.* 12:637956. doi: 10.3389/fpls.2021.637956
- Kohavi, R. (1995). “A study of cross-validation and bootstrap for estimation and model selection,” in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, ed. C. S. Mellish (San Francisco, CA: Morgan Kaufmann Publishers), 1137–1143.
- Lamkey, K. R., and Edwards, F. (1999). “Quantitative genetics of heterosis,” in *The Genetics and Exploitation of Heterosis in Crops*, eds J. G. Coors and S. Pandey (Madison, WI: ASA/CSSA/SSSA), 29–43.
- Lande, R., and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756.
- Lara, L., de, C., Santos, M. F., Jank, L., Chiari, L., Vilela, M. M., et al. (2019). Genomic selection with allele dosage in *Panicum maximum* Jacq. *G3* 9, 2463–2475. doi: 10.1534/g3.118.200986
- Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F., and Fritz, S. (2011). Improved Lasso for genomic selection. *Genet. Res.* 93, 77–87. doi: 10.1017/S0016672310000534
- Lima, L. P., Azevedo, C. F., Resende, M. D. V., Silva, F. F., Suela, M. M., Nascimento, M., et al. (2019a). New insights into genomic selection through population-based non-parametric prediction methods. *Sci. Agric.* 76, 290–298. doi: 10.1590/1678-992x-2017-0351
- Lima, L. P., Azevedo, C. F., Resende, M. D. V., Silva, F. F., Viana, J. M. S., and Oliveira, E. J. (2019b). Triple categorical regression for genomic selection: application to cassava breeding. *Sci. Agric.* 76, 368–375. doi: 10.1590/1678-992x-2017-0369
- Lin, Z., Hayes, B., and Daetwyler, H. (2014). Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* 65, 1177–1191. doi: 10.1071/CP13363
- Lovell, J. T., Jenkins, J., Lowry, D. B., Mamidi, S., Sreedasyam, A., Weng, X., et al. (2018). The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*. *Nat. Commun.* 9:5213. doi: 10.1038/s41467-018-07669-x
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., et al. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9:e1003215. doi: 10.1371/journal.pgen.1003215
- Lush, J. (1937). *Animal Breeding Plans*. Ames: Iowa State College Press.
- Machado, J. C., Pereira, J. F., Azevedo, A. L. S., Mittelman, A., Pereira, A. V., Sobrinho, F. S., et al. (2019). “Melhoramento genético de forrageiras e o uso de ferramentas genômicas,” in *Melhoramento de Forrageiras na Era Genômica*, eds A. L. S. Azevedo, J. F. Pereira, and J. C. Machado (Brasília: Embrapa), 11–42.
- Mackay, I., Piepho, H. P., and Garcia, A. A. F. (2019). “Statistical methods for plant breeding,” in *Handbook of Statistical Genomics*, eds D. J. Balding, I. Moltke, and J. Marion (Hoboken, NJ: John Wiley & Sons), 501–521.
- Matias, F. I., Alves, F. C., Meireles, K. G. X., Barrios, S. C. L., Valle, C. B., Endelman, J. B., et al. (2019a). On the accuracy of genomic prediction models considering multi-trait and allele dosage in *Urochloa* spp. interspecific tetraploid hybrids. *Mol. Breed.* 39:100. doi: 10.1007/s11032-019-1002-7
- Matias, F. I., Meireles, K. G. X., Nagamatsu, S. T., Barrios, S. C. L., Valle, C. B., Carazzolle, M. F., et al. (2019b). Expected genotype quality and diploidized

- marker data from genotyping-by-sequencing of *Urochloa* spp. tetraploids. *Plant Genome* 12:190002. doi: 10.3835/plantgenome2019.01.0002
- McCormick, R. F., Truong, S. K., and Mullet, J. E. (2015). RIG: recalibration and interrelation of genomic sequence data with the GATK. *G* 3, 655–665. doi: 10.1534/g3.115.017012
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytksy, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Melo, A. T. O., Bartaula, R., and Hale, I. (2016). GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics* 17:29. doi: 10.1186/s12859-016-0879-y
- Mendes-Bonato, A. B., Pagliarini, M. S., and Valle, C. B. (2007). Meiotic arrest compromises pollen fertility in an interspecific hybrid between *Brachiaria ruziziensis* x *Brachiaria decumbens* (Poaceae: Paniceae). *Braz. Arch. Biol. Technol.* 50, 831–837. doi: 10.1590/S1516-89132007000500011
- Mendes-Bonato, A. B., Risso-Pascotto, C., Pagliarini, M. S., and Valle, C. B. (2006). Cytogenetic evidence for genome elimination during microsporogenesis in an interspecific hybrid between *Brachiaria ruziziensis* and *B. brizantha* (Poaceae). *Genet. Mol. Biol.* 29, 711–714. doi: 10.1590/S1415-47572006000400021
- Meuwissen, T. H. E. (2007). Genomic selection: marker assisted selection on genome-wide scale. *J. Anim. Breed. Genet.* 124, 321–322. doi: 10.1111/j.1439-0388.2007.00708.x
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Miles, J. W., Cardona, C., and Sotelo, G. (2006). Recurrent selection in a synthetic *Brachiariagrass* population improves resistance to three spittlebug species. *Crop Sci.* 46, 1008–1093. doi: 10.2135/cropsci2005.06-0101
- Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92, 4648–4655. doi: 10.3168/jds.2009-2064
- Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W. R., Fajardo-Flores, S., et al. (2021). A review of deep learning applications for genomic selection. *BMC Genomics* 22:19. doi: 10.1186/s12864-020-07319-x
- Morais, R. F. D., Souza, B. J. D., Leite, J. M., Soares, L. H. D. B., Alves, B. J. R., Boddey, R. M., et al. (2009). Elephant grass genotypes for bioenergy production by direct biomass combustion. *Pesq. Agrop. Bras.* 44, 133–140. doi: 10.1590/S0100-204X2009000200004
- Morales, F. J. (2009). “Introduction to tropical agriculture and outlook for tropical crops in a globalized economy,” in *Tropical Biology and Conservation Management*, eds K. Del Claro, P. S. Oliveira, and V. Rico-Gray (Oxford: EOLSPublishers Co Ltd), 1–27.
- Mrode, R., Coffey, M., and Berry, D. P. (2010). Understanding genomic evaluations from various evaluation methods and GMACE. *Interbull. Bull.* 42, 52–55.
- Mrode, R. A. (2014). *Linear Models for the Prediction of Animal Breeding Values*. Wallingford: CAB International.
- Muktar, M. S., Teshome, A., Hanson, J., Negawo, A. T., Habte, E., Entfellner, J. B. D., et al. (2019). Genotyping by sequencing provides new insights into the diversity of Napier grass (*Cenchrus purpureus*) and reveals variation in genome-wide LD patterns between collections. *Sci. Rep.* 9:6936. doi: 10.1038/s41598-019-43406-0
- Negawo, A. T., Jorge, A., Hanson, J., Teshome, A., Muktar, M. S., Azevedo, A. L. S., et al. (2018). Molecular markers as a tool for germplasm acquisition to enhance the genetic diversity of a Napier grass (*Pennisetum purpureum*) collection. *Trop. Grassl. Forrajes Trop.* 6, 58–69. doi: 10.17138/TGFT(6)58-69
- Nejati-Javaremi, A., Smith, C., and Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Anim. Sci.* 75, 1738–1745. doi: 10.2527/1997.7571738x
- Nunes, S. G., Boock, A., Pentead, M. I., de, O., and Gomes, D. T. (1984). *Brachiaria brizantha* cv. *Marandu*. *Documentos Embrapa*, No. 21. Campo Grande: Embrapa.
- Ozias-Akins, P., and van Dijk, P. J. (2007). Mendelism genetics of apomixis in plants. *Ann. Rev. Genet.* 41, 509–537. doi: 10.1146/annurev.genet.40.110405.090511
- Pagliarini, M. S., Riso-Pascotto, C., Sauza-Kaneshima, A. M., and Valle, C. B. (2008). Analysis of meiotic behavior in selecting potential genitors among diploid and artificially induced tetraploid accessions of *Brachiaria ruziziensis* (Poaceae). *Euphytica* 164, 181–187. doi: 10.1007/s10681-008-9697-2
- Paudel, D., Kannan, B., Yang, X., Harris-Shultz, K., Thudi, M., Varshney, R. K., et al. (2018). Surveying the genome and constructing a high-density genetic map of napiergrass (*Cenchrus purpureus* Schumacher). *Sci. Rep.* 8:14419. doi: 10.1038/s41598-018-32674-x
- Paul, B. K., Koge, J., Maass, B. L., Notenbaert, A., Peters, M., Groot, J. C. J., et al. (2020). Tropical forage technologies can deliver multiple benefits in Sub-Saharan Africa. A meta-analysis. *Agron. Sustain. Dev.* 40:22. doi: 10.1007/s13593-020-00626-3
- Pengelly, B. C., and Maass, B. L. (2019). Tropical and subtropical forage germplasm conservation and science on their deathbed! 2. Genebanks, FAO and donors must take urgent steps to overcome the crisis. *Outlook Agric.* 48, 210–219. doi: 10.1177/0030727019867955
- Pereira, A. V., Léo, F. J. S., and Machado, J. C. (2017). BRS Kurumi and BRS Capiçã – New elephant grass cultivars for grazing and cut-and-carry system. *Crop Breed. Appl. Biotechnol.* 17, 59–62. doi: 10.1590/1984-70332017v17n1c9
- Pereira, J. F., Azevedo, A. L. S., Pessoa-Filho, M., Romanel, E. A. D. C., Pereira, A. V., Vigna, B. B. Z., et al. (2018). Research priorities for next-generation breeding of tropical forages in Brazil. *Crop Breed. Appl. Biotechnol.* 18, 314–319. doi: 10.1590/1984-70332018v18n3n46
- Pessim, C., Pagliarini, M. S., Jank, L., Kaneshima, M. A. S., and Mendes-Bonato, A. B. (2010). Meiotic behavior in *Panicum maximum* Jacq. (Poaceae: Panicoideae: Paniceae): hybrids and their genitors. *Acta Sci. Agron.* 32, 417–422. doi: 10.4025/actasciagron.v32i3.6461
- Pessim, C., Pagliarini, M. S., Silva, N., and Jank, L. (2015). Chromosome stickiness impairs meiosis and influences reproductive success in *Panicum maximum* (Poaceae) hybrid plants. *Genet. Mol. Res.* 14, 4195–4202. doi: 10.4238/2015.April.28.2
- Rao, I., Peters, M., Castro, A., Schultze-Kraft, R., White, D., Fisher, M., et al. (2015). LivestockPlus – The sustainable intensification of forage-based agricultural systems to improve livelihoods and ecosystem services in the tropics. *Trop. Grassl. Forrajes Trop.* 3, 59–82. doi: 10.17138/TGFT(3)59-82
- Reis, M. C., Sobrinho, F. S., Ramalho, M. A. P., Ferreira, D. F., Ledo, F. J. S., and Pereira, A. V. (2008). Allohexaploid pearl millet x elephantgrass population potential for a recurrent selection program. *Pesq. Agropec. Bras.* 43, 195–199. doi: 10.1590/S0100-204X2008000200006
- Resende, M. D. V. (2007). *Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético*. Colombo: Embrapa Florestas, 561.
- Resende, M. D. V. (2015). *Genética Quantitativa e de Populações*. Visconde do Rio Branco: Suprema.
- Resende, M. D. V., and Alves, R. S. (2020). Linear, generalized, hierarchical, Bayesian and random regression mixed models in genetics/genomics in plant breeding. *Funct. Plant Breed. J.* 2, 1–31.
- Resende, M. D. V., Lopes, P. S., Silva, R. L., and Pires, I. E. (2008). Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesq. Flor. Bras.* 56, 63–78.
- Resende, M. D. V., Silva, F. F., and Azevedo, C. F. (2014). *Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência*. Visconde do Rio Branco: Suprema.
- Resende, M. D. V., Silva, F. F., Lopes, P. S., and Azevedo, C. F. (2012). *Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial*. Viçosa: Universidade Federal de Viçosa.
- Resende, M. D. V., Silva, F. F., Resende, M. F. R. Jr., and Azevedo, C. F. (2013). “Genome-wide selection (GWS),” in *Biotechnology and Plant Breeding*, eds A. Borém and R. Fritsche-Neto (Amsterdam: Elsevier), 105–134.
- Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020). Breeder friendly phenotyping. *Plant Sci.* 295:1103962. doi: 10.1016/j.plantsci.2019.110396
- Rincint, R., Charpentier, J. P., Faivre-Rampant, P., Paux, E., Gouis, J. L., Bastien, C., et al. (2018). Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar. *G3* 8:3961. doi: 10.1534/g3.118.200760

- Risso-Pascotto, C., Pagliarini, M. S., and Valle, C. B. (2005). Meiotic behavior in interspecific hybrids between *Brachiaria ruziziensis* and *Brachiaria brizantha* (Poaceae). *Euphytica* 145, 155–159. doi: 10.1007/s10681-005-0893-z
- Rocha, J. R. A. S. C., Marçal, T. S., Salvador, F. V., da Silva, A. C., Carneiro, P. C. S., de Resende, M. D. V., et al. (2019). Unraveling candidate genes underlying biomass digestibility in elephant grass (*Cenchrus purpureus*). *BMC Plant Biol.* 19:548. doi: 10.1186/s12870-019-2180-5
- Sandhu, J. S., Kumar, D., Yadav, V. K., Singh, T., Sah, R. P., and Radhakrishna, A. (2015). “Recent trends in breeding of tropical grass and forage species,” in *Proceedings of the 23rd International Grassland Congress*, eds D. Vijay, M. K. Srivastava, C. K. Gupta, D. R. Malaviya, M. M. Roy, S. K. Mahanta, et al. (Jhansi: Range Management Society of India), 337–348.
- Santantonio, N., Atanda, S. A., Beyene, Y., Varshney, R. K., Olsen, M., Jones, E., et al. (2020). Strategies for effective use of genomic information in crop breeding programs serving Africa and South Asia. *Front. Plant Sci.* 11:353. doi: 10.3389/fpls.2020.00353
- Shamshad, M., and Sharma, A. (2018). The usage of genomic selection strategy in plant breeding. *Next Gener. Plant Breed.* 26:93. doi: 10.5772/intechopen.76247
- Shi, J., Ma, X., Zhang, J., Zhou, Y., Liu, M., Huang, L., et al. (2019). Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nat. Commun.* 10:464. doi: 10.1038/s41467-018-07876-6
- Silva, R. O., Barioni, L. G., Hall, J. A. J., Matsuura, M. F., Albertini, T. Z., Fernandes, F. A., et al. (2016). Increasing beef production could lower greenhouse gas emissions in Brazil if decoupled from deforestation. *Nat. Clim. Change* 6, 493–497. doi: 10.1038/nclimate2916
- Simeão, R., Silva, A., Valle, C., Resende, M. D., and Medeiros, S. (2015). Genetic evaluation and selection index in tetraploid *Brachiaria ruziziensis*. *Plant Breed.* 135, 246–253. doi: 10.1111/pbr.12353
- Simeão-Resende, R. M., Casler, M. D., and Resende, M. D. V. (2013). Selection methods in forage breeding: a quantitative appraisal. *Crop Sci.* 53, 1925–1936. doi: 10.2135/cropsci2013.03.0143
- Simeão-Resende, R. M., Casler, M. D., and Resende, M. D. V. (2014). Genomic selection in forage breeding: accuracy and methods. *Crop Sci.* 54, 143–156. doi: 10.2135/cropsci2013.05.0353
- Simeão-Resende, R. M., Jank, L., Valle, C. B., and Bonato, A. L. V. (2004). Biometrical analysis and selection of tetraploid progenies of *Panicum maximum* using mixed model methods. *Pesq. Agropec. Bras.* 39, 335–341. doi: 10.1590/S0100-204X2004000400006
- Singh, B. P., Singh, H. P., and Obeng, E. (2013). “Elephant grass,” in *Biofuel Crops: Production, Physiology and Genetics*, ed. B. P. Singh (Fort Valley, GA: CAB International), 271–291.
- Stebbins, G. L. (1947). Types of polyploids: their classification and significance. *Adv. Genet.* 1, 403–429. doi: 10.1016/S0065-2660(08)60490-3
- Talukder, S. K., and Saha, M. C. (2017). Toward genomics-based breeding in C3 cool-season perennial grasses. *Front. Plant Sci.* 8:1317. doi: 10.3389/fpls.2017.01317
- Taylor, M. G., and Vasil, I. K. (1987). Analysis of DNA size, content and cell cycle in leaves of napier grass (*Pennisetum purpureum* Schum.). *Theor. Appl. Genet.* 74, 681–686. doi: 10.1007/BF00247541
- Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., et al. (2014). A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* 15:823. doi: 10.1186/1471-2164-15-823
- Valle, C. B., Jank, L., and Resende, R. M. S. (2009). O melhoramento de forrageiras tropicais no Brasil. *Revist. Ceres* 56, 460–472.
- Valle, C. B., and Savidan, Y. H. (1996). “Genetics, cytogenetics, and reproductive biology of *Brachiaria*,” in *Brachiaria: Biology, Agronomy, and Improvement*, eds J. W. Miles, B. L. Maass, and C. B. do Valle (Colombia: Embrapa), 147–163.
- Valle, C. B. D. (2001). “Genetic resources for tropical areas: achievements and perspectives,” in *Proceedings of the 19<sup>o</sup> International Grassland Congress*, (Piracicaba: Fundação de Estudos Agrários Luiz de Queiroz), 477–482.
- Van Raden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Vigna, B. Z. B., Jungmann, L., Francisco, P. M., Zucchi, M. I., Valle, C. B., and Souza, A. P. (2011). Genetic diversity and population structure of the *Brachiaria brizantha* germplasm. *Trop. Plant Biol.* 4, 157–169. doi: 10.1007/s12042-011-9078-1
- Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the genomics era: concepts and misconceptions. *Nat. Rev. Genet.* 9, 255–266. doi: 10.1038/nrg2322
- Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., et al. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2:e41. doi: 10.1371/journal.pgen.0020041
- Visscher, P. M., Yang, J., and Goddard, M. E. (2010). A commentary on ‘Common SNPs explain a large proportion of the heritability for human height’ by Yang et al. (2010). *Twin Res. Hum. Genet.* 13, 517–524. doi: 10.1375/twin.13.6.517
- Vitezica, Z. G., Varona, L., and Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* 195, 1223–1230. doi: 10.1534/genetics.113.155176
- Wang, C., Li, J., Zhou, S., Liu, T., Zhang, X., and Huang, L. (2018). Genome survey sequencing of purple elephant grass (*Pennisetum purpureum* Schum. ‘Zise’) and identification of its SSR markers. *Mol. Breed.* 38:94. doi: 10.1007/s11032-018-0849-3
- Weller, J. I. (2016). *Genomic Selection in Animals*. Hoboken, NJ: John Wiley & Sons.
- Whittaker, J. C., Thompson, R., and Denham, M. C. (2000). Marker assisted selection using ridge regression. *Genet. Res.* 75, 249–252. doi: 10.1017/S0016672399004462
- Worthington, M., Heffelfinger, C., Bernal, D., Quintero, C., Zapata, Y. P., Perez, J. G., et al. (2016). A parthenogenesis gene candidate and evidence for segmental allopolyploidy in apomictic *Brachiaria decumbens*. *Genetics* 203, 1117–1132. doi: 10.1534/genetics.116.190314
- Worthington, M., and Miles, J. W. (2015). “Reciprocal full-sib recurrent selection and tools for accelerating genetic gain in apomictic *Brachiaria*,” in *Molecular Breeding of Forage and Turf*, eds H. Budak and G. Spangenberg (Cham: Springer International Publishing), 19–30. doi: 10.1007/978-3-319-08714-6\_3
- Worthington, M., Perez, J. G., Mussurova, S., Silva-Cordoba, A., Castiblanco, V., Cardoso Arango, J. A., et al. (2020). A new genome allows the identification of genes associated with natural variation in aluminum tolerance in *Brachiaria* grasses. *J. Exp. Bot.* 16:eraa469. doi: 10.1093/jxb/eraa469
- Würschum, T. (2012). Mapping QTL for agronomic traits in breeding populations. *Theor. Appl. Genet.* 125, 201–210. doi: 10.1007/s00122-012-1887-6
- Yan, Q., Wu, F., Xu, P., Sun, Z., Li, J., Gao, L., et al. (2020). The elephant grass (*Cenchrus purpureus*) genome provides insights into anthocyanidin accumulation and fast growth. *Mol. Ecol. Resour.* 21, 526–542. doi: 10.1111/1755-0998.13271
- Zhang, S., Xia, Z., Zhang, W., Li, C., Wang, X., Lu, X., et al. (2020). Chromosome-scale genome assembly provides insights into speciation of allotetraploid and massive biomass accumulation of Elephant Grass (*Pennisetum purpureum* Schum.). *bioRxiv* [Preprint] doi: 10.1101/2020.02.28.970749
- Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284. doi: 10.1038/s41588-018-0041-z
- Zou, C., Li, L., Miki, D., Li, D., Tang, Q., Xiao, L., et al. (2019). The genome of broomcorn millet. *Nat. Commun.* 10:436. doi: 10.1038/s41467-019-08409-5

**Conflict of Interest:** The authors declare that this review was written in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Simeão, Resende, Alves, Pessoa-Filho, Azevedo, Jones, Pereira and Machado. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.