G3 Genes | Genomes | Genetics

GENOMIC PREDICTION

Genomic Selection with Allele Dosage in *Panicum maximum* Jacq.

Letícia A. de C. Lara,* Mateus F. Santos,† Liana Jank,† Lucimara Chiari,† Mariane de M. Vilela,† Rodrigo R. Amadeu,* Jhonathan P. R. dos Santos,* Guilherme da S. Pereira,‡ Zhao-Bang Zeng,‡ and Antonio Augusto F. Garcia*

*Luiz de Queiroz College of Agriculture / University of São Paulo (ESALQ/USP), Piracicaba, SP, Brazil, †Embrapa Beef Cattle, Campo Grande, MS, Brazil, and ‡North Carolina State University (NCSU), Raleigh, NC

ORCID IDs: 0000-0002-0593-5003 (L.A.C.L.); 0000-0001-5324-906X (M.F.S.); 0000-0001-9436-3678 (L.J.); 0000-0003-3030-7532 (L.C.); 0000-0001-5127-4448 (R.R.A.); 0000-0002-7106-8630 (G.S.P.); 0000-0002-3115-1149 (Z.-B.Z.); 0000-0003-0634-3277 (A.A.F.G.)

ABSTRACT Genomic selection is an efficient approach to get shorter breeding cycles in recurrent selection programs and greater genetic gains with selection of superior individuals. Despite advances in genotyping techniques, genetic studies for polyploid species have been limited to a rough approximation of studies in diploid species. The major challenge is to distinguish the different types of heterozygotes present in polyploid populations. In this work, we evaluated different genomic prediction models applied to a recurrent selection population of 530 genotypes of *Panicum maximum*, an autotetraploid forage grass. We also investigated the effect of the allele dosage in the prediction, *i.e.*, considering tetraploid (GS-TD) or diploid (GS-DD) allele dosage. A longitudinal linear mixed model was fitted for each one of the six phenotypic traits, considering different covariance matrices for genetic and residual effects. A total of 41,424 genotyping-by-sequencing markers were obtained using 96-plex and *Pst*1 restriction enzyme, and quantitative genotype calling was performed. Six predictive models were generalized to tetraploid species and predictive ability was estimated by a replicated fivefold cross-validation process. GS-TD and GS-DD models were performed considering 1,223 informative markers. Overall, GS-TD data yielded higher predictive abilities than with GS-DD data. However, different predictive models had similar predictive ability performance. In this work, we provide bioinformatic and modeling guidelines to consider tetraploid dosage and observed that genomic selection may lead to additional gains in recurrent selection program of *P. maximum*.

KEYWORDS
Plant Breeding
Guinea Grass
Quantitative Genotyping
Polyploidy
Genotyping-by-sequencing (GBS)
Recurrent Genomic Selection
Genomic Prediction
GenPred
Shared Data Resources

Many agricultural and forage crops of economic importance are autotetraploids, such as potato (*Solanum tuberosum*) (Allard 1960), alfalfa (*Medicago sativa*) (McCoy and Bingham 1988), and guinea grass (*Panicum maximum* Jacq.) (Warmke 1954; Jank *et al.* 2014). For these species, marker alleles can be represented with different dosages ranging from 0 to 4. This allele dosage refers to the number of copies of the reference allele, ranging from *aaaa* for nulliplex, *Aaaa* for simplex (single dose), and so on up to *AAAA* (quadruplex), where *A* is the reference allele, for a biallelic marker, such as Single Nucleotide Polymorphisms (SNPs).

Historically, genetic analysis in polyploids has been based on single dose markers, mostly for building genetic maps (Wu *et al.* 1992; Hackett and Luo 2003). With the development of Next-Generation Sequencing (NGS) technologies and the advance of genetic and statistical methods, new possibilities have arisen to enable studies of the complex polyploid genomes for many crops (Garcia *et al.* 2013). In these species, the evaluation of SNP throughout the genome allows one to assess the

doi: https://doi.org/10.1534/g3.118.200986
Manuscript received December 19, 2018; accepted for publication May 23, 2019; published Early Online June 6, 2019.

Supplemental material available at FigShare: https://doi.org/10.25387/g3.7762958.
[1]Corresponding authors: Department of Genetics, Luiz de Queiroz College of Agriculture / University of São Paulo (ESALQ/USP), Piracicaba, SP, Brazil. E-mail: augusto.garcia@usp.br. Embrapa Beef Cattle, Campo Grande, MS, Brazil. E-mail: mateus.santos@embrapa.br

relative abundance of each allele (Voorrips *et al.* 2011; Serang *et al.* 2012; Garcia *et al.* 2013; Hackett *et al.* 2013; Mollinari and Serang 2013; Schmitz Carley *et al.* 2017; Gerard *et al.* 2018).

According to Osborn *et al.* (2003), the allele dosage is expected to be related to different expression levels of some target traits. Thus, polyploidy can increase the potential variation in its genic expression, reflecting in phenotypic variation. Therefore, the inclusion of allele dosage information has become important for genetic studies in polyploid species (Garcia *et al.* 2013; Hackett *et al.* 2013; Endelman *et al.* 2018). In addition, it will allow the development of new statistical models in autopolyploid species, such as in the forage *P. maximum*.

Panicum maximum Jacq. (Syn. *Megathyrsus maximum* Jacq. B. K. Simon & S. W. L. Jacobs) is a perennial tropical autotetraploid forage grass ($2n = 4x = 32$) that reproduces by facultative apomixis (Warmke 1954), which means a high percentage of asexual reproduction by seeds. Cross-breeding started with the doubling of chromosomes by colchicine of rare diploid sexual plants ($2n = 2x = 16$) found in the center of origin in East Africa (Pernès 1975). This enabled the successful crossing between apomictic and sexual plants on a tetraploid level, giving rise to a progeny of sexual and apomictic hybrids in a 1:1 ratio (Savidan 1978). This means that it is possible to fix the hybrid vigor by apomixis in the F_1 generation. Also, superior sexual hybrids can be used as new parents in sexual × apomictic crosses or as donors of favorable alleles in sexual populations.

Few breeding generations separate the founder diploid accessions from the current sexual parents, and increase in the frequency of favorable alleles by selection for traits such as disease resistance, regrowth, biomass and seasonal yield, and nutritive value is necessary. Currently, conventional methods such as half-sib recurrent selection (RS) scheme (Hallauer and Carena 2012) have been carried out in sexual populations to increase the general combining ability of sexual parents. However, RS is a long term program as each cycle requires from three to five years for developing, evaluating, selecting and recombining superior plants. The evaluation phase needs two years for phenotyping the target traits in different harvests and seasons (rainy and dry seasons), and is the most time and expensive phase of the RS program. Thus, methods that accelerate the RS cycles will speed up the development of superior sexual parents for crossing with apomictic plants and increase the probability of releasing improved apomictic cultivars in the market. This is one of the goals of EMBRAPA, the Brazilian Agricultural Research Corporation. The unit Embrapa Beef Cattle is responsible for forage breeding programs, where the main crops are *Brachiaria* spp. and *P. maximum* (Jank *et al.* 2011). This research was established into the *P. maximum* recurrent genomic selection program.

Initially proposed by Meuwissen *et al.* (2001), genomic selection (GS) is an approach that uses statistical methods to predict breeding values from markers distributed throughout the genome. It uses all markers simultaneously to calculate the genomic estimated breeding values (GEBVs) of individuals for ranking and progeny selection (Meuwissen *et al.* 2001; Bernardo 2014). GS has recently been proposed as a useful tool for rapid genetic gains in RS (Massman *et al.* 2013; Müller *et al.* 2017; Zhang *et al.* 2017). It is a promising approach to accelerate cycles of recurrent selection since an established prediction equation can be used repeatedly for multiple cycles of selection (Müller *et al.* 2017). Furthermore, GS promises to increase genetic gain per unit of time, reduce costs for phenotyping, and increase the accuracy of selection.

Most of the available GS models were developed for diploids, and are not well adapted and evaluated for polyploid species. The huge challenge in working with polyploid species is in the correct distinction among different types of heterozygotes. Hidden heterozygotes lead to pseudo-diploid models, that are commonly applied in polyploid species (Annicchiarico *et al.* 2015; Biazzi *et al.* 2017). However, the impact of this approximation for GS models has only been reported in a few studies (Endelman *et al.* 2018; Oliveira *et al.* 2019).

The goal of this research was to develop and to evaluate predictive models for genomic selection in *P. maximum*. For this purpose, we compared different predictive models considering tetraploid and diploid allele dosage to obtain the GEBVs. The genotype calling was performed to allow the identification of different dose levels, showing all heterozygous genotypes and six different GS models were extended to incorporate the tetraploid allele dosage. To our knowledge, this is the first study that includes tetraploid dosage in whole-genome regression models for genomic selection in this tropical forage grass, combined with a high throughput genotyping in a sexual recurrent selection population of *P. maximum*.

MATERIALS AND METHODS

Panicum maximum population

We generated *P. maximum* multi-parental population using 20 selected plants (JA, S7, S13, S16, A42, B87, T103, T4610, A47, A72, B107, C48, C16, B22, Y34, C54, B74, B96, BX4, and B103) as male parents and 19 sexual plants (the same set of plants except B107) as female parents in a polycross mating design. These parents were selected based on relevant agronomic performance in the forage breeding program at Embrapa Beef Cattle during the last 30 years. After the crossing process, we synthesized 19 half-sibs progeny families, each family composed of 30 individuals, resulting in a total of 570 tetraploid sexual plants.

Phenotypic evaluation

The multi-parental population was evaluated in a split-plot randomized complete block design with six replications at Embrapa Beef Cattle, in Campo Grande city, Mato Grosso do Sul state, Brazil (20°27'S, 54°37'W, 530m). Half-sib progenies (represented by each female parent) were treated as the whole plot factor, while individuals were treated as subplot factors. Experimental units were 5 *m* by 2 *m* sized plots containing 5 individual plants of a family or 5 clones of one of three apomictic cultivars (B107, Mombaça, and Tanzânia). Thus, each block consisted of a total of 22 (19 half-sib families and three check cultivars) and a total of 110 individual plants (Figure S1). Therefore 660 plants were evaluated in the experiment, in which 570 were individuals from the 19 half-sib progenies and 90 were clones from the apomictic cultivars.

Each individual plant was evaluated for six traits: leaf dry matter (LDM - g/plant), regrowth capacity (RC), percentage of leaf blade (PLB - %), organic matter (OM - %), crude protein (CP - %), and *in vitro* digestibility of organic matter (IVD - %). The RC trait was evaluated seven days after the harvest using density notes and speed notes of sprouted stems. We considered these traits due their agronomic importance for forage breeding, especially for better forage quality properties for animal consumption. Thus, forage value is measured indirectly by being converted into animal products (such as meat, milk, leather, and furs) (Jank *et al.* 2011). The first three traits were evaluated for eight harvests, during the years of 2013 (four harvests), 2014 (one harvest), and 2015 (three harvests), and the last three were evaluated for four harvests, during the years of 2013 and 2015 (two harvests each).

Statistical analysis of phenotypic data

For each trait, we fitted the following longitudinal linear mixed model to obtain predicted means later used in genomic selection analysis. The model is

$$y_{ijkl} = \mu + h_l + b_{k(l)} + p_{j(l)} + bp_{kj(l)} + t_{il} + e_{ijkl} \quad (1)$$

where y_{ijkl} was the phenotypic value of the i^{th} plant from the parent j in the block k at the harvest l, μ was the fixed overall mean, h_l was the fixed effect of the l^{th} harvest ($l = 1, \ldots, L$, with $L = 8$ for LDM, RC, and PLB traits, and $L = 4$ for OM, CP, and IVD traits), $b_{k(l)}$ was the random effect of the k^{th} block ($k = 1, \ldots, K$, with $K = 6$) at the harvest l, $p_{j(l)}$ was the random effect of the j^{th} parent ($j = 1, \ldots, J$, with $J = 22$) at the harvest l, $bp_{kj(l)}$ was the random interaction effect of block k and parent j at the harvest l, t_{il} was the effect of the i^{th} plant ($i = 1, \ldots, I$, with $I = 573$) at the harvest l, and e_{ijkl} was the random environmental error. The plant effects (t_{il}) were separated into two groups, in which g_{il} was the random effect of the i^{th} individual genotype ($i = 1, \ldots, I_g$, with $I_g = 570$) at the harvest l, and c_{il} was the fixed effect of the i^{th} apomictic cultivar ($i = I_g + 1, \ldots, I_g + I_c$, with $I_c = 3$) at the harvest l. For genotypes, the vector $\boldsymbol{g} = (g_{11}, \ldots, g_{I_gL})'$ was assumed to follow a multivariate normal distribution with zero mean and genetic VCOV matrix $\boldsymbol{G} = \boldsymbol{G}_L \otimes \boldsymbol{I}_{I_g}$, *i.e.*, $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{G})$. For residuals, $\boldsymbol{e} \sim N(0, \boldsymbol{R})$, where $\boldsymbol{e} = (e_{1111}, \ldots, e_{IJKL})'$ and $\boldsymbol{R} = \boldsymbol{R}_L \otimes \boldsymbol{I}_{I \cdot J \cdot K}$. For parents, $p_{j(kl)} \sim N(0, \sigma_p^2)$ and for blocks, $b_{k(l)} \sim N(0, \sigma_b^2)$.

The choice of the genetic ($\boldsymbol{G}_L$) and residual ($\boldsymbol{R}_L$) VCOV matrices was performed hierarchically as follows. First, we picked the model with the best fit for the VCOV structures of the genetic effects among nine different VCOV structures (Table 1). This was done assuming an identity matrix (ID) as the VCOV matrix of the residual effects. Then, this picked model was evaluated assuming nine different VCOV structures for residual effects. Finally, the model with the best fit for the VCOV structure of the residual effect was selected. The models' goodness of fitness were computed with Akaike Information Criterion (AIC) (Akaike 1974) and Bayesian Information Criterion (BIC) (Schwarz 1978). These analyses were performed in the R package ASReml-R (Butler *et al.* 2009). The generalized heritability (broad-sense heritability) for each trait was calculated using the same model as before mentioned but considering $\boldsymbol{G}_L$ matrix as CS and $\boldsymbol{R}_L$ as ID, according to Cullis *et al.* (2006) as:

$$\hat{H}_C^2 = 1 - \frac{PEV}{2\sigma_G^2} \quad (2)$$

where, *PEV* is the prediction error variance (average variance of individual plant comparisons) and σ_G^2 is the genetic variance.

Table 1 Variance and covariance structures examined for genetic (G_L matrix) and residual effects (R_L matrix)

Model	n_{PAR}[a]	Description
ID	1	Identical variation
DIAG	*L*	Heterogeneous variation
CS	2	Compound symmetry with homogeneous variation
CS_{Het}	*L* + 1	Compound symmetry with heterogeneous variation
AR1	2	First-order autoregressive model with homogeneous variation
$AR1_{Het}$	*L* + 1	First-order autoregressive model with heterogeneous variation
Po	2	Power model with homogeneous variation
Po_{Het}	*L* + 1	Power model with heterogeneous variation
US	*L*(*L*+1)/2	Unstructured model

[a]The number of parameters for the models, where *L* is the number of harvests.

Molecular data

Due to losses of individuals in the field, we sequenced a total of 530 offspring. DNA of these 530 offspring were extracted using the DNeasy Plant kit (QIAGEN) and sequenced along with female parents each repeated twice. To provide a higher sequence depth, genotyping-by-sequencing (GBS) was conducted in NextSeq 500 platform for 96-plex *Pst*1 libraries and following the protocol from Genomic Diversity Facility, Cornell University (Elshire *et al.* 2011). Raw data were analyzed using Tassel-GBS pipeline (Glaubitz *et al.* 2014) modified to obtain the exact read counts for each SNP allele (Pereira *et al.* 2018). As this pipeline requires a reference genome and *P. maximum* does not have one, we aligned the GBS tags against six pseudo-references from other diploid and tetraploid forage genomes, and *P. maximum* transcriptomes: (i) *Panicum hallii* genome (v. 2.0, DOE-JGI; ~554 Mb arranged in 9 chromosomes and 8,405 scaffolds; diploid forage); (ii) *Panicum virgatum* genome (v 1.0, DOE-JGI; ~1,230 Mb arranged in total of 18 chromosomes, 9 chromosomes named as A and B each, and 220,646 contigs; allotetraploid forage); (iii) *Setaria italica* genome (v 2.2; ~405.7 Mb arranged in 336 scaffolds; diploid forage) (Bennetzen *et al.* 2012); (iv) *Setaria viridis* genome (v 1.0, DOE-JGI; ~394.9 Mb arranged in 9 chromosomes and 724 scaffolds; diploid forage); (v) transcriptome of *P. maximum* obtained by EMBRAPA institution (43,803 transcripts with mean length of 841.334 bases); and (vi) transcriptome of *P. maximum* obtained by UNICAMP institution (138,853 transcripts with mean length of 695.658 bases). All the genomes are available in Phytozome website (http://www.phytozome.jgi.doe.gov/) (Goodstein *et al.* 2011). The transcriptomes were provided directly by the mentioned institutions. The transcriptome obtained by UNICAMP was published by Toledo-Silva *et al.* (2013). The transcriptome obtained by EMBRAPA has not yet been published. Bowtie2 algorithm (Langmead and Salzberg 2012) was used to align tags against each reference with -D and -R parameters defined as 20 and 4, respectively, and with very-sensitive-local argument. Common reads that aligned to the different genomes were called as different SNP markers. Then, we performed a preliminary analysis evaluating the prediction accuracy of models containing all the markers against models containing only non-redundant markers in genomic prediction models. As predictive accuracies between them did not differ, we chose to use all the markers in the subsequent analyses.

In Tassel-GBS pipeline, the minimum minor allele frequency (mnMAF) considered was 1%. The counting information derived from the Tassel-GBS pipeline was used to estimate the tetraploid allele dosage for each loci using SuperMASSA (Serang *et al.* 2012; Pereira *et al.* 2018). This software is designed especially for SNP calling in polyploid species based on probabilistic graphical models. In SuperMASSA software, the minimum overall depth considered was 25 reads and the model used was "Generalized Population Model". Markers were fitted and filtered to ploidy 4. Triallelic SNPs and markers with more than 5% of missing data were filtered out. Since the selection of markers with up to 5% of missing data are a very strict filter, it is expected to include few percentage of errors due to the imputation process. The imputation was made using random sampling considering the frequency of each dose within each marker. Redundancy among markers was analyzed and displayed using R package circlize (Gu *et al.* 2014). The name of the markers was formed by: reference genome plus chromosome number plus position of SNP in the chromosome.

As linkage disequilibrium (LD) can affect the prediction accuracy of GS, the LD was estimated using squared Pearson correlation, r^2 (Vos *et al.* 2017). Correlations were calculated on tetraploid dosage (0, 1, 2, 3, and 4) between marker pairs with up 1500 bp of distance

for each reference genome. The average r^2 between adjacent markers were calculated and, subsequently, the pairwise correlations were pooled over all chromosomes for each reference genome. A principal component analysis (PCA) was performed on the genotype data to detect population structure and displayed using ggplot2 (Wickham 2009). The relatedness among individuals was assessed computing (i) a pedigree-based additive relationship matrix (Amadeu *et al.* 2016), and (ii) a marker-based genomic relationship matrix (VanRaden 2008) using the R package AGHmatrix (Amadeu *et al.* 2016).

Genomic prediction models using tetraploid dosage

Here, we generalized well known GS models for diploid to tetraploid species using the information of tetraploid allele dosage. Our response variable was the marginal predicted values of individual plants across harvests (BLUPs considering all harvests simultaneously). We evaluated both bayesian and frequentist approaches: Bayesian Ridge Regression (BRR) (Whittaker *et al.* 2000; Meuwissen *et al.* 2001); Bayes A (BA) (Meuwissen *et al.* 2001); Bayes B (BB) (Meuwissen *et al.* 2001); Bayes C (BC) (Habier *et al.* 2011); Bayesian LASSO (BL) (Park and Casella 2008); and Genomic Best Linear Unbiased Predictor (GBLUP) (VanRaden 2008).

All bayesian models expanded for tetraploid allele dosage (TD models) share the same predictive multiple linear regression model (a complete description of these models can be seen at Pérez and de los Campos (2014) and de los Campos *et al.* (2013)),

$$\boldsymbol{y} = 1_n\mu + \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{3}$$

where, $\boldsymbol{y}$ $(n\times 1)$ is the n adjusted entry mean response vector (mean values across harvests obtained from the phenotypic analysis); $\mathbf{1}_n$ is a vector of 1's; μ is a scalar representing the population mean; $\boldsymbol{X}$ $(n\times p)$ is the tetraploid allele dosage incidence matrix of p marker loci coded as $x_{ij}\subset\{0,1,2,3,4\}$ according to the copy number of the reference allele for individual i at marker j; $\boldsymbol{\beta}$ $(p\times 1)$ is the vector of (unknown) marker with tetraploid dosage genetic (TDG) effects; and $\boldsymbol{\varepsilon}$ $(n\times 1)$ is the vector of residual effects, $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}_n, \boldsymbol{I}\sigma^2_\varepsilon)$.

Different assumptions of TDG effects were evaluated. BRR-TD model assumes that all marker loci share the same normal prior distribution, $\beta_j\left|\sigma^2_\beta \right. \sim N(0,\sigma^2_\beta)$, where the common genetic variance hyperparameter (σ^2_β) follows a scaled inverse chi-squared hyperprior distribution $\sigma^2_\beta|d.f._\beta, S_\beta \sim \chi^{-2}(d.f._\beta, S_\beta)$, where $d.f._\beta$ is the number of degrees of freedom and S_β is the scale parameter of the distribution.

BA-TD model is an extension of the above model, which assumes that each TDG prior effects follows specific normal densities, $\beta_j\left|\sigma^2_{\beta_j} \right. \sim N(0,\sigma^2_{\beta_j})$. As before, each specific genetic variance hyperparameter $(\sigma^2_{\beta_j})$ follows a scaled inverse chi-squared distribution. Due to its property, we expect that BA-TD model tends to shrink TDG effects with different prior strength (desirable for highly parameterized models, $p \gg n$), as opposed to the BRR-TD that assumes a common genetic variance hyperparameter. Genetically, these assumptions mean that the analyzed traits are controlled by many genes of small effects and few genes of large effects.

BB-TD model is an extension of BA-TD model, that takes into account the TDG effects prior as a mixture of two normal densities, $\beta_j|\delta,\sigma^2_{\beta_j},\pi \sim \pi\ N(0,\delta) + (1-\pi)\ N(0,\sigma^2_{\beta_j})$, where δ is the genetic variance hyperparameter assumed as a known infinitesimal small value. In this model, we can interpret the mixture proportion (π) as a known expectation of a Bernoulli random variable, that is, the expectation of which mixture component best describes the TDG effects (Dos Santos *et al.* 2016).

BC-TD model is a parsimonious variant of the BB-TD model, which considers that all TDG effects follows a common mixture of two normal distributions, $\beta_j|\delta,\sigma^2_\beta,\pi \sim \pi\ N(0,\delta) + (1-\pi)\ N(0,\sigma^2_\beta)$.

BL-TD model assumes that all markers follows specific normal priors, with genetic variance hyperprior given by the product $\sigma^2_\varepsilon\sigma^2_{\beta_j}$. However, the key difference of BL-TD is the assumption that one component of genetic variance hyperparameters follows exponential distributions, $\sigma^2_{\beta_j}\left|\lambda_\beta \right. \sim Exp(\lambda_\beta)$. The hyperparameter λ_β measures the knowledge (precision) about the genetic variance hyperparameter. As the usual procedure for all the above models, we assume the residual genetic variance hyperparameter follows $\sigma^2_\varepsilon\left|d.f._\varepsilon, S_\varepsilon \right. \sim \chi^{-2}(d.f._\varepsilon, S_\varepsilon)$.

The frequentist model GBLUP-TD was:

$$\boldsymbol{y} = 1_n\mu + \boldsymbol{Zg} + \boldsymbol{\varepsilon} \tag{4}$$

where $\boldsymbol{y}$, 1_n, μ, and ε_{ij} are the same as previously defined; $\boldsymbol{Z}$ $(n\times n)$ is an identity indicence matrix, and $\boldsymbol{g}$ $(n\times 1)$ is the vector mapping the individuals total dosage genetic effects (random effect). The GBLUP-TD model assumes that the random variable $\boldsymbol{g}$ follows a multivariate normal distribution, $\boldsymbol{g} \sim MVN(\mathbf{0}_n, \boldsymbol{K}^*\sigma^2_g)$, where σ^2_g represents the genetic variance of the population and $\boldsymbol{K}^* = 0.99\boldsymbol{K} + 0.01\boldsymbol{A}$, where $\boldsymbol{K}$ is the genomic relationship matrix and $\boldsymbol{A}$ the additive relationship matrix (VanRaden 2008; Aguilar *et al.* 2010; Isik *et al.* 2017). This normalization is necessary to obtain a invertible genomic relationship matrix (Isik *et al.* 2017) and the use of such weight does not result in critical differences in the outputs (VanRaden 2008; Aguilar *et al.* 2010). The $\boldsymbol{K}$ matrix, for tetraploid organisms, can be expressed as $\boldsymbol{K} = \frac{\boldsymbol{WW}^T}{tr(\boldsymbol{WW}^T)/n}$, were $\boldsymbol{W}$ is the centered marker score matrix. Each $k_{i'i}$ on $\boldsymbol{K}$ can be interpreted as a correlation between genotypes of different individuals (genomic relationship), and each k_{ii} as the correlation of the genotypes of an individual with itself (inbreeding). $\boldsymbol{K}$ and $\boldsymbol{A}$ matrices were computated using R package AGHmatrix (Amadeu *et al.* 2016).

To fit all bayesian models, we used the R package BGLR (Pérez and de los Campos 2014), choosing the default package settings for all known hyperparameters. To obtain the posterior distribution of the unknown parameters and hyperparameters, we used the Gibbs sampler with 20,000 iterations; the first 2,000 cycles were discarded as burn in. The GBLUP-TD model was analyzed using the R package ASReml-R (Butler *et al.* 2009).

Model evaluation

The best statistical model was selected for each of the phenotypic traits. Cross-validation with fivefolds has been repeated 100 times for bayesian approaches (computationally intensive) and 1,000 times for the frequentist approach, to obtain an asymptotic empirical distribution of the predictive ability. In each replication, the population was randomly split into five disjoint subsets of genotypes. Whereas one subset was used as validation population (20% or 106 individuals), the remaining four were combined as training population (80% or 424 individuals) to predicte the left-out genotypes in the first population. Subsequently, another subset was used as validation population and the left-out genotypes of this set were predicted. These steps were repeated until all five subsets were used as validation population once.

We calculated Pearson correlation between observed $(\boldsymbol{y})$ and predicted $(\hat{\boldsymbol{y}})$ adjusted entry means for the validation sets, considering, simultaneously, all five cross-validations of each replication. Predictive ability was calculated as the mean of these correlations. In addition, we also derived the empirical distribution of the predicted residual error sums of squares (PRESS), given by the sum of squares of the difference between the predicted and observed adjusted entry means. The narrow-sense heritability was calculated as:

$$\hat{h}^2 = \frac{\sigma^2_A}{\sigma^2_P} \tag{5}$$

where, σ_A^2 is the additive genetic variance and σ_P^2 is the phenotypic variance from GBLUP analysis. Therefore, this estimate corresponds to a genomic heritability, using the additive relationship matrix from molecular markers.

Genomic prediction models using diploid dosage

We also performed a comparison between GS models using diploid and tetraploid allele dosage (GS-DD and GS-TD models, respectively). To do so, the three possible heterozygotes (*Aaaa*, *AAaa*, and *AAAa*) were coded as a single heterozygote (*Aa*), while the two tetraploid homozygotes (*AAAA* and *aaaa*) as diploid homozygotes (*AA* and *aa*). Molecular data were filtered to eliminate markers containing only doses 0 and 1. This step was necessary since the offspring were closely related among them and most of the genotypes in the full data were classified as *aaaa* or *Aaaa* (84.27% and 12.57%, respectively). Therefore, the molecular data set decreased from 41,424 to 1,223 markers. In these new data, tetraploid molecular matrix contains 1,223 markers coded as 0, 1, 2, 3, and 4, and diploid molecular matrix contains the same 1,223 markers coded as 0, 1, and 2. Usual GBLUP model with diploid dosage (GBLUP-DD) was evaluated using R package ASReml-R (Butler *et al.* 2009), considering the genomics relationship matrix as described by VanRaden (2008). Bayesian models with diploid dosage were evaluated using R package BGLR (Pérez and de los Campos 2014). The model evaluation was performed as previously described.

Data Availability

Phenotypic and molecular data and R code files can be found at: <https://github.com/leticia-lara/PM_data>. GBS data have been submitted to the NCBI Sequence Read Archive with the BioProject ID: PR-JNA510446. Additional information can also be found in the supplementary material. Figure S1 contains an illustrative scheme of the experimental design. Figure S2 contains the proportion of missing data in molecular dataset. Figure S3 contains the heatmaps of relationship matrices. Figure S4 contains the principal component analysis. Figure S5 contains the linkage disequilibrium analysis. Table S1 contains model selection for phenotypic traits. Table S2 contains the alignment rate for each reference genome. Table S3 contains the PRESS of predictive models for each trait. Table S4 contains the mean predictive ability of genomic selection models using tetraploid dosage for all markers. Table S5 contains the comparison between individual plant values based on half-sib family means and the genomic prediction of individuals. Supplemental material available at FigShare: https://doi.org/10.25387/g3.7762958.

RESULTS

Phenotypic models

All VCOV structures selected for $\boldsymbol{G}_L$ and $\boldsymbol{R}_L$ matrices allowed for heterogeneous variation and/or correlations across harvests for genetic and residual effects (Table S1). When the selection by AIC was not in agreement with the selection by BIC, we calculated the differences between these two selected models considering both criteria and selected by the criteria that obtained the higher difference. For example, considering the OM trait and $\boldsymbol{G}_L$ matrix, the Po_{Het} had the lowest AIC value (with AIC criteria of 2774.400 and BIC criteria of 2826.809) and the Po had the lowest BIC value (with AIC criteria of 2782.262 and BIC criteria of 2817.201). The differences between these two selected models were 7.862 for AIC (2782.262 - 2774.400) and 9.608 for BIC (2826.809 - 2817.201). As BIC criteria had the higher difference, this criteria was used and the Po matrix was selected for $\boldsymbol{G}_L$ matrix in OM trait. The most commom selected matrix for genetic effects was AR1_{Het}, which was selected for CP and IVD, and different VCOV matrices were selected for residual effects. Generalized heritabilities were 0.66, 0.42, 0.48, 0.89, 0.85, and 0.43 for OM, CP, IDV, LDM, RC, and PLB, respectively. On average, traits evaluated in four harvests had lower heritability than traits evaluated in eight harvests.

Genotype calling

Approximately 485 million of reads per lane were obtained from short read sequencing, where 81.73% were good barcoded reads. The alignment of 6,596,939 read tags using *P. hallii*, *P. virgatum*, *S. italica*, *S. viridis*, and two transcriptomes of *P. maximum* (Table 2) showed that the overall alignment ranged from 19.05 to 24.24%. Although transcriptomes obtained by UNICAMP had the highest overall alignment rate, transcriptomes obtained by EMBRAPA had the highest unique alignment rate, followed by *S. viridis* and *S. italica* (Table S2). A total of 476,904 markers were obtained as the sum of the final number of markers for each reference genome (Table 2).

Due to the nature of GBS technique, the sequencing coverage of different samples are random. It is possible that the same genomic region was not sequenced for all samples. Furthermore, sequences that have mutation in the restriction site of the enzyme are also not observed (Elshire *et al.* 2011). Therefore, a large amount of missing data are expected. The reference genomes of *P. hallii*, *P. virgatum*, *S. italica*, and *S. viridis* had around 5,000 markers with 87% missing data (Figure S2). Markers with a high proportion of missing values were eliminated in subsequent steps. Besides, most part of the selected markers had 0% missing data after the Tassel-GBS pipeline. *S. italica*, *S. viridis*, and transcriptome obtained by EMBRAPA had more than 15,000 markers with 0% missing data.

VCF files obtained from Tassel-GBS pipeline were used as input in the SuperMASSA software. A total of 78,289 markers was selected with minimum average depth for the population of 25 reads (Table 2). From this, 32,619 markers had more than 100 minimum overall depth. For example, the scatterplot of marker Hallii.1_3461595 (Figure 1) shows the intuition of how SuperMASSA uses count reads of two alleles to classify individuals according to their genotype using a probabilistic graphical model (Serang *et al.* 2012).

High level of missing data can impact the performance of GS models, and precise imputation of missing entries can be complex, especially when considering polyploid genomes. Therefore, markers were selected with up to 5% of missing data, aiming to reduce imputation bias. The final number of markers was 41,424, which were used in GS models (Table 2). Subsequently, imputation was made using random sampling and considering the dose frequency for each marker.

The redundancy among markers was inspected (Figure 2) and a large similarity was verified within three specific groups of reference: *Panicum* genus, *Setaria* genus, and transcriptomes. This result is expected due to phylogenetic proximity of the groups (Aliscioni *et al.* 2003; Bennetzen *et al.* 2012). More than half of the markers identified by the different reference genomes have non-redundant information, and 31,046 markers were classified as unique (74.95%). This may be due to the great genomic variability still persistent in each genome, since they are relatively new in terms of breeding.

Population structure and linkage disequilibrium

A trace of population structure was detected for the additive relationship matrix based on pedigree information (Figure S3A) as expected, since these pedigrees consider only female parental information. However, no clear population structure was detected for the genomic relationship matrix based on molecular markers (Figure S3B). We also confirmed a

Table 2 Overall alignment rate and total number of markers obtained after Tassel-GBS pipeline, allele dosage in SuperMASSA software, and selection of up to 5% of missing data (Filter NA), for each reference genome

Reference Genome	Overall alignment rate	Tassel-GBS	SuperMASSA	Filter NA
Panicum hallii	19.05%	77,105	12,835	6,945
Panicum virgatum	22.66%	84,119	11,230	5,598
Setaria italica	22.04%	92,494	15,047	8,066
Setaria viridis	22.07%	92,591	15,271	8,118
Transcriptome (EMBRAPA)	20.11%	74,049	14,129	7,665
Transcriptome (UNICAMP)	24.24%	56,546	9,777	5,032
Total	–	476,904	78,289	41,424

relevant trace of population structure for female parents by principal component analysis of their genotype data, in which the three first principal components explained, respectively, 20.88, 11.98, and 10.23% of the total variability in these parents (Figure S4A). On the other hand, the principal component analysis for the 19 half-sib families (530 individuals) did not detected the population structure in these population. Their first three principal components explained, respectively, 6.20, 4.91, and 4.03% of the total variability in the breeding population (Figure S4B).

Average linkage disequilibrium (LD) between marker pairs was 0.3063, 0.3321, 0.3155, 0.3165, 0.2401, and 0.2810 for markers identified from *P. hallii*, *P. virgatum*, *S. italica*, *S. viridis*, and transcriptomes of *P. maximum* obtained by EMBRAPA and UNICAMP, respectively (Figure S5). The LD decay was lower for *P. hallii*, *S. italica*, and *S. viridis* than for others.

Genomic prediction using tetraploid dosage

Genomic selection models were evaluated using all markers (41,424 markers) and only non-redundant ones (31,046 markers). The predictive accuracy did not differ between these two data sets (results not shown), since the predictive models deal well with multicollinearity. From now on, only predictive models using all markers will be presented.

Mean values of predictive ability ranged from 0.1610 (BL-TD for LDM) to 0.4229 (BB-TD for OM) (Table S4). LDM showed the lowest accuracies for all analyzed GS-TD models and OM showed the highest ones (Figure 3). No clear difference in predictive ability was observed among models (Figure 3A). The BL-TD model for LDM showed the highest estimates of standardized PRESS (Figure 3B, Table S3).

Comparison between GS-TD and GS-DD models

The mean predictive ability showed clear differences between GS-TD and GS-DD models (Figure 4 and Table 3). Almost all GS-TD models were superior to GS-DD models for most of the analyzed traits. The higest superiority was for LDM, with an average superiority of 50.96% for GS-TD model in relation to GS-DD model. The narrow-sense heritability ranged from 0.11 for LDM to 0.53 to OM. These results were consistent with those obtained for the predictive abilities, which is expected since both were calculated considering the additive effects (Table 3).

DISCUSSION

The aim of this study was to develop predictive models considering allele dosage for autotetraploid species, with applications in *P. maximum*. We extended six different GS models to autotetraploid species and compared the accuracy of predicted breeding values of these models. To the best of our knowledge, this is for the first time where the efficiency of genomic selection models considering the information from hidden heterozygotes in the autotetraploid tropical forage *P. maximum* has been reported. Furthermore, we evaluated strategies for modeling residual effects during phenotypic analysis, and performing quantitative genotype calling in autotetraploid species. This methodology can be applied to other autotetraploid species as well as can be extended to species with higher ploidy levels.

Before the development of GS models considering tetraploid allele dosage (GS-TD models), it is important to perform a phenotypic analysis carefully since the sucess of GS in breeding for quantitative traits largely depends of phenotyping process (Cabrera-Bosquet *et al.* 2012). We performed a two stage approach for genomic selection, where we considered only phenotypic data in the first stage and incorporated molecular data in the second stage. In the first stage, for each trait, we fitted a longitudinal linear mixed model and treated genotypes as random. We recognized there is a discussion about fit genotype as random in the first stage (Schulz-Streeck *et al.* 2013). However, as we have multiple harvests and we are modeling genotype effects nested within harvests, with VCOV structure for genotype values across harvests, the model requires to treat genotype as random. The gain in information by incorporating the correlations among harvests is corroborated in the VCOV structures selection, in which the matrices allowing heterogeneous genetic and residual variation as well as allowing correlation among harvests provided a better fit than other models for most analyzed traits (Table S1). Moreover, there was reported in the literature that the use of BLUPs (Best Linear Unbiased Predictor, and in this case, it refers to the marginal predicted values of individual plants across harvests) does not result in significant differences for selection purposes (Galli *et al.* 2018), and it has been already applied in other GS studies as a simple approach to include correlations among traits/harvests/environments (Asoro *et al.* 2011; Resende *et al.* 2017; Ferrão *et al.* 2017; de Oliveira *et al.* 2018). Furthermore, we calculated the deregressed BLUPs, but the GS models had lower predictive ability than

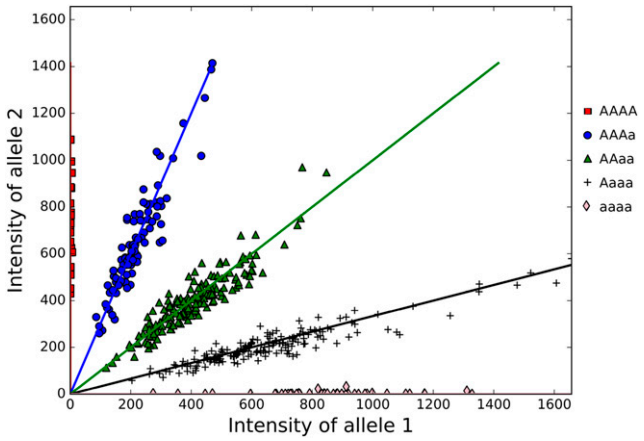

Figure 1 Tetraploid allele dosage for marker Hallii.1_3461595.

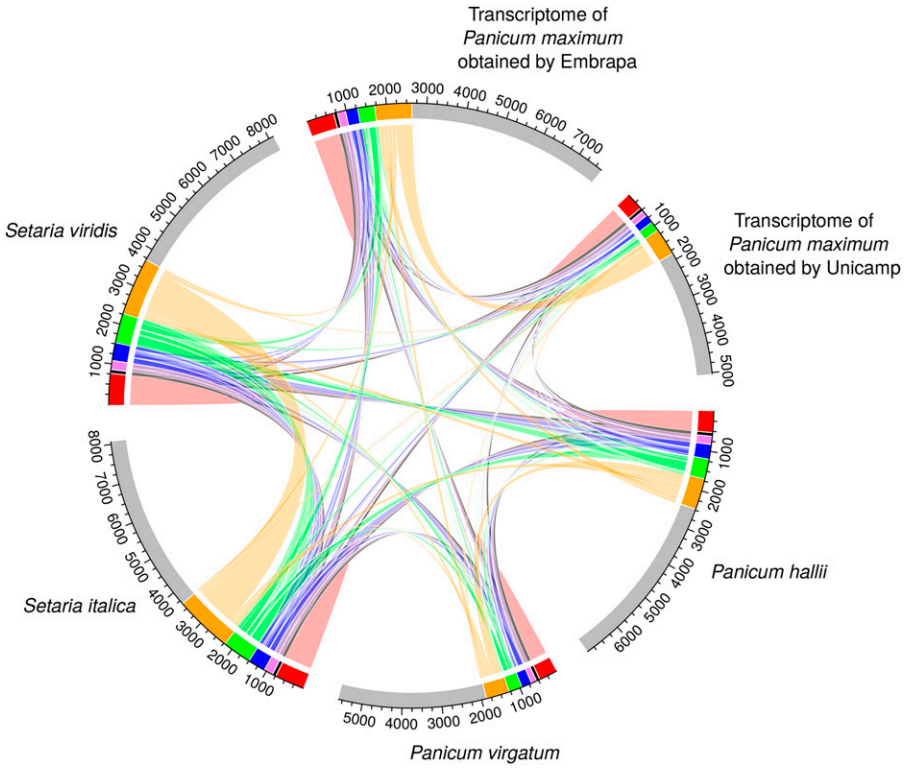

Figure 2 Redundancy among markers. Regions in red represent redundant markers within each reference, while regions in black, pink, blue, green, and orange represent redundant markers among six, five, four, three, and two references, respectively. Gray regions represent markers with unique information for each reference.

using the marginal predicted values of individual plants across harvests (results not shown).

In the second stage, molecular markers were considered in the predictive models. For this, the genotype calling took into consideration the allele dosage, discriminating among the five possible genotypes. According to Uitdewilligen *et al.* (2013), a high sequence depth is required to identify the genotypic class accurately for autotetraploid species, where 60–80 depth leads to 98.4% accuracy in genotypic calls. Reads were also selected with minimum overall depth of 100 reads. However, as the predictive ability of GS-TD models was similar for both criteria (results not shown), we chose to keep only the analysis with an overall depth of 25 reads (approximately 78.7% of markers selected with minimum overall depth of 25 reads were also selected with minimum overall depth of 100 reads). This is a less strict filter for

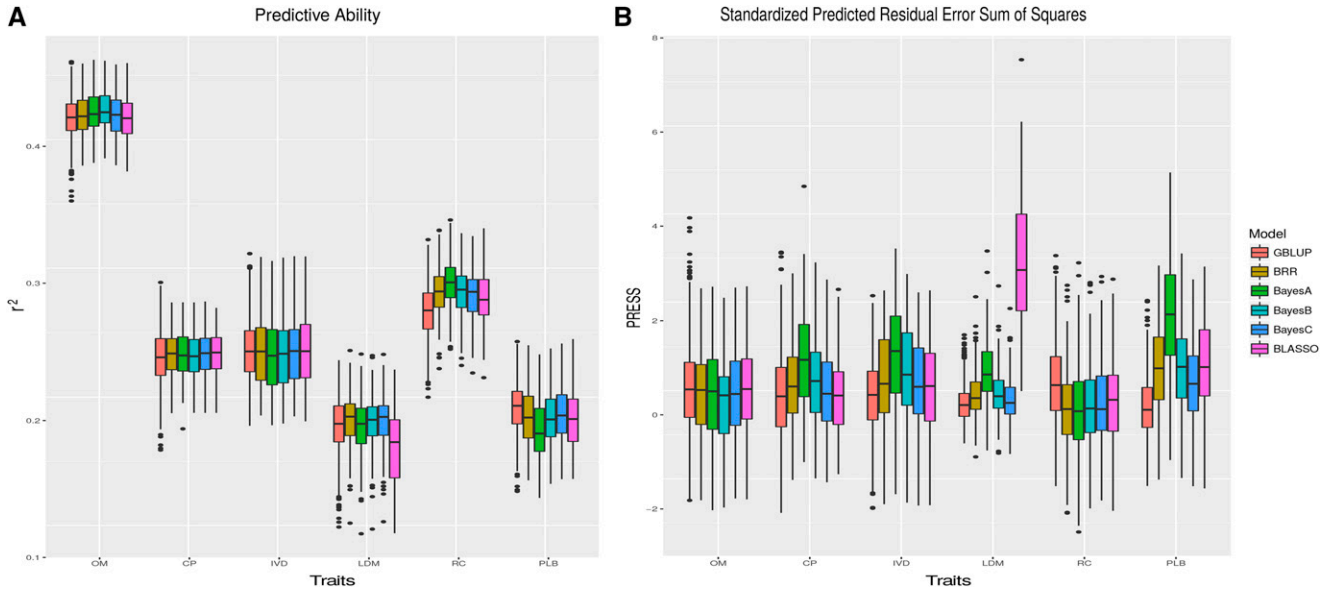

Figure 3 Comparison among six genomic selection models using tetraploid dosage (GS-TD) for organic matter (OM), *in vitro* digestibility of organic matter (IVD), crude protein (CP), leaf dry matter (LDM), regrowth capacity (RC), and percentage of leaf blade (PLB). Molecular data contains 41,424 markers. (A) Predictive ability. (B) Standardized Predicted Residual Error Sum of Squares.

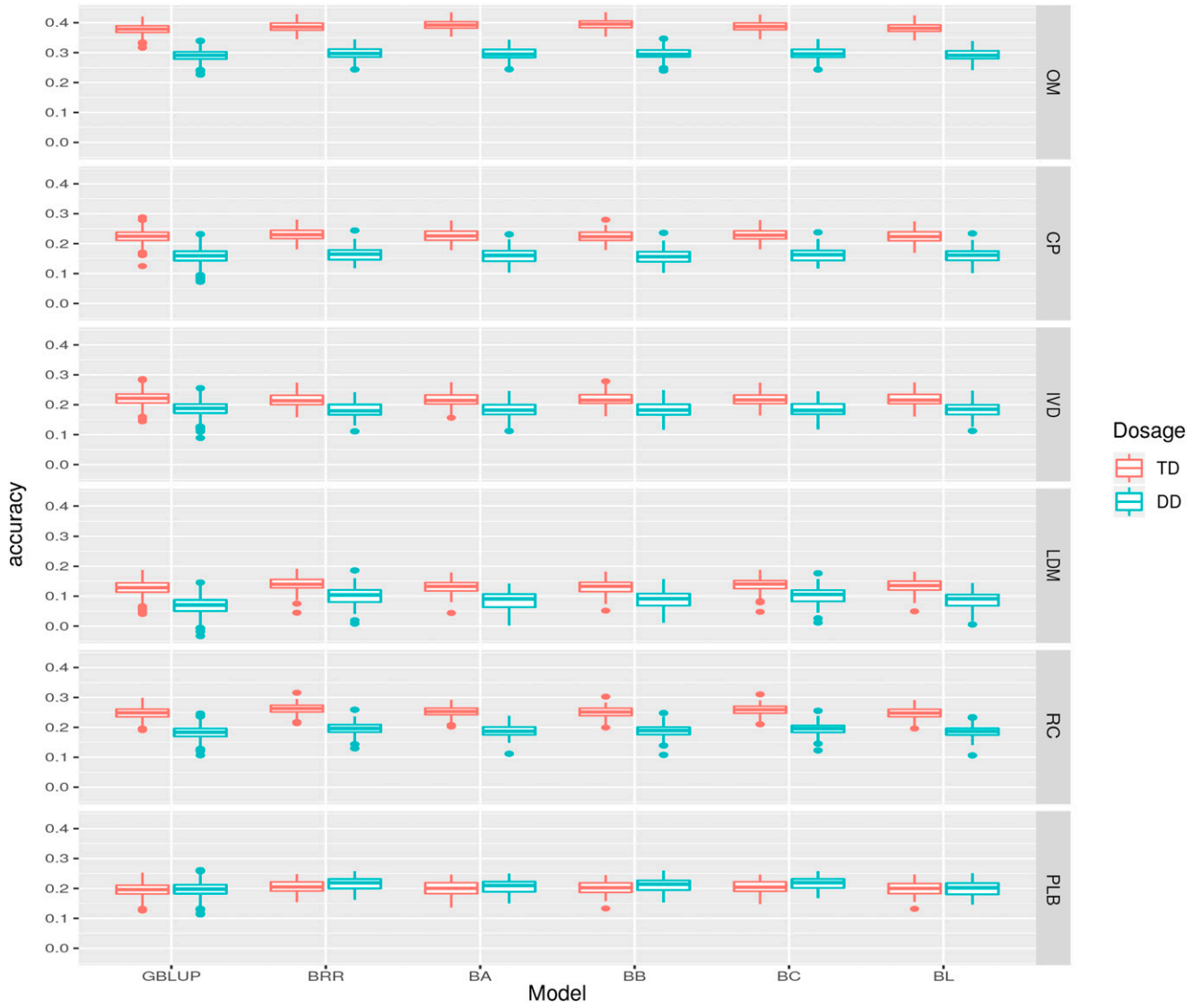

Figure 4 Comparison between genomic selection models using tetraploid dosage (GS-TD) and diploid dosage (GS-DD), for organic matter (OM), crude protein (CP), *in vitro* digestibility of organic matter (IVD), leaf dry matter (LDM), regrowth capacity (RC), and percentage of leaf blade (PLB). Molecular matrices containing 1,223 markers for each data set.

the practical pipeline of breeding and contains a larger number of markers.

The underlying assumption of genomic selection is the presence of SNPs at some loci in linkage disequilibrium (LD) with Quantitative Trait Loci (QTL) alleles that affect traits which are subject to selection (Calus *et al.* 2008). LD represents the non-random association between alleles at different loci and it can be estimated using the correlation among markers when the SNP alleles at those loci are coded with numerical values. Vos *et al.* (2017) calculated several estimators for LD in a simulated and real panel of tetraploid potato and concluded that $LD_{1/2,90}$ values provides the most consistent estimates of LD decay. This estimator consists of 90% percentile r^2 the short-range LD. Short-range LD is calculated across a defined interval of physical distances between marker pairs (Vos *et al.* 2017). One major reason for the minor differences in prediction accuracies among prediction models is the high level of LD found in breeding populations (Riedelsheimer *et al.* 2012). Riedelsheimer *et al.* (2012) obtained similar accuracies of GS models in elite maize germplasm, which had high level of LD. Accuracies did not differ regardless whether the effect of large QTL were precisely captured or spread over a larger region. In this work, we also obtained a high level of LD between markers pairs with up 1500 bp of distance (r^2 from 0.2401 to 0.3321), which can explain the similarity among GS-TD models for prediction purposes.

Population structure is another important factor in GS and can result in biased estimates of the predict ability when it is not taken into account (Riedelsheimer *et al.* 2013). As our population is composed by half-sib families, a stratified sampling was performed for the cross-validation process, in which 20% of individuals in each family were taken from the validation population and 80% from the training population. Therefore, all families were represented in both populations. The GS models and the evaluation process were performed as described in Materials and Methods section. We did not observe differences in the predictive abilities between stratified sampling and random sampling (results not shown) and we are presenting in this paper only results using random sampling. This similarity is expected, since no clear population structure was detected in this population of *P. maximum*

Table 3 Mean predictive ability of genomic selection models using tetraploid dosage (GS-TD) and diploid dosage (GS-DD) for organic matter (OM), crude protein (CP), *in vitro* digestibility of organic matter (IVD), leaf dry matter (LDM), regrowth capacity (RC), and percentage of leaf blade (PLB). GS-TD and GS-DD models with the highest mean predictive ability for each trait are indicated in bold. Molecular matrices containing 1,223 markers for each data set. The broad-sense heritability is $\hat{H}_C^2$ and the narrow-sense heritability is $\hat{h}^2$

Model	OM	CP	IVD	LDM	RC	PLB
GBLUP-TD	0.3782	0.2237	**0.2211**	0.1282	0.2475	0.1957
GBLUP-DD	0.2905	0.1578	0.1869	0.0685	0.1829	0.1968
BRR-TD	0.3866	**0.2299**	0.2157	**0.1390**	**0.2621**	0.2058
BRR-DD	0.2966	0.1633	0.1837	0.0997	0.1951	0.2155
BA-TD	0.3931	0.2251	0.2174	0.1308	0.2514	0.2001
BA-DD	0.2958	0.1585	0.1835	0.0854	0.1876	0.2067
BB-TD	**0.3955**	0.2242	0.2190	0.1311	0.2501	0.2004
BB-DD	0.2960	0.1560	0.1843	0.0886	0.1885	0.2110
BC-TD	0.3879	0.2282	0.2186	0.1374	0.2577	0.2049
BC-DD	0.2958	0.1612	0.1855	0.1006	0.1944	**0.2161**
BL-TD	0.3821	0.2233	0.2187	0.1333	0.2468	0.1991
BL-DD	0.2923	0.1594	0.1849	0.0870	0.1853	0.1998
Average GS-TD	0.3872	0.2257	0.2184	0.1333	0.2526	0.2010
Average GS-DD	0.2945	0.1594	0.1848	0.0883	0.1889	0.2076
Average superiority	31.48	41.59	18.18	50.96	33.72	−3.18
$\hat{H}_C^2$	0.66	0.42	0.48	0.89	0.85	0.43
$\hat{h}^2$	0.53	0.23	0.25	0.11	0.31	0.12

(Figure S3 and S4B). When we observe the genotypes of the parents, few parents are different from the others (Figure S4A). The absence of population structure occurs because most parents are closely related and the individuals analyzed have a high level of relationship among them. Similar results have been reported in wheat (Arruda *et al.* 2015) and in potato (Sverrisdóttir *et al.* 2017), with typical values for data with family structure, but without substantial population structure. As reported by Sverrisdóttir *et al.* (2017), the absence of population structure could in principle be caused by an extremely narrow genetic base of the parents. This is in line with our study since all parents were derived from a common female tetraploidized ancestor.

Synthetic populations are usually formed by crossing several parents and posteriorly cross-pollinating the F_1 individuals for one or several generations (Falconer and Mackay 1996). These populations have played an important role in quantitative genetic research on gene action in complex heterotic traits and comparison of selection methods (Hallauer *et al.* 2010; Schopp *et al.* 2017). As in our population, the number of parents is usually small and parents are often related, leading to small effective size (N_e) (Müller *et al.* 2017). We estimated broad and narrow sense heritabilities for the evaluated traits. Besides we did not expect high heritabilities for traits related with yield, we observed a high broad sense heritability (0.89) and low narrow sense heritability (0.11) for LDM (Table 3). Such low value reflects the difficulty to breed for this trait. Differences in magnitude between these estimates were previously reported by Resende *et al.* (2014b), which observed broad-sense heritability ranging from 0.47 to 0.75 and narrow-sense heritability ranging from 0.15 to 0.24 for LDM considering different years of evaluation. For individual harvests, the same authors found a narrow-sense heritability of 0.05 for LDM. Similar to our results, broad-sense heritabilities were reported in *Paspalum* for LDM with 0.82 and PLB with 0.55 (Pereira *et al.* 2017). In respect to the nutritive traits, moderate values were reported by Matias *et al.* (2016) for narrow-sense heritability considering the mean of seven harvests in *Brachiaria decumbens* progenies for CP (0.31) and IVD (0.14). Similar moderate values for CP and IVD were also reported in our study (Table 3).

The prediction of the breeding values was made using the marginal predicted values of individual plants across harvests that were obtained in the first stage of the analysis. The goals were to select individuals in the present recurrent selection cycle (already phenotypically evaluated) as well as to select non-phenotyped individuals from the next cycle. Since one selection cycle requires three to five years to complete (Resende *et al.* 2014a), *P. maximum* breeding program with genomic selection will reduce from five to one year for each recurrent cycle. This one year is necessary to grow the selected plants and cross them to obtain the new population. In addition, the *P. maximum* breeding programs for releasing cultivars will be benefited since superior sexual plants can be selected every year to, posteriorly, cross with apomictic plants. From these crosses, new apomictic hybrid combinations will be obtained and tested as new cultivars for releasing the best one in the market. Lipka *et al.* (2014) applied genomic selection in *P. virgatum* species with the objective of evaluating genomic selection efficiency to accelerate breeding cycles in this species. Since *P. virgatum* is an allotetraploid species, the authors used diploid dosage and obtained high prediction accuracy for most of the traits, using association panel and considering seven morphological and thirteen biomass quality trais. Although analyzed traits were different from ours, the range of values were similar. The higher mean prediction accuracy obtained by Lipka *et al.* (2014) was 0.52 for standability and the lower was -0.08 for minerals. Our higher mean predictive ability was 0.4229 for organic matter and the lower was 0.1610 for leaf dry matter (Table S4). Similar accuracies were also obtained for oats (Asoro *et al.* 2011), maize (González-Camacho *et al.* 2012), and rice (Spindel *et al.* 2015). Despite the difference among traits, the accuracy did not differ among models. Besides that, we suggest to use GBLUP-TD model for all traits, since it is more intuitive for breeders and it is less computational demanding.

We also compared the predictions of individual plant values based on genomic prediction at the level of half-sib family means with that at the individual level. To achieve this, we removed the maternal effect of the phenotypic model and included it in the genomic selection model. Then, we predicted individuals based only on the maternal effects and compared with prediction of individuals (Table S5). The genomic predicted values are more accurate because they capture not only information on the family means, but also the within-family deviations. The accuracy of prediction using only maternal effect ranged from 0.0452 for PLB to

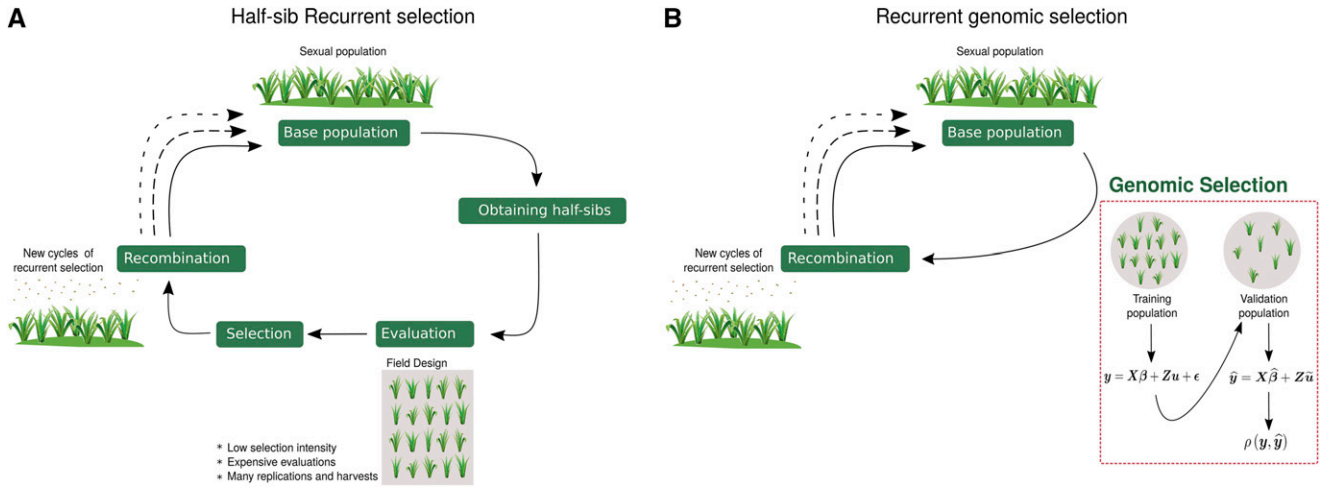

Figure 5 Forage breeding program in a: (A) recurrent phenotypic selection. (B) recurrent genomic selection.

0.1843 for LDM. The prediction of individuals using molecular markers and family effect ranged from 0.1786 for LDM to 0.3466 for OM. Therefore, the GS can operate on individual plants without needing trait data on sibs with considerable accuracy of prediction and genetic gains with selection.

There are few studies on GS for autotetraploid species using allele dosage, most of them are for potato (Slater *et al.* 2016; Endelman *et al.* 2018; Stich and Inghelandt 2018). To our knowledge, this is the first study that applied genomic selection with allele dosage in an autotetraploid tropical forage grass. Slater *et al.* (2016) developed an extension of genomic relationship matrix proposed by Yang *et al.* (2010) for autotetraploids and applied genomic selection in potato. The authors found accuracies ranging from 0.2, under conditions of low heritability and small reference populations, to 0.8 in larger reference populations. Endelman *et al.* (2018) and Stich and Inghelandt (2018) analyzed the inclusion of additive and non-additive components (digenic dominance and two order epistasis) for the accuracy of predictive models. The inclusion of dominance effects in GS when selecting the per-se performance of the individuals is fundamental, mainly when working with species that are highly heterozygous, which allows dominance effects to contribute to phenotypic variation (Stich and Inghelandt 2018). Concordantly, Endelman *et al.* (2018) extended the genomic relationship matrix proposed by VanRaden (2008) for autotetraploids and highlighted the importance of genotypic and additive values for selection. However, for both studies, when the goals is selecting new parents for several cycles, only the additive value needs to be considered since the contribution of digenic dominance diminishes exponentially over the generations (Gallais 2003; Endelman *et al.* 2018), in other words, non-additive effects are less efficiently transmitted to progeny. In our work, we modeled only additive effects since our study aims to select superior sexual parents in a recurrent genomic selection program in the current and next selection cycles.

A comparison of GS-TD models with GS models considering the usual diploid allele dosage (GS-DD models) was also performed. Although being a rougly approximation, this strategy has been applied in several polyploid crops, such as alfafa (Annicchiarico *et al.* 2015; Biazzi *et al.* 2017), and sugarcane (Gouy *et al.* 2013), as GS software for polyploids have only recently emerged. To investigate the impact of using diploidized markers, Endelman *et al.* (2018) also compared diploid and tetraploid dosage. The authors observed that the accuracy of the diploidized model was consistently lower. In our study, the accuracy of GS-DD models were also lower than for GS-TD models for all traits (Table 3 and Figure 4). This reinforces that GS models using tetraploid allele dosage are superior for autotetraploid populations, and extensions can and should be applied to other ploidy levels.

Implementation of genomic selection in forage breeding

Recent research shows the potential of genomic selection to reshape crop breeding programs. In particular, the results obtained here imply that GS has great potential for *P. maximum* breeding programs, especially in a recurrent genomic selection program.

Usually, one cycle of half-sib recurrent selection is split in: i) development of progenies; ii) phenotypic evaluation; and iii) selection and recombination of the best selected individuals to obtain an improved population (Figure 5A). The recurrent genomic selection is a modification to get shorter breeding cycles and greater genetic gains with selection. Therefore, genomic prediction is implemented by genotyping and phenotyping the base population and estimating marker effects to predict hybrid performance in the subsequent recurrent selection cycles and recombine the best individuals based only on the GEBVs (Figure 5B). The persistency of predictive accuracy in GS is fundamental for practical breeding because it determines the number of cycles that can be advanced until it is necessary to retrain the predictive models (Müller *et al.* 2017). This is because in each cycle of recombination and selection, the individuals of breeding population can accumulate genetic diversity and gene frequencies may differ from the training population (Heffner *et al.* 2010; Bassi *et al.* 2016). Therefore the breeder needs to update the GS models by phenotyping the population and re-estimating marker effects each three recurrent selection cycles (Heffner *et al.* 2009) or whenever necessary.

To compare the efficiency of GS approach with phenotypic breeding programs, the breeders' equation can be used. In this equation, expected genetic gains per unit of time is defined as $\Delta_G = (ir\sigma_A)/T$, where i is the selection intensity, r is the selection accuracy, σ_A is the square root of the additive genetic variance, and T is the length of time to complete one breeding selection cycle (Falconer and Mackay 1996). The success of GS approach is determined by its ability to predict phenotypes that were not evaluated as well as by its ability to increase the rate of Δ_G while maintaining affordable costs. Assuming similar selection intensity and similar genetic variance for both methods, greater gains per unit of time can be achieved as long as the reduction in the time of each

breeding cycle by GS compensates for the reduction in selection accuracy (Bassi *et al.* 2016). Furthermore, GS becomes cheaper than phenotypic selection for traits that have a long generation time or are difficult and expensive to evaluate.

Another crucial aspect for implementation of genomic selection in forage breeding programs is the generation of accurate genotypic data. Several softwares have been developed to perform quantitative genotype calling for autotetraploid species, such as R packages fitTetra (Voorrips *et al.* 2011) and ClusterCall (Schmitz Carley *et al.* 2017); for "diploidizing" tetraploid species, such as R package breadarrayMSV (Gidskehaug *et al.* 2010); and species with any ploidy level, such as R packages fitPoly (Voorrips *et al.* 2011), updog (Gerard *et al.* 2018), and software SuperMASSA (Serang *et al.* 2012; Pereira *et al.* 2018). Therefore, the evaluation of allele dosage for polyploid species has become more accessible by several available softwares.

This is the first work of genomic selection in the tropical forage grass *P. maximum* which uses a high throughput genotyping and considers tetraploid allele dosage in Bayesian and frequentist models. GBS and allele dosage showed to be promising strategies for genomic analysis in autotetraploid species. Furthermore, the accuracy of predictive models and the time reduction in the breeding cycles justifies the implementation of genomic selection in *P. maximum* breeding programs.

ACKNOWLEDGMENTS

This work was supported by FAPESP (São Paulo Research Foundation), grants 2015/20659-2 and 2016/01279-7. Additional support was provided by CNPq ("Conselho Nacional do Desenvolvimento Científico e Tecnológico"), EMBRAPA ("Empresa Brasileira de Pesquisa Agropecuária") number 02.14.01.013.00.03.003, UNIPASTO ("Associação para o Fomento à Pesquisa de Melhoramento de Forrageiras"), and CAPES ("Coordenação de Aperfeiçoamento de Pessoal de Nível Superior"). L. A. C. Lara was supported by FAPESP, grants 2015/20659-2 and 2016/01279-7. A. A. F. Garcia was supported by a productivity scholarship from CNPq and CAPES ("Bio Computacional 051/2013"). J. P. R. dos Santos was supported by FAPESP, grants 2017/03625-2 and 2017/25674-5. The authors thank all trainees of Embrapa Beef Cattle that helped on the phenotypic evaluations.

LITERATURE CITED

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta *et al.*, 2010 *Hot topic*: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93: 743–752. https://doi.org/10.3168/jds.2009-2730

Akaike, H., 1974 A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19: 716–723. https://doi.org/10.1109/TAC.1974.1100705

Aliscioni, S. S., L. M. Giussani, F. O. Zuloaga, and E. A. Kellogg, 2003 A molecular phylogeny of *Panicum* (Poaceae: Paniceae): tests of monophyly and phylogenetic placement within the Panicoideae. Am. J. Bot. 90: 796–821. https://doi.org/10.3732/ajb.90.5.796

Allard, R. W., 1960 Inheritance in autotetraploids. In *Principles of Plant Breeding*, edited by J. W. Sons, pp. 385–399, Wiley.

Amadeu, R. R., C. Cellon, J. W. Olmstead, A. A. F. Garcia, M. F. R. Resende *et al.*, 2016 AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. Plant Genome 9: 1–10. https://doi.org/10.3835/plantgenome2016.01.0009

Annicchiarico, P., N. Nazzicari, X. Li, Y. Wei, L. Pecetti *et al.*, 2015 Accuracy of genomic selection for alfalfa biomass yield in different reference populations. BMC Genomics 16: 1020. https://doi.org/10.1186/s12864-015-2212-y

Arruda, M. P., P. J. Brown, A. E. Lipka, A. M. Krill, C. Thurber *et al.*, 2015 Genomic selection for predicting Head Blight Resistance in a Wheat Breeding Program. Plant Genome 8: 1–12. https://doi.org/10.3835/plantgenome2015.01.0003

Asoro, F. G., M. A. Newell, W. D. Beavis, M. P. Scott, and J.-L. Jannink, 2011 Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome 4: 132–144. https://doi.org/10.3835/plantgenome2011.02.0007

Bassi, F. M., A. R. Bentley, G. Charmet, R. Ortiz, and J. Crossa, 2016 Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). Plant Sci. 242: 23–36. https://doi.org/10.1016/j.plantsci.2015.08.021

Bennetzen, J. L., J. Schmutz, H. Wang, R. Percifield, J. Hawkins *et al.*, 2012 Reference genome sequence of the model plant Setaria. Nat. Biotechnol. 30: 555–561. https://doi.org/10.1038/nbt.2196

Bernardo, R., 2014 Genomewide selection when major genes are known. Crop Sci. 54: 68–75. https://doi.org/10.2135/cropsci2013.05.0315

Biazzi, E., N. Nazzicari, L. Pecetti, E. C. Brummer, A. Palmonari *et al.*, 2017 Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. PLoS One 12: e0169234. https://doi.org/10.1371/journal.pone.0169234

Butler, D. G., B. R. Cullis, A. R. Gilmour, and B. J. Gogel, 2009 *ASReml-R reference manual.*

Cabrera-Bosquet, L., J. Crossa, J. von Zitzewitz, M. D. Serret, and J. L. Araus, 2012 High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. J. Integr. Plant Biol. 54: 312–320. https://doi.org/10.1111/j.1744-7909.2012.01116.x

Calus, M. P. L., T. H. E. Meuwissen, A. P. W. Roos, and R. F. Veerkamp, 2008 Accuracy of genomic selection using different methods to define haplotypes. Genetics 178: 553–561. https://doi.org/10.1534/genetics.107.080838

Cullis, B. R., A. B. Smith, and N. E. Coombes, 2006 On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11: 381–393. https://doi.org/10.1198/108571106X154443

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus, 2013 Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193: 327–345. https://doi.org/10.1534/genetics.112.143313

de Oliveira, A. A., M. M. Pastina, R. A. da Costa Parrella, R. W. Noda, M. L. F. Simeone *et al.*, 2018 Genomic prediction applied to high-biomass sorghum for bioenergy production. Mol. Breed. 38: 49. https://doi.org/10.1007/s11032-018-0802-5

Dos Santos, J. P. R., L. P. M. Pires, R. C. C. Vasconcellos, G. S. Pereira, R. G. V. Pinho *et al.*, 2016 Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. BMC Genet. 17: 86. https://doi.org/10.1186/s12863-016-0392-3

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379. https://doi.org/10.1371/journal.pone.0019379

Endelman, J. B., C. A. S. Carley, P. C. Bethke, J. J. Coombs, M. E. Clough *et al.*, 2018 Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. Genetics 209: 77–87. https://doi.org/10.1534/genetics.118.300685

Falconer, D. S., and T. F. C. Mackay, 1996 *Quantitative genetics in maize breeding*, Pearson, London.

Ferrão, L. F. V., R. G. Ferrão, M. A. G. Ferrão, A. F. A. Fonseca, and A. A. F. Garcia, 2017 A mixed model to multiple harvest-location trials applied to genomic prediction in *Coffea canephora*. Tree Genet. Genomes 13: 95. https://doi.org/10.1007/s11295-017-1171-7

Gallais, A., 2003 *Quantitative genetics and breeding methods in autopolyploid plants*, Institut National de la Recherche Agronomique, Paris, France.

Galli, G., D. H. Lyra, F. C. Alves, S. C. Granato, R. Fritsche-Neto *et al.*, 2018 Impact of phenotypic correction method and missing phenotypic data on genomic prediction of maize hybrids. Crop Sci. 58: 1481–1491. https://doi.org/10.2135/cropsci2017.07.0459

Garcia, A. A. F., M. Mollinari, T. G. Marconi, O. R. Serang, R. R. Silva *et al.*, 2013 SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Sci. Rep. 3: 3399. https://doi.org/10.1038/srep03399

Gerard, D., L. F. V. Ferrão, A. A. F. Garcia, and M. Stephens, 2018 Genotyping polyploids from messy sequencing data. Genetics 210: 789–807. https://doi.org/10.1534/genetics.118.301468

Gidskehaug, L., M. Kent, B. J. Hayes, and S. Lien, 2010 Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. Bioinformatics 27: 303–310. https://doi.org/10.1093/bioinformatics/btq673

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9: e90346. https://doi.org/10.1371/journal.pone.0090346

González-Camacho, J. M., G. De los Campos, P. Pérez, D. Gianola, J. E. Cairns *et al.*, 2012 Genome-enabled prediction of genetic values using radial basis function neural networks. Theor. Appl. Genet. 125: 759–771. https://doi.org/10.1007/s00122-012-1868-9

Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes *et al.*, 2011 Phytozome: a comparative platform for green plant genommics. Nucleic Acids Res. 40: D1178–D1186. https://doi.org/10.1093/nar/gkr944

Gouy, M., Y. Rousselle, D. Bastianelli, P. Lecomte, L. Bonnal *et al.*, 2013 Experimental assessment of the accuracy of genomic selection in sugarcane. Theor. Appl. Genet. 126: 2575–2586. https://doi.org/10.1007/s00122-013-2156-z

Gu, Z., L. Gu, R. Eils, M. Schlesner, and B. Brors, 2014 *circlize* implements and enhances circular visualization in R. Bioinformatics 30: 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick, 2011 Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12: 186–197. https://doi.org/10.1186/1471-2105-12-186

Hackett, C. A., and Z. W. Luo, 2003 TetraploidMap: Construction of a linkage map in autotetraploid species. J. Hered. 94: 358–359. https://doi.org/10.1093/jhered/esg066

Hackett, C. A., K. McLean, and G. Bryan, 2013 Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PLoS One 8: e63939. https://doi.org/10.1371/journal.pone.0063939

Hallauer, A. R., and M. J. Carena, 2012 Recurrent selection methods to improve germplasm in maize. Maydica 57: 266–283.

Hallauer, A. R., M. J. Carena, and J. B. Miranda Filho, 2010 *Quantitative genetics in maize breeding*, Springer, New York.

Heffner, E. L., A. Lorenz, J.-L. Jannink, and M. E. Sorrells, 2010 Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50: 1681–1690. https://doi.org/10.2135/cropsci2009.11.0662

Heffner, E. L., M. E. Sorrells, and J.-L. Jannink, 2009 Genomic selection for crop improvement. Crop Sci. 49: 1–12. https://doi.org/10.2135/cropsci2008.08.0512

Isik, F., J. Holland, and C. Maltecca, 2017 *Genetic data analysis for plant and animal breeding*, Springer Nature, New York. https://doi.org/10.1007/978-3-319-55177-7

Jank, L., S. C. Barrios, C. B. Valle, R. M. Simeão, and G. F. Alves, 2014 The value of improved pastures to Brazilian beef production. Crop Pasture Sci. 65: 1132–1137. https://doi.org/10.1071/CP13319

Jank, L., C. B. Valle, and R. M. S. Resende, 2011 Breeding tropical forages. Crop Breed. Appl. Biotechnol. 11: 27–34. https://doi.org/10.1590/S1984-70332011000500005

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359. https://doi.org/10.1038/nmeth.1923

Lipka, A. E., F. Lu, J. H. Cherney, E. S. Buckler, M. D. Casler *et al.*, 2014 Accelerating the switchgrass (*Panicum virgatum* L.) breeding cycle using genomic selection approaches. PLoS One 9: e112227. https://doi.org/10.1371/journal.pone.0112227

Massman, J. M., H.-J. G. Jung, and R. Bernardo, 2013 Genomewide selection *vs.* marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci. 53: 58–66. https://doi.org/10.2135/cropsci2012.02.0112

Matias, F. I., S. C. L. Barrios, C. B. Valle, R. G. Mateus, L. B. Martins *et al.*, 2016 Estimate of genetic parameters in *Brachiaria decumbens* hybrids. Crop Breed. Appl. Biotechnol. 16: 115–122. https://doi.org/10.1590/1984-70332016v16n2a18

McCoy, T. J., and E. T. Bingham, 1988 Cytology and cytogenetics of alfalfa, pp. 737–776 in *Alfalfa and alfalfa improvement*, Vol. 29, edited by Hanson, A. A., D. K. Barnes, and R. R. Hill. American Society of Agronomy, Madison, WI.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Mollinari, M., and O. Serang, 2013 Quantitative SNP genotyping of polyploids with MassARRAY and other platforms, pp. 215–241 in *Methods in Molecular Biology*, edited by Walker, J. M. Springer, New York.

Müller, D., P. Schopp, and A. E. Melchinger, 2017 Persistency of prediction accuracy and genetic gain in synthetic populations under recurrent genomic selection. G3 (Bethesda) 7: 801–811. https://doi.org/10.1534/g3.116.036582

Oliveira, I. d. B., M. F. Resende, L. F. V. Ferrão, R. R. Amadeu, J. B. Endelman *et al.*, 2019 Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. G3 (Bethesda) 9: 1189–1198. https://doi.org/10.1534/g3.119.400059

Osborn, T. C., J. Chris Pires, J. A. Birchler, D. L. Auger, Z. Jeffery Chen *et al.*, 2003 Understanding mechanisms of novel gene expression in polyploids. Trends Genet. 19: 141–147. https://doi.org/10.1016/S0168-9525(03)00015-5

Park, T., and G. Casella, 2008 The Bayesian Lasso. J. Am. Stat. Assoc. 103: 681–686. https://doi.org/10.1198/016214508000000337

Pereira, E. A., M. DallAgnol, K. M. Saraiva, C. Simioni, A. P. S. Leães *et al.*, 2017 Genetic gain in apomictic species of the genus *Paspalum*. Rev. Ceres 64: 60–67. https://doi.org/10.1590/0034-737x201764010009

Pereira, G. S., A. A. F. Garcia, and G. R. A. Margarido, 2018 A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids. BMC Bioinformatics 19: 398. https://doi.org/10.1186/s12859-018-2433-6

Pérez, P., and G. de los Campos, 2014 Genome-wide regression and prediction with the BGLR statistical package. Genetics 198: 483–495. https://doi.org/10.1534/genetics.114.164442

Pernès, J., 1975 *Organisation évolutive d'un groupe agamique: la section des Maximae du genre Panicum*. Ph.D. thesis, Orstom.

Resende, R. M. S., M. D. Casler, and M. D. V. de Resende, 2014a Genomic selection in forage breeding: Accuracy and methods. Crop Sci. 54: 143–156. https://doi.org/10.2135/cropsci2013.05.0353

Resende, R. M. S., L. Jank, C. B. d. Valle, and A. L. V. Bonato, 2014b Biometrical analysis and selection of tetraploid progenies of *Panicum maximum* using mixed model methods. Crop Sci. 54: 143–156.

Resende, R. T., M. D. V. Resende, F. F. Silva, C. F. Azevedo, E. K. Takahashi *et al.*, 2017 Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. Heredity 119: 245–255. https://doi.org/10.1038/hdy.2017.37

Riedelsheimer, C., J. B. Endelman, M. Stange, M. E. Sorrells, J.-L. Jannink *et al.*, 2013 Genomic predictability of interconnected bi-parental maize populations. Genetics 194: 493–503. https://doi.org/10.1534/genetics.113.150227

Riedelsheimer, C., F. Technow, and A. E. Melchinger, 2012 Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. BMC Genomics 13: 452. https://doi.org/10.1186/1471-2164-13-452

Savidan, Y., 1978 Genetic control of facultative apomixis and application in breeding *Panicum maximum*.

Schmitz Carley, C. A., J. J. Coombs, D. S. Douches, P. C. Bethke, J. P. Palta *et al.*, 2017 Automated tetraploid genotype calling by hierarchical clustering. Theor. Appl. Genet. 130: 717–726. https://doi.org/10.1007/s00122-016-2845-5

Schopp, P., D. Müller, F. Technow, and A. E. Melchinger, 2017 Accuracy of genomic prediction in synthetic populations depending on the number of parents, relatedness, and ancestral linkage disequilibrium. Genetics 205: 441–454. https://doi.org/10.1534/genetics.116.193243

Schulz-Streeck, T., J. O. Ogutu, and H.-P. Piepho, 2013 Comparisons of single-stage and two stage approaches to genomic selection. Theor. Appl. Genet. 126: 69–82. https://doi.org/10.1007/s00122-012-1960-1

Schwarz, G., 1978 Estimating a dimension of a model. Ann. Stat. 6: 461–464. https://doi.org/10.1214/aos/1176344136

Serang, O., M. Mollinari, and A. A. F. Garcia, 2012 Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. PLoS One 7: e30906. https://doi.org/10.1371/journal.pone.0030906

Slater, A. T., N. O. I. Cogan, J. W. Forster, B. J. Hayes, and H. D. Daetwyler, 2016 Improving genetic gain with genomic selection in autotetraploid potato. Plant Genome 9: 1–15. https://doi.org/10.3835/plantgenome2016.02.0021

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard *et al.*, 2015 Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait, genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding line. PLoS One 11: e1004982. https://doi.org/10.1371/journal.pgen.1004982

Stich, B., and D. V. Inghelandt, 2018 Prospects and potential uses of genomic prediction of key performance traits in tetraploid potato. Front. Plant Sci. 9: 159. https://doi.org/10.3389/fpls.2018.00159

Sverrisdóttir, E., S. Byrne, E. H. R. Sundmark, H. O. Johnsen, K. H. Grethe *et al.*, 2017 Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. Theor. Appl. Genet. 130: 2091–2108. https://doi.org/10.1007/s00122-017-2944-y

Toledo-Silva, G., C. B. Cardoso-Silva, L. Jank, and A. P. Souza, 2013 De Novo transcriptome assembly for the tropical grass *Panicum maximum* Jacq. PLoS One 8: e70781. https://doi.org/10.1371/journal.pone.0070781

Uitdewilligen, J. G. A. M. L., A.-M. A. Wolters, B. B. D'hoop, T. J. A. Borm, R. G. F. Visser *et al.*, 2013 A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One 8: 1–14. https://doi.org/10.1371/journal.pone.0062355

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–4423. https://doi.org/10.3168/jds.2007-0980

Voorrips, R. E., G. Gort, and B. Vosman, 2011 Genotype calling in tetraploid species from bi-allelic marker data using mixture models. BMC Bioinformatics 12: 172. https://doi.org/10.1186/1471-2105-12-172

Vos, P. G., M. João Paulo, R. E. Voorips, R. G. F. Visser, H. J. van Eck *et al.*, 2017 Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theor. Appl. Genet. 130: 123–135. https://doi.org/10.1007/s00122-016-2798-8

Warmke, H. E., 1954 Apomixis in *Panicum maximum*. Am. J. Bot. 41: 5–11. https://doi.org/10.1002/j.1537-2197.1954.tb14297.x

Whittaker, J. C., R. Thompson, and M. C. Denham, 2000 Marker-assisted selection using ridge regression. Genet. Res. 75: 249–252. https://doi.org/10.1017/S0016672399004462

Wickham, H., 2009 *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wu, K. K., W. Burnquist, M. E. Sorrells, T. L. Tew, P. H. Moore *et al.*, 1992 The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83: 294–300. https://doi.org/10.1007/BF00224274

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569. https://doi.org/10.1038/ng.608

Zhang, X., P. Pérez-Rodríguez, J. Burgueño, M. Olsen, E. Buckler *et al.*, 2017 Rapid cycling genomic selection in a multiparental tropical maize population. G3 (Bethesda) 7: 2315–2326. https://doi.org/10.1534/g3.117.043141

Communicating editor: J. Holland