Mol Breeding (2019) 39: 100
https://doi.org/10.1007/s11032-019-1002-7

On the accuracy of genomic prediction models considering multi-trait and allele dosage in *Urochloa* spp. interspecific tetraploid hybrids

Filipe Inácio Matias · Filipe Couto Alves · Karem Guimarães Xavier Meireles · Sanzio Carvalho Lima Barrios · Cacilda Borges do Valle · Jeffrey B. Endelman · Roberto Fritsche-Neto

Received: 28 September 2018 / Accepted: 11 June 2019 / Published online: 28 June 2019

Abstract Currently, there is a lack of information regarding the employment of genomic prediction in tropical forages when compared to other crops and temperate forages. Moreover, genomic prediction models have been extensively developed for diploid species, whereas to apply those to polyploids most studies consider the genotypic information parametrized for diploids. This simplification can reduce the accuracy to estimate genetic effects and, consequently, the genomic breeding values. Another challenge is that agronomical and nutritional traits in forages frequently are negatively correlated and may have low heritability. To circumvent those problems one attractive alternative is the use of multi-trait prediction models. Therefore, we compared the impact of the ploidy parametrization over the prediction accuracy of agronomical and nutritional traits in *Urochloa* spp. hybrids using single and multi-trait models. GBLUP-A (additive) and GBLUP-AD (additive + dominance) showed similar prediction abilities in both single and multi-trait models. Conversely, combining GBLUP-AD and tetraploid information may improve the selection coincidence. Furthermore, the multi-trait Validation Scheme 2, where one trait is not evaluated for some individuals, can provide an increment of up to 30% of prediction ability. Therefore, it is an excellent strategy for traits with low heritability. Overall, all genomic selection models provided greater genetic gains than phenotypic selection. Similarly, the allele dosage associated with additive, dominance, and multi-trait factors increased the accuracy of genomic prediction models for interspecific polyploid hybrids. Finally, genomic prediction should be used in tropical forages breeding programs in order to reduce time.

Key message The allele dosage associated with additive, dominance, and multi-trait factors increases the accuracy of genomic prediction models for interspecific polyploid hybrids.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11032-019-1002-7) contains supplementary material, which is available to authorized users.

F. I. Matias (✉) · R. Fritsche-Neto
Genetics Department, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Av. Pádua Dias, 11, C. P. 9, 13.418-900, Piracicaba, São Paulo, Brazil
e-mail: filipematias23@usp.br

F. C. Alves
Departments of Epidemiology & Biostatistics, Michigan State University, East Lansing, MI 48824, USA

K. G. X. Meireles · S. C. L. Barrios · C. B. do Valle
Embrapa Beef Cattle, Av. Rádio Maia, 830, Zona Rural, 79.106-550, Campo Grande, MS, Brazil

J. B. Endelman
Department of Horticulture, University of Wisconsin, Madison, WI 53706, USA

Keywords Polyploid · Genotyping-by-sequencing · Tropical forage · Dominance

Introduction

Genomic prediction (GP) has been employed successfully in several species of plants and animals (Daetwyler et al. 2013; Jonas and De Koning 2013; Desta and Ortiz

2014; Meuwissen et al. 2016). This technique uses the whole-genome markers to predict complex traits (Meuwissen et al. 2001) and offers the opportunity to reduce the cost per cycle and the time required for variety development (Crossa et al. 2017). The accuracy of GP depends on the training and testing population sizes, trait heritability, number of markers, and statistical model (Heffner et al. 2009). Conversely, it is inversely proportional to the number of segregating chromosome sections in the target crop (Hayes et al. 2009). Some of these factors are considered challenges to be overcome in order to apply the genome prediction, mainly in polyploid species (Hayes et al. 2013).

Even playing an essential role in economy and contributing to the food security worldwide, the number of genomic prediction studies in polyploidy species is modest and mainly done in autopolyploid species (Gouy et al. 2013; Annicchiarico et al. 2015; Li et al. 2015; Biazzi et al. 2017; Sverrisdóttir et al. 2017; Endelman et al. 2018; You et al. 2018; Nyine et al. 2018). The problems in these species start with the reference genome, since most of the polyploid species do not have a complete genome sequence. Consequently, it is necessary to use the closest diploid species to compare and to make inferences (You et al. 2018). Also, the majority of current genomic sequencing tools are specific for diploids, and statistical approaches are necessary to predict the polyploids genotypes (Serang et al. 2012; Schmitz Carley et al. 2017). Those difficulties are due to the complexity of polyploid genomes and the necessity to use the allele dosage information for autopolyploid locus (Uitdewilligen et al. 2013). In most of the studies applying GP in polyploid species, the genotypic information is parametrized in a diploid level (Annicchiarico et al. 2015; Biazzi et al. 2017), having a few studies that consider a polyploid parametrization (Sverrisdóttir et al. 2017; de Bem Oliveira et al. 2018; Enciso-Rodriguez et al. 2018; Zingaretti et al. 2018; Nyine et al. 2018).

The influence of allele dosage in autopolyploid species was recently evaluated in genomic association studies (Ferrão et al. 2018; Sharma et al. 2018). These studies highlight that different gene action can be evaluated considering diploid and tetraploid models. Furthermore, the missed information of heterozygous is one of the most critical problems caused when the autopolyploid genome is simplified in diploidized data (Voorrips et al. 2011; Hackett et al. 2013). For example, in autotetraploid species, the genotypic classes simplex (*Aaaa*), duplex (*AAaa*), and triplex (*AAAa*) are summarized in a class by diploid data (*Aa*). Hence, this simplication may affect the correct estimation of allele substitution effects, dominance deviations, and consequently, the genomic breeding values.

The genotypic value in autotetraploid species is orthogonally decomposed among additive effects of each allele, digenetic dominance effects between the pair of alleles, trigenic, and quadrigenic interaction effects (Kempthorne 1957). In order to understand the influence and use of the allele dosage in genomic predictions in these situations, new methods to build the kinship matrix have been developed (Endelman et al. 2018).

The use of molecular breeding techniques is relatively undeveloped in tropical forage compared to other crops (Hayes et al. 2013). A common group of polyploid forage species in tropical climates is *Urochloa* (*syn. Brachiaria*). This genus has vast importance to feed cattle in the livestock business (Montagner et al. 2012; Euclides et al. 2016). Additionally, it has promoted the competitive of animal products in the international scenario and has been used to improve animal welfare (Jank et al. 2014). The primary commercial species are *U. brizantha*, *U. decumbens*, *U. humidicola*, and *U. ruziziensis*, and they were classified before as *Brachiaria* (Keller-Grein et al. 1996). The most important cultivars in tropical countries are apomictic and segmental allotetraploids, such as *U. brizantha* cv Marandu with part of the genome with allopolyploid behavior and part with autopolyploid behavior (Jank et al. 2014; Worthington et al. 2016; Bourke et al. 2018).

Usually, the whole selection process, from the generation of segregating populations to the releasing of new cultivars in tropical perennial forages, takes around 10–15 years. Furthermore, it is a hard-working and expensive process due to the evaluation of the animals' performance apart from plant performance (Jank et al. 2014). For instance, one selection cycle in these species demands in average 2 years, where phenotypic records of seven to ten cuttings are employed to evaluate the genetic value, stability, and adaptability of genotypes. Hence, genomic prediction methods can be a useful tool to reduce the costs due to the phenotyping expenses and the length of *Urochloa* spp. breeding cycle. In this sense, a simulation study of the feasibility of GP in a traditional forages breeding program (Resende et al. 2014) concluded that the individual genomic prediction method (INDG) could be useful when marker effects have been previously estimated. However, the genomic prediction may be ineffective depending on the

heritability of the target trait (de los Campos et al. 2013). Thus, an alternative is to exploit the correlation between traits to improve the predictive ability of the models by using the multi-trait approach (MTM). Through this approach, it is possible to use traits with higher heritability to improve the power to predict the other traits (Bauer and Léon 2008; Dos Santos et al. 2016; Fernandes et al. 2018).

Many traits of *Urochloa* are negatively correlated and have different heritabilities, such as crude protein and field green weight (Figueiredo et al. 2012; Matias et al. 2016, 2018). Consequently, the use of MTM could provide higher simultaneous selection gains for the inversely correlated traits (Bauer and Léon 2008; Guo et al. 2014). Thus, our main goal was to empirically evaluate the influence of multi-trait and the allele dosage information in genomic prediction accuracy in a diversity panel of *Urochloa* spp. hybrids.

Materials and methods

Genotypes

A representative subset of 272 individuals was selected from a larger population of tetraploid interspecific hybrids of *Urochloa* spp. This population was generated from crosses among apomictic cultivars of *U. brizantha* and tetraploid sexual access of *U. ruziziensis* in Embrapa Beef Cattle, Mato Grosso do Sul, Brazil (Matias et al. 2018). The genomic DNA was extracted by Qiagen® kit and genotyped by sequencing (GBS) (Elshire et al. 2011) using ApeKI enzyme and Illumina Hi-Seq 2500 platform. The sequencing data were evaluated using FastQC software (Andrews 2010) to determinate the quality by Phred score. The Cutadapt software (Martin 2011) was used to remove the barcodes and then the software Genome Analysis Toolkit (GATK) was used to discover single nucleotides polymorphisms using ploidy = 4 for genotype calling in the tetraploid level (McKenna et al. 2010; Depristo et al. 2011) and then made "diploidized" calls based on the genotype likelihoods. Furthermore, the software Burrows-Wheeler Alignment tool (BWA) (Li and Durbin 2009), SAMtools (Li et al. 2009; Li 2011), and Picard (http://broadinstitute.github.io/picard/) were used to align the reads, mark duplicate reads, and estimate the average insert size of the single-end reads, respectively. *Urochloa* spp. does not have complete reference genome available; then, six different genome references were used to the alignment step: *Setaria viridis* (Sv) (DOE-JGI 2018a), *Setaria italica* (Si) (Bennetzen et al. 2012), *Sorghum bicolor* (Sb) (DOE-JGI 2018b), *Oryza sativa* (Os) (Ouyang et al. 2006), *Zea mays* (Zm) (Schnable et al. 2009), and the *Urochloa* mock reference (Um) (data not shown).

All aligned markers with median depth (MedianDP) ≤ 8, minimum allele depth (MAD) ≤ 2, minor allele frequency (MAF) ≤ 0.01, and missing data ≥ 50% were eliminated. Also, samples with DepthPerSample (DP) < 8 were set as missing, and samples with genotype quality (QD) ≤ ten were eliminated. The filtration criteria described above were adequate for the diploid level. However, we had the interest to evaluate the performance of permissive filtration criteria in predictions of the greater polyploid level, then all markers selected in diploid level was extended to tetraploid level. Therefore, a total of 26,535 SNPs were selected and used in the diploid and tetraploid level. The remained missing data were imputed by the package Random Forest from R software (Liaw and Wiener 2002) for each ploidy; in particular, all markers that had $r^2 \geq 0.1$ with the inputting locus were used as predictors and 300 trees were used to fit the algorithm. Genotype data set is available at http://www.genetica.esalq.usp.br/alogamas/data.html.

Phenotypes

The population was evaluated in the field (20° 27′ S; 54° 57′ W) for 2 years (2013 and 2014) over seven cuttings using an incomplete block design with ten blocks. The plots consisted of squares covering 2.25 m^2. Nine genotypes were added to each block as checks and used to estimate the environmental effect of the statistical design. The checks were *U. brizantha* cultivar "Marandu," *U. brizantha* cultivar "Paiaguás," *U. decumbens* cultivar "Basilisk," the interspecific commercial hybrid "Mulato II," the accession "B140" of *U. brizantha*, and the *U. ruziziensis* sexual hybrids "BS9," "BS15," "336-T1," and "336-T2." Additional information about experimental design and biological material (hybrids and checks) are available on Matias et al. (2018).

The materials were agronomically evaluated by cutting the plants around 10 cm above the soil surface and weighed using a dynamometer to set the field green weight (FGW) in kg ha^{-1}. Seven days after each cutting, the regrowth capacity (REG) was obtained by the combination of scores for the density of regrown tillers and

regrowth speed (Figueiredo et al. 2012). The nutritional traits were evaluated only on the third and fourth cutting, where the crude protein (CP) and neutral detergent fiber (NDF) were measured by infrared reflectance spectroscopy (NIRS) (Marten et al. 1984). The calibration of the NIRS was performed previously by comparing the results obtained in the wet chemical analyses; the spectrum read from these same samples in the NIRS for several nutritional characteristics (unpublished data).

We employed a two-step approach to adjust the phenotypic record of each hybrid. First, the block effect was estimated considering a complete randomized block design as described in the Eq. 1. Posteriorly, the block effects were deducted from observed data of each hybrid in function of the field position. Then, the new corrected trait data were evaluated according to the Eq. 2.

$$y_{\mathrm{bcd}} = \mu + p_b + q_c + s_d + u_{b_{\times}c} + \varepsilon_{\mathrm{bcd}} \quad (1)$$

$$y^{*}_{\mathrm{gc}} = \mu + p^{*}_g + q_c + u^{*}_{g_{\times}c} + \varepsilon_{\mathrm{gc}} \quad (2)$$

where $\boldsymbol{y}$ is the vector of checks phenotypic data; $\boldsymbol{y^*}$ is the vector of hybrid's corrected phenotypes; $\boldsymbol{\mu}$ is the intercept; $\boldsymbol{p}$ is the vector of check effects, considered as fixed, with $b = \{1, 2, \ldots, 9\}$; $\boldsymbol{p^*}$ is the vector of hybrids effects, considered as fixed, with $g = \{1, 2, \ldots, 272\}$; $\boldsymbol{q}$ is the vector of cut effects, considered as fixed, with $c = \{1, 2, \ldots, 7\}$ for agronomical traits and $c = \{3, 4\}$ for nutritional traits; $\boldsymbol{s}$ is the vector of block effects, considered as fixed, with $d = \{1, 2, \ldots, 10\}$; $\boldsymbol{u}$ is the vector of the check by cut interaction effects, considered as random, with $\boldsymbol{u}\sim N\left(0, \boldsymbol{I}\sigma^2_{b_{\times}c}\right)$ where $\boldsymbol{I}$ is the identity matrix and $\sigma^2_{b_{\times}c}$ is the variance component of described interaction; $\boldsymbol{u^*}$ is the vector of the hybrid by cut interaction effects, considered as random, with $\boldsymbol{u}^{*}\sim N\left(0, \boldsymbol{I}\sigma^2_{g_{\times}c}\right)$ where $\sigma^2_{g_{\times}c}$ is the variance component of the described interaction; and $\boldsymbol{\varepsilon}$ is the residual vector with $\boldsymbol{\varepsilon}\sim N\left(0, \boldsymbol{I}\sigma^2_{\varepsilon}\right)$ where σ^2_{ε} is the variance component of error.

All models used to obtain the genetic value of each hybrid and the significance tests were fitted using the *ASreml-R* package (Butler et al. 2009).

Regression models applied to study the dosage information

Genomic values were predicted using the additive and additive+dominance GBLUP (GBLUP-A and GBLUP-AD, respectively) assuming the model:

$$y = 1\mu + Za + Td + \varepsilon \quad (3)$$

where $\boldsymbol{y}$ is the vector of genetic hybrids values from the Eq. 2, $\boldsymbol{\mu}$ is the intercept, $\boldsymbol{a}$ is the vector of additive effect with $\boldsymbol{a}\sim N(0, \boldsymbol{G}\sigma^2_a)$, $\boldsymbol{d}$ is the vector of dominance effect with $\boldsymbol{d}\sim N(0, \boldsymbol{D}\sigma^2_d)$, $\boldsymbol{\varepsilon}$ is the residual vector with $\boldsymbol{\varepsilon}\sim N\left(0, \boldsymbol{I}\sigma^2_{\varepsilon}\right)$. σ^2_{ε}, σ^2_a, and σ^2_d is the variance component of error, additivity, and dominance, respectively. $\boldsymbol{G}$ and $\boldsymbol{D}$ are the covariance matrices associated with the additive and dominance effects, respectively. $\boldsymbol{I}$ is the identity matrix. $\boldsymbol{Z}$ and $\boldsymbol{T}$ are the incidence matrices of each assumed genetic effect. The genomic kinship matrices for additive and dominant effects for diploid genetic configurations were estimated according to Vitezica et al. (2013) and tetraploid according to Endelman et al. (2018) following the equations:

Diploid additive:

$$W_{\mathrm{Dip}} = \left(X_{\mathrm{Dip}} - 2p_i\right)$$
$$G_{\mathrm{Dip}} = \frac{W_{Dip} W_{Dip}'}{\sum 2p_i(1-p_i)}$$

Diploid dominance:

$$S_{\mathrm{Dip}} = 2p_i X_{\mathrm{Dip}} - 2p_i^2 - X_{\mathrm{Dip}}\left(X_{\mathrm{Dip}} - 1\right)$$
$$D_{\mathrm{Dip}} = \frac{S_{\mathrm{Dip}} S_{\mathrm{Dip}}'}{\sum 4p_i^2(1-p_i)^2}$$

Tetraploid additive:

$$W_{\mathrm{Tetra}} = \left(\mathrm{X}_{\mathrm{Tetra}} - 4\mathrm{p_i}\right)$$
$$G_{\mathrm{Tetra}} = \frac{W_{\mathrm{Tetra}} W_{\mathrm{Tetra}}'}{\sum 4p_i(1-p_i)}$$

Tetraploid dominance:

$$S_{\mathrm{Tetra}} = 6{p_i}^2 - 3p_i X_{\mathrm{Tetra}} + \frac{X_{\mathrm{Tetra}}\left(X_{\mathrm{Tetra}} - 1\right)}{2}$$
$$D_{\mathrm{Tetra}} = \frac{S_{\mathrm{Tetra}} S_{\mathrm{Tetra}}'}{\sum 6p_i^2(1-p_i)^2}$$

where p_i is the reference allele frequency, and $\boldsymbol{X}$ is the allele dosage matrix with genotypes ($\boldsymbol{X}_{\mathrm{Dip}}$ to diploid level and $\boldsymbol{X}_{\mathrm{Tetra}}$ to tetraploid level).

Selection approaches and validation systems

1- Single trait model (INDG)

To estimate the predictive ability ($r_{\hat{y}y}$) of each scenario (Trait+Ploidy+GBLUP) using the INDG scheme,

we randomly splitted the population into training (TP, 75% of the individuals) and validation (VP, 25% of the genotypes) sets. This process was repeated 100 times for each scenario. For each random sample replicate, we assessed the prediction ability by estimating the Pearson's correlation among the predicted and observed phenotypes of the individuals from VP ($r_{\hat{y}y}$). Finally, we compared the evaluated scenarios by the average of the 100 prediction ability estimates. The genomic prediction analyses were carried out using the BGLR-R package (Pérez and de los Campos 2013) assuming 30,000 Gibbs samples, a burn-in of 5000, and thinning of 5.

2- Multi-trait model (MTM)

The four traits (FGW, REG, CP, and FDN) were evaluated in a Bayesian Multivariate Gaussian Models using the MTM package in R software (de los Campos, http://quantgen.github.io/MTM/vignette.html), following the equation described by Lehermeier et al. (2015):

$$y_{\mathrm{ni}} = \mu + \beta_i + u_A + u_D + \varepsilon_{\mathrm{ni}} \quad (4)$$

where $y_{\mathrm{ni}} = (y_{1i}, \ldots, y_{\mathrm{ni}})'$ is the phenotypic data with i equal the number of traits $i = \{1, 2, 3, 4\}$ and n the number of hybrids $n = \{1, 2, \ldots, 272\}$; $\boldsymbol{\mu}$ is the model intercept; $\boldsymbol{\beta}$ is the vector of the i^{th} trait effect; $\boldsymbol{u}$ is the genetic vector of hybrids, considered as random, with $\boldsymbol{u_A} \sim MVN(0, \sum_a \otimes \boldsymbol{G})$ and $\boldsymbol{u_D} \sim MVN(0, \sum_d \otimes \boldsymbol{D})$, where $\boldsymbol{G}$ is the additive matrix (G_{Dip} and G_{Tetra}), $\boldsymbol{D}$ is the dominance matrix (D_{Dip} and D_{Tetra}) and $\otimes$ denotes the Kronecker product; $\sum_a = \boldsymbol{B_a}p$ and $\sum_d = \boldsymbol{B_d}p$ are the additive and dominance genomic variance-covariance matrices among traits; ε is the residual vector with $\varepsilon \sim MVN\left(0, \mathbf{I}\sigma_\varepsilon^2\right)$.

For the multi-trait genomic method, we considered two different validation schemes. The first (VS1) considers a scenario in which an individual is not evaluated for any trait. This scheme mimics the situations in which the breeder desires to predict the performance of newly developed materials, without any phenotypic record (Supplementary Fig.S1.A). The second (VS2), assume that the breeders aim to predict the performance of a particular individual for a determined trait (i.e., neutral detergent fiber–NDF) based on the phenotypic records of other traits in which this material was previously phenotyped (Supplementary Fig.S1.B). For each trait we tested three scenarios assuming a training set (TP) of 75, 50, and 25% of the total population size. The sampling process was repeated 100 times for each scenario (MTM + VP size+Ploidy+GBLUP). We assessed the prediction ability by estimating the Pearson's correlation among the predicted and observed phenotypes and compared the average of the 100 prediction ability estimates. We assumed 30,000 Gibbs samples, a burn-in of 5000 and thinned of 5.

The validation error bar was calculated by $\mathrm{SE} = \mathrm{SD} \times \sqrt{\frac{1}{\mathrm{n}} + \frac{\mathrm{n2}}{\mathrm{n1}}}$, where SE is the standard error, SD is the standard deviation, $n = 272$, and $\frac{n2}{n1}$ is the ratio of $\frac{\mathrm{VP}}{\mathrm{TP}}$ size (Bouckaert and Frank 2004).

Selection coincidence, genetic gains, and validation approach comparations

The selection coincidence was estimated considering the coincidence of the 10% best selected materials in both genomic and phenotypic selection. The genetic gains were estimated by $\Delta G = \frac{ir\sigma_a}{L}$, where i is the standardized selection intensity, r is the model accuracy, σ_a is the genetic standard deviation, and L is the generation interval (Hayes et al. 2013). r is the average prediction ability of all scenarios (Trait+Ploidy+GBLUP) for the genomic approaches, and the square root of heritability for phenotypic selection (PS). L is the equal to 4.5 years and represents the summation between the necessary time to conclude the breeding's first stage in a traditional tropical forage breeding program (4 years for phenotypic selection) and to get seedlings to extract DNA (6 months for genotypic prediction). The standardized selection intensity was fixed at 10% ($i = 1.76$). The genomic prediction approaches INDG and MTM were compared with the phenotypic selection by the ratio among genetic gains obtained using the GP and PS ($\Delta G_{\mathrm{genomic:PS}} = {}^{\Delta G_{\mathrm{genomic}}}/_{\Delta G_{\mathrm{PS}}}$). As the values of σ_a and i are equal for both GP and PS, the genetic gains were estimated by $\Delta G = \frac{r}{L}$. The heritability was estimated by the following equation: $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{g_x c}^2}{c} + \frac{\sigma_{\bar{\varepsilon}}^2}{c*b}}$, where σ_g^2 is the genetic variance of hybrids, $\sigma_{g_x c}^2$ in the interaction between hybrids and cuttings, c is the number of cuttings, and b in the number of blocks.

Results

Phenotypic selection

Significant genetic effects were found for all studied traits (Table 1), highlighting the possibility of genetic gains when applying the phenotypic selection in this *Urochloa* spp. population. The heritabilities' estimates were high for agronomical traits, 0.81 and 0.75 for field green weight and regrowth capacity, respectively. On the other hand, the heritabilities were moderate for nutritional traits, 68% for crude protein and 52% for neutral detergent fiber. The highest and lowest genetic gains were 9.17% and 0.52% for CP and NDF, respectively (Table 1). It is important to point out that, the selection in tropical forages is made to improve FGW, REG, and CP and to reduce the NDF during the breeding cycles. Regarding the univariate phenotypic selection, the population performance for a given trait in the next generation is estimated by the response to selection $\boldsymbol{SG}=\Delta\,\overline{\boldsymbol{X}}\,\boldsymbol{*H^2}$ plus the original population mean ($\overline{\boldsymbol{X}}_{\boldsymbol{pop}}$). The phenotypic correlation between the traits varied from moderate to low (0.49 and −0.24 for CP×NDF and FGW×NDF, respectively, Supplementary Fig.S2 and Supplementary Table.S1).

Single-trait genomic prediction

The average single trait's prediction ability (r) across genomic models was 0.20, 0.15, 0.13, and 0.31 for crude protein, green weight, fiber, and regrow capacity, respectively. Slightly differences were observed among the r estimates obtained by the GBLUP-A and GBLUP-AD using 75/25 cross-validation (INDG) (Fig. 1a). For CP and FGW, the GBLUP-A prediction ability was equal to or larger than those obtained when fitting the GBLUP-AD (Fig. 1a). Regarding NDF and REG, the r estimates were not as consistent as for the other traits, and the best prediction model varied according to the ploidy considered to build the kinship matrices. The tetraploid level had slightly lower prediction ability than the diploid level for CP and FGW. However, the tetraploid level led to better results to predict REG and NDF (Fig. 1a).

Breeders should choose the genomic selection model not by only considering the prediction ability but also the selection coincidence. With the exception of CP considering the diploid parametrization, the GBLUP-AD performed better than GBLUP-A (Fig. 1b). On average, GBLUP-AD showed 0–5% higher selection coincidence than GBLUP-A for crude protein, and 5% for regrow capacity. Concerning the dosage information, the tetraploid level combined to the GBLUP-AD, in general, improved the selection coincidence for all traits (Fig. 1b).

Multi-trait genomic prediction

The differences in the prediction ability within trait using the Validation Scheme 1 (VS1) were modest, even varying the training population sizes (Fig. 2a). The smallest prediction accuracy was observed for NDF, ranging from 0.05 (TP = 25%, GBLUP-A-Diploid) to 0.15 (TP = 75%, AD-Tetraploid). Conversely, it was the trait that showed the larger variation on the prediction ability according to the training population size. For most scenarios, the additive model led to slightly higher prediction accuracies for both diploid and tetraploid parametrization. On the other hand, for the majority of the traits, the combination between GBLUP-AD and tetraploid parametrization provided the largest selection coincidence estimates.

The Validation Scheme 2 (VS2) showed the highest prediction abilities for all evaluated scenarios (Figs. 2a and 3a). The prediction abilities for crude protein in VS2 were, in average, ~43 to 76% (training set sizes of 25 and 75%, respectively) higher than those obtained in VS1. The same trend was observed for the other traits, where the prediction ability highly increased according to the training population size and scheme.

Regrowth ability showed the highest prediction values in both validation schemes (Figs. 2a and 3a). However, its difference in relation to the other traits reduced in VS2. For instance, the difference of prediction ability between REG and NDF, considering VS1 was around 0.10–0.20 and reduced to 0.02–0.06 using VS2.

Overall, the GBLUP-AD provided higher prediction ability estimates than GBLUP-A in VS2. However, as observed for single-trait predictions, the differences between the additive and additive-dominance GBLUP were modest, being the most considerable differences observed in small training populations sizes, such as 25% (Fig. 3a). Regarding the ploidy information, the differences in prediction abilities between diploid and tetraploid parametrizations were small for all scenarios. Although, as one can observe, the selection coincidence increases according to the training set size and the ploidy considered to build the relationship matrices (Fig. 3b). Furthermore, the tetraploid matrix led to

Table 1 Wald test for fixed effects of genotype, broad heritability (H^2), average of population ($\overline{X}_{pop}$), average of 10% best hybrids ($\overline{X}_{10\%Best}$), average of 10% worst hybrids ($\overline{X}_{10\%Worst}$), selection differential ($\Delta\overline{X}=\overline{X}_{10\%Best}-\overline{X}_{pop}$), and response to selection ($SG\% = \Delta\overline{X}*H^2/\overline{X}_{pop}$) of field green weight (FGW), regrowth capacity (REG), crude protein (CP), and neutral detergent fiber (NDF)

Parameters	FGW	REG	CP	NDF
Genotype	8181.20**	5035.10**	3374.10**	2267.70**
H^2	0.81	0.75	0.68	0.52
$\overline{X}_{pop}$	1468.26	3.23	15.80	65.91
$\overline{X}_{10\%Best}$	1614.57	3.47	17.93	59.32
$\overline{X}_{10\%Worst}$	447.88	1.83	13.11	70.81
$\Delta\overline{X}$	146.31	0.24	2.13	−6.59
$SG\%$	8.07	5.57	9.17	−0.52

** significant at 0.05

highest coincidence estimates than the diploid matrix, achieving a maximum of 0.63 for FGW.

Genetic gains

Using single-trait's genomic selection (INDG) increased the response to selection in 1.543, 3.322, 2.431, and 2.038 times in comparison to phenotypic selection for FGW, REG, CP, and NDF, respectively (Table 2). The gains obtained by multi-trait prediction in VS1 were similar to those observed by the INDG approach. The highest genetic gains per unit of time were obtained using the VS2 with $\Delta G_{\text{genomic}:\text{PS}}$ varying from ~3.6 to ~6.7 for FGW and NDF, respectively (Table 2).

It is interesting to highlight the positive and high correlation between INDG and MTM methods (Supplementary

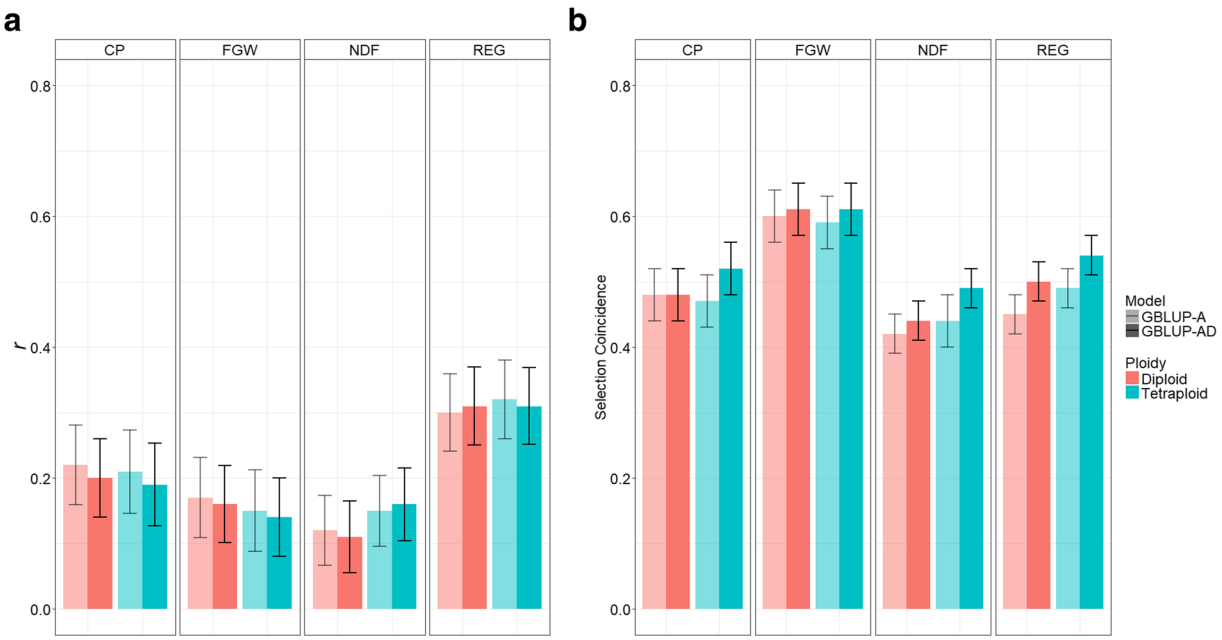

Fig. 1 INDG univariate (75/25) genomic prediction approach. **a** Predictive ability of INDG genomic prediction and **b** selection coincidence between the10% of the best hybrids selected by phenotypic selection and by genomic prediction carried out using two prediction models (GBLUP-A and GBLUP-AD) and two levels of ploidy (diploid and tetraploid) for (FGW) field green weight, (REG) regrowth ability, (CP) crude protein, and (NDF) neutral detergent fiber in a *Urochloa* spp. hybrid panel

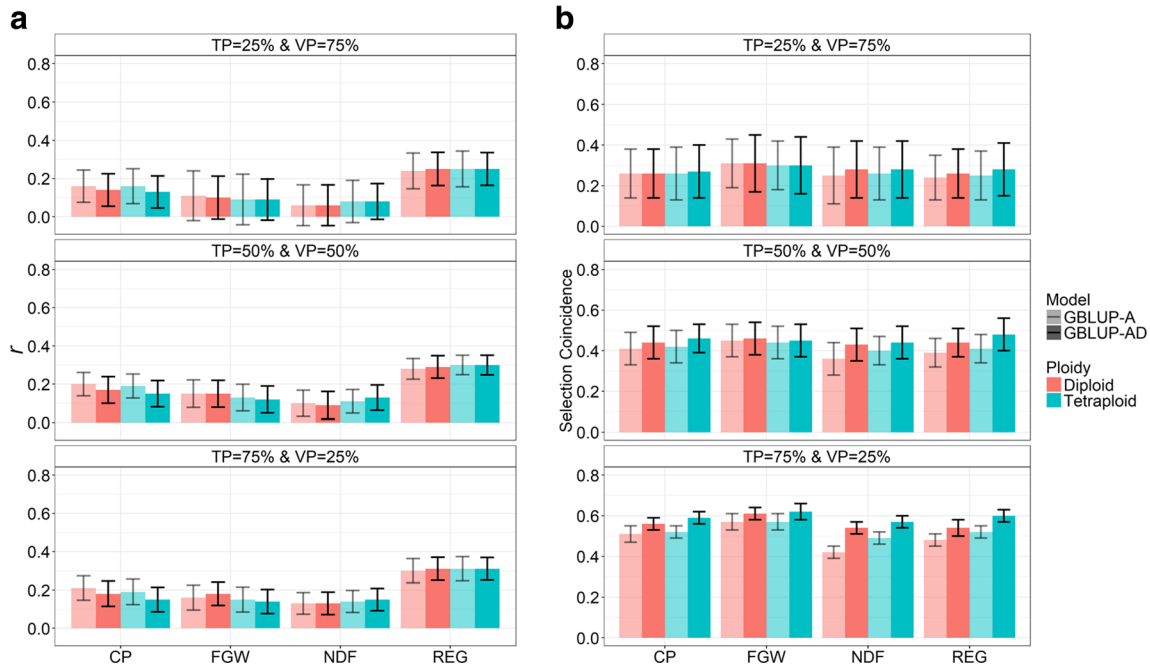

Fig. 2 Validation Scheme 1 (VS1) for multi-trait genomic prediction approach. **a** Predictive ability of INDG genomic prediction and **b** selection coincidence between the 10% of the best hybrids selected by phenotypic selection and by genomic prediction carried out using two prediction models (GBLUP-A and GBLUP-AD) and two levels of ploidy (diploid and tetraploid) for (FGW) field green weight, (REG) regrowth ability, (CP) crude protein, and (NDF) neutral detergent fiber in a *Urochloa* spp. hybrid panel. Three sizes of training population (TP) and validation population (VP) were evaluated 1 - {TP = 75% and VP = 25%}, 2 - {TP = 50% and VP = 50%}, and 3 - {TP = 25% and VP = 75%}

Table.S2). In this case, the genomic prediction performance of individuals is similar among the methods. Non significance was observed between FGW and CP using the phenotypic correlation (Supplementary Table.S1); however, all genomic correlations between these traits were significant (Supplementary Table.S2). Also, in general, the correlation values were greater using the genomic values predicted by MTM methods than the correlation estimated by the phenotypic values, for example the correlation between FGW and REG using the MTM-V2 was 73%.

Discussion

Genomic prediction in polyploids

Currently, there is a lack of information regarding the employment of genomic prediction in tropical forages breeding. Furthermore, the genomic prediction has been extensively applied for diploid species in detriment of polyploid species (You et al. 2018). The use of allele dosage in genomic prediction is important since, as described by Kempthorne (1957), the genetic value (GV_{ijkl}) of a tetraploid *genotype* $A_iA_jA_kA_l$ *can be partitioned into:* $GV_{ijkl} = \mu + \alpha i + \alpha j + \alpha k + \alpha l + \beta ij + \beta ik + \beta il + \beta jk + \beta jl + \beta kl + \gamma ijk + \gamma ijl + \gamma ikl + \gamma jkl + \delta ijk$, where μ is the population mean, α is the main effects of each allele, β is the diallelic interaction effect, γ is the triallelic interaction effect, and δ is the tetra-allelic interactions effect. Conversely, the genetic value in a diploid genotype (GV_{ij}) is decomposed only in $\mu + \alpha i + \alpha j + \beta ij$. Therefore, simplifying a tetraploid genotype as a diploid may include bias on the genomic predictions due to high order interactions and allele substitution effects to be estimated. In our study, we compared the ploidy's parametrization impact on the prediction accuracy of agronomical traits in *Urochloa* spp. hybrids using single and multi-trait models.

To apply the tetraploid dosage information in genomic prediction, we assumed that the studied traits are controlled by many genes with small effects distributed on the whole genome. In this case, the locus dosage should be diminished by the number of markers and genome coverage. Usually, high-density marker panel shows the best predictions accuracy (Guo et al. 2012; Pérez-Rodríguez et al.

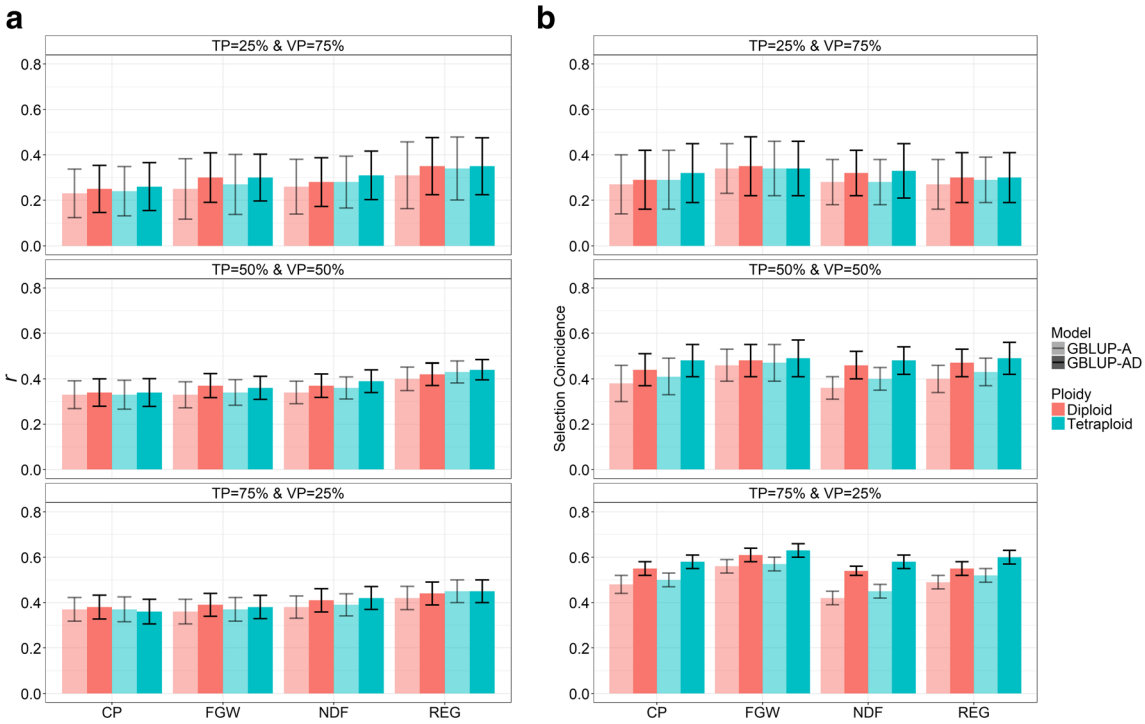

Fig. 3 Validation Scheme 2 (VS2) for multi-trait genomic prediction approach. **a** Predictive ability of INDG genomic prediction and **b** selection coincidence between the 10% of the best hybrids selected by phenotypic selection and by genomic prediction carried out using two prediction models (GBLUP-A and GBLUP-AD) and two levels of ploidy (diploid and tetraploid) for (FGW) field green weight, (REG) regrowth ability, (CP) crude protein, and (NDF) neutral detergent fiber in a *Urochloa* spp. hybrid panel. Three sizes of training population (TP) and validation population (VP) were evaluated 1 - {TP = 75% and VP = 25%}, 2 - {TP = 50% and VP = 50%}, and 3 - {TP = 25% and VP = 75%}

2012; Combs and Bernardo 2013). Thus, as *Urochloa* spp. are segmental allotetraploid species, we applied a high-quality filtering for diploid level to select markers in the diploid regions of the *Urochloa* genome during the quality control process. Our results indicated that the influence of the ploidy in our population depends on the trait. For example, diploid and tetraploid levels showed little differences in prediction ability between them for FGW, CP, and REG (Figs. 1a, 2a, and 3a). However, the prediction ability using tetraploid level was at least 2–5% greater than the diploid for NDF.

Generally, the dosage diagnostic in GBS genotype calls for polyploid species demands high read depths, as 100x for sugarcane (Song et al. 2016), 48.7x for strawberry (Bassil et al. 2015), and 60–80x for potato (Uitdewilligen et al. 2013). High read depths are necessary due to the difficulty in calling the genontypes in polyploid species. In tetraploid species, as the *Urochloa* spp., there are five different genotype categories to be distinguished: nulliplex (0, aaaa), simplex (1, Aaaa), duplex (2, AAaa), triplex (3, AAAa), and tetraplex (4, AAAA) (Uitdewilligen et al. 2013). In these situations, low read depth is a barrier to cross and identify correctly the simplex, duplex, and triplex genotypes; once identifying them is more challenging than the nulliplex and the tetraplex (Serang et al. 2012; Rosyara et al. 2016; Schmitz Carley et al. 2017). Therefore, it is possible to infer that in scenarios where the genomic data has low reading depth grouping, the different intermediate genotype classes (duplex, triplex, tetraplex) in one class, i.e., considering the diploid parametrization, seem to be a good strategy to circumvent the problem and apply genomic tools as observed by the slight difference of prediction ability between ploidy levels (Figs. 1a, 2a, and 3a).

As previously pointed, the dosage information in genomic tools needs to be evaluated carefully. *Urochloa* spp. in this study are segmental allotetraploids hybrids; it means that their genome is organized part as

Table 2 Genetic gain per unit of time comparing the phenotypic selection (PS) to 75/25 cross-validation (INDG) and multi-trait prediction (MTM) with validation population size equal to 25% of population for field green weight (FGW), regrowing ability (REG), crude protein (CP), and neutral detergent fiber (NDF) from in interspecific *hybrid Urochloa* spp. panel

Methods	Parameters	FGW	REG	CP	NDF
PS	H^2	0.810	0.750	0.680	0.520
	$r = \sqrt{H^2}$	0.900	0.866	0.825	0.721
	ΔG_{PS}	0.225	0.217	0.206	0.180
INDG	$r = {}^{r_{y\hat{y}}}/_{\sqrt{H^2}}$	0.174	0.360	0.251	0.184
	$\Delta G_{\mathrm{genomic}}$	0.347	0.719	0.501	0.367
	$\Delta G_{\mathrm{genomic:PS}}$	1.543	3.322	2.431	2.038
MTM-VS1	$r = {}^{r_{y\hat{y}}}/_{\sqrt{H^2}}$	0.156	0.346	0.255	0.208
	$\Delta G_{\mathrm{genomic}}$	0.311	0.693	0.509	0.416
	$\Delta G_{\mathrm{genomic:PS}}$	1.383	3.200	2.471	2.308
MTM-VS2	$r = {}^{r_{y\hat{y}}}/_{\sqrt{H^2}}$	0.411	0.439	0.485	0.610
	$\Delta G_{\mathrm{genomic}}$	0.822	0.878	0.970	1.220
	$\Delta G_{\mathrm{genomic:PS}}$	3.654	4.053	4.706	6.769

Additive variance (H^2), accuracy (r), genetic gain using PS (ΔG_{PS}), genetic gain using genomic selection ($\Delta G_{\mathrm{genomic}}$), and ratio gain ($\Delta G_{\mathrm{genomic:PS}}$)

autotetraploid and part allotetraploid (Mendes-Bonato et al. 2002; Worthington et al. 2016). This suggests that using genotype matrices obtained through diploid species quality filters in *Urochloa* spp. for GS studies is feasible due to the nature of Brachiaria's genome organization. Slight differences in prediction accuracies between tetra and diploid matrices were found in our population. These results differ from Nyine et al. (2018) findings, who observed a significant reduction in the prediction accuracy when considering allele dosage in triploid banana. According to the authors, the reduction on prediction acuracy can be due to the variation on the minor allele frequency across loci. As pointed by them, the MAF has a significant impact on the estimation of SNP effects. However, the selection coincidence also should be accounted by the breeder; as one can observe, the more considerable selection coincidence provided by tetraploid information is a strong argument to accounting the tetraploid level in the genotyping calling step for allotetraploid segmental species (Figs. 1, 2, and 3). It is important to highlight that for a complete autotetraploid species, the results could be different.

In this context, our study is the first to compare both diploid and tetraploid matrices for interspecific hybrids, although we were very permissive for the tetraploid markers calling. Therefore, further studies accounting for a "recommended" read depth in polyploid markers calling are necessary.

Genomic prediction in *Urochloa* spp. hybrids

Several problems can be circumvented in a breeding program by using genomic prediction. The most common is the situation in which the material was not evaluated in field trials, and the breeder aims to determine the best materials for the field evaluations. This scenario is mimicked by two of the tested validation schemes we used, the INDG and MTM-Validation Scheme 1 (VS1). In this case, all recovered information for the trait prediction will mainly be due to the genetic relationship between the training and testing sets (Burgueño et al. 2012; Crossa et al. 2017). Another typical situation observed in a breeding program is that a developed material was not evaluated for all traits. In this case, the material's performance for a determined trait can be easily obtained by using multi-trait prediction analyses and can take advantage of the genetic correlations among the considered traits (Guo et al. 2014; Lyra et al. 2017) and can broadly increase the prediction accuracy of genomic models. This scenario can be represented by the MTM-Validation Scheme 2

(VS2), and it is similar to the "trait-assisted GS approach" proposed by Fernandes et al. (2018). However, different with the proposed approach by Fernandes, in our study, we considered three traits as assisting the prediction of one. Furthermore, we considered a variation on the training sets the size for the MTM models.

As expected, INDG and VS1 did not show significative differences in their prediction accuracies for the same training population size (Figs. 1a, and 2a, TP = 75%). This result is in accordance with some previous studies in annual crops (Lyra et al. 2017; Fernandes et al. 2018) but differs of several others (Jia and Jannink 2012; Guo et al. 2014; Schulthess et al. 2016). As pointed before, the information recovered in these validation schemes is mainly due to the relationship among genotypes within traits. It is similar to the CV1 proposed by Burgueño et al. (2012), a validation scheme commonly employed in multi-environment prediction models in crops (Lopez-Cruz et al. 2015; Souza et al. 2017).

The prediction ability of INDG and MTM were lower than 0.45 for all traits (Figs. 1 and 2). Frequently, low accuracy of genomic prediction was found for agronomic and nutritional traits in polyploid forages. For instance, alfalfa forage quality traits show low accuracy for neutral detergent fiber (leaf and stem) and crude protein (stem) (Biazzi et al. 2017). In our case, this fact was not verified due to green weight which had high heritability around 80% but the prediction capacity was 0.15 and 0.40 using INDG and MTM-VS2 approaches, respectively. Moreover, the markers density influences the predictions (Resende et al. 2014). Hence, the absence of reference genome and the complexity of *Urochloa* genome make it challenging to cover the whole genome.

The prediction of newly developed materials using single or multi-trait models (INDG and MTM-VS1) showed smaller prediction accuracy than the observed by the MTM-VS2 (Figs. 1a, 2a, and 3a). This result is in accordance with the findings of Fernandes et al. (2018), who found in sorghum more substantial prediction accuracy of the "trait-assisted GS" when compared to single and multi-trait prediction models for biomass yield. Also, it is interesting to note that, even considering the smallest training population size (TP = 25%), the MTM-VS2 provided higher prediction than the other validation schemes (INDG and VS1) which shows the efficiency of this scheme in comparison to the other validation methods.

GBLUP-A and GBLUP-AD models presented similar prediction abilities (Figs. 1a, 2a, and 3a). In general, GBLUP-A was slightly superior to GBLUP-AD using the INDG approach. Probably, this dataset is not large enough to capture and estimate the nonadditive effects accurately, such as was observed in alfalfa (Biazzi et al. 2017). In our study, we cannot conclude precisely the influence of additive and nonaddictive marker effects on the prediction of new materials. Despite this, the inclusion of nonadditive effects in the genomic model can improve the prediction ability in diploid species as pine (De Almeida Filho et al. 2016), eucalyptus (Tan et al. 2018), maize (Dias et al. 2018; Alves et al. 2019), and also tetraploid species as potato (Endelman et al. 2018). On the other hand, the scenarios with GBLUP-AD showed higher selection coincidence than GBLUP-A for all traits, mainly when combined with tetraploid information (Figs. 1b, 2b, and 3b). We observed that more levels of allele dosages provide subtle but different configurations of additive and non-additive kinship matrix. These differences were not large enough to improve the prediction models' accuracy but could better explain the genetic variability and approximate the genomic prediction rank to the real rank for all traits.

Furthermore, we believe that there is a trade-off between the amount and quality of information and accuracy. Thus, even if biologically correct, depending on the number and depth of markers, it is not possible to estimate all the genetic effects accurately. Therefore, this bias on the estimates led the models to perform similarly under poor conditions. However, the differences between them tend to appear regarding the data improvement.

The genomic prediction in *Urochloa* spp. breeding programs

Forages breeding programs focus on improving yield and quality of herbage to support feed conversion into meat or milk (Jank et al. 2014). Field green weight, regrowth ability, crude protein, and fiber require expensive and destructive measurement, which make them good candidates for GP (Hayes 2013). Also, reducing costs and increasing genetic gain *per unit* of time are common aims of any breeding program. According to Resende et al. (2014), during the early stages of forage breeding programs, several traits must be selected simultaneously. In this case, it is possible to use markers associated with multiple traits as a tool for a multivariate

selection. Our study provided an insight of the use of genomic prediction models on early stages of a traditional *Urochloa* spp. breeding program (Jank et al. 2014), but also could be extended to later stages.

Genomic prediction should be used when the phenotypic selection (PS) has lower predictability using sward conditions, low meaningful selection pressure within families, and long and expensive phenotyping cycle (Resende et al. 2014). Different schemes of application of GP in forages were described in the literature (Hayes et al. 2013; Resende et al. 2014; Biazzi et al. 2017). However, this is the first GS study applied to tropical forages. We compared genetic gains achieved for critical agronomic traits when GP is used in a tropical forage breeding program (Table 2). Despite the low values of prediction accuracy, the gains in unit of time provided for all scenarios, larger genetic gains for genomic selection (GS) when compared to the phenotypic selection (Table 2). This is in accordance with several authors (Heffner et al. 2009; Hayes et al. 2013; Crossa et al. 2017) and indicates that the employment of genomic tools for *Urochloa* spp. hybrids prediction should be adopted in breeding programs.

The genetic knowledge of forage crops is small compared with cereals. Few studies applying genomic prediction to real polyploid forage dataset are available (Annicchiarico et al. 2015; Biazzi et al. 2017), and we believe that our work can improve the understanding of different selection processes on apomictic forage breeding programs. We evaluated the influence of polyploidy and the quality of genotyping call on predictions. No difference of prediction ability was observed using GBLUP-A and GBLUP-AD; however, combining GBLUP-AD and tetraploid information can improve the selection coincidence. Also, to improve prediction ability, other different strategies were recommended in the literature as using multi-trait models (Guo et al. 2014). For our dataset, the MTM approach improved significantly the prediction values, manly using MTM-VS2 that also provided the most significant genetic gains (Table 2). INDG and MTM-VS1 had similar prediction abilities and genetic gains.

The *Urochloa* breeding program from the *Empresa Brasileira de Pesquisa Agropecuária* (EMBRAPA Beef-Cattle) followed the breeding scheme described by Jank et al. (2014). Each stage takes at least 2 years and adding 1 year for seed multiplication between them makes the whole process take 10–15 years. In a traditional program, the focus in the initial stage is to obtain new genotypes, in the intermediate stage is selection, and in the final stage is the recommendation of superior genotypes (Jank et al. 2014). Genomic prediction and selection tools can be applied to skip the stage 1, and the selected genotypes should be evaluated directly on stage 2. Consequently, at least 2 or 4 years will be reduced. Also, the breeder can use the information of costs and more accessible traits to be measured as a tool to decide which one will be evaluated in field trials and which will be predicted in the MTM approaches. The costs of GBS methods have the price around $35 per sample (Peng et al. 2017); this value is less than observed to phenotype tropical grasses which is around $180 per sample (personal communication). Therefore, by accounting the costs on the genetic gains equation, the advantages to using genomic prediction approaches would be even higher than the described above (Table 2). Although the challenges detected for SNP discovery and genotyping in this polyploid interspecific population, noticeable progress has been developed to help the application of genomic tools in polyploids for computational (Serang et al. 2012; Schmitz Carley et al. 2017) and statistics analysis (Endelman et al. 2018). Finally, genomic prediction should be used in forages breeding programs to reduce the time and the costs of recommending a new cultivar.

Acknowledgments National Council for Scientific and Technological Development (CNPq), Brazilian Agricultural Research Corporation (EMBRAPA) for financial support. National Center for High-Performance Processing in São Paulo (CENAPAD) and Center for High Throughput Computing (CHTC) for computing support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Author contributions FIM and RFN designed the study. SCLB and CBdoV conducted the field experiment and collected the phenotypic data. KGXM performed the DNA extraction. FIM performed the SNP calling and filtering. FIM and FCA performed the data analyses and interpretation. FIM, FCA, JBE, and RFN wrote the paper. JBE provided analytical expertise and edited the manuscript. RFN supervised the whole study. All authors read and approved the final version of the manuscript for publication.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Alves FC, Granato ÍSC, Galli G, Lyra DH, Fritsche-Neto R, de los Campos G (2019) Bayesian analysis and prediction of hybrid performance. Plant Methods 15:14. https://doi.org/10.1186/s13007-019-0388-x

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. In: line. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Annicchiarico P, Nazzicari N, Li X, Wei Y, Pecetti L, Brummer EC (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

Bassil NV, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, Alperin ES, Rosyara UR, Koehorst-vanc Putten H, Monfort A, Sargent DJ, Amaya I, Denoyes B, Bianco L, van Dijk T, Pirani A, Iezzoni A, Main D, Peace C, Yang Y, Whitaker V, Verma S, Bellon L, Brew F, Herrera R, van de Weg E (2015) Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria × ananassa. BMC Genomics 16. https://doi.org/10.1186/s12864-015-1310-1

Bauer AM, Léon J (2008) Multiple-trait breeding values for parental selection in self-pollinating crops. Theor Appl Genet 116:235–242. https://doi.org/10.1007/s00122-007-0662-6

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, Jenkins J, Barry K, Lindquist E, Hellsten U, Deshpande S, Wang X, Wu X, Mitros T, Triplett J, Yang X, Ye C-Y, Mauro-Herrera M, Wang L, Li P, Sharma M, Sharma R, Ronald PC, Panaud O, Kellogg EA, Brutnell TP, Doust AN, Tuskan GA, Rokhsar D, Devos KM (2012) Reference genome sequence of the model plant Setaria. Nat Biotechnol 30(6):555-561. https://doi.org/10.1038/nbt.2196

Biazzi E, Nazzicari N, Pecetti L, Brummer EC, Palmonari A, Tava A, Annicchiarico P (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

Bouckaert RR, Frank E (2004) Evaluating the replicability of significance tests for comparing learning algorithms. In: Dai H, Srikant R, Zhang C (eds) Advances in knowledge discovery and data mining. PAKDD 2004. Lecture Notes in Computer Science, vol 3056. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24775-3_3

Bourke PM, Voorrips RE, Visser RGF, Maliepaard C (2018) Tools for genetic studies in experimental populations of polyploids. Front Plant Sci. https://doi.org/10.3389/fpls.2018.00513

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci 52:707–719. https://doi.org/10.2135/cropsci2011.06.0299

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R reference manual mixed models for S language environments. Train. Ser. QE02001 149

Combs E, Bernardo R (2013) Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. Plant Genome 6:0. https://doi.org/10.3835/plantgenome2012.11.0030

Crossa J, Pérez-Rodriguez P, Cuevas J, et al (2017) Genomic selection in plant breeding: methods, models, and perspectives

Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193:347–365

De Almeida Filho JE, Guimarães JFR, E Silva FF et al (2016) The contribution of dominance to phenotype prediction in a pine breeding and simulated population. Heredity (Edinb) 117:33–41. https://doi.org/10.1038/hdy.2016.23

de Bem Oliveira I, Resende MF, Ferrao F, et al (2018) Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. bioRxiv 432179. doi: https://doi.org/10.1101/432179

Figueiredo UJ, Nunes JAR, do Valle CB (2012) Estimation of genetic parameters and selection of Brachiaria humidicola progenies using a selection index. Crop Breed Appl Biotechnol 12:237–244. https://doi.org/10.1590/S1984-70332012000400002

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345. https://doi.org/10.1534/genetics.112.143313

Depristo MA, Banks E, Poplin R et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–501. https://doi.org/10.1038/ng.806

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592–601

Dias KODG, Gezan SA, Guimarães CT, et al (2018) Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity (Edinb). 1–14. https://doi.org/10.1038/s41437-018-0053-6

Dos Santos JPR, De Castro Vasconcellos RC, Pires LPM et al (2016) Inclusion of dominance effects in the multivariate GBLUP model. PLoS One 11:e0152045. https://doi.org/10.1371/journal.pone.0152045

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. https://doi.org/10.1371/journal.pone.0019379

Enciso-Rodriguez F, Douches D, Lopez-Cruz M, et al (2018) Genomic selection for late blight and common scab resistance in tetraploid potato (*Solanum tuberosum*). G3: genes|genomes|genetics g3.200273.2018. doi: https://doi.org/10.1534/g3.118.200273

Endelman JB, Carley CAS, Bethke PC, Coombs JJ, Clough ME, da Silva WL, de Jong WS, Douches DS, Frederick CM, Haynes KG, Holm DG, Miller JC, Muñoz PR, Navarro FM, Novy RG, Palta JP, Porter GA, Rak KT, Sathuvalli VR, Thompson AL, Yencho GC (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

Euclides VPB, Montagner DB, Barbosa RA, Valle CB, Nantes NN (2016) Animal performance and sward characteristics of two cultivars of Brachiaria brizantha (BRS Paiagu{á}s and BRS Piat{ã}). Rev Bras Zootec 45:85–92

Fernandes SB, Dias KOG, Ferreira DF, Brown PJ (2018) Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor Appl Genet 131:747–755. https://doi.org/10.1007/s00122-017-3033-y

Ferrão LFV, Benevenuto J, Oliveira I de B et al (2018) Insights into the genetic basis of blueberry fruit-related traits using diploid and polyploid models in a GWAS context. Front Ecol Evol 6:107. https://doi.org/10.3389/fevo.2018.00107

Gouy M, Rousselle Y, Bastianelli D, Lecomte P, Bonnal L, Roques D, Efile JC, Rocher S, Daugrois J, Toubi L, Nabeneza S, Hervouet C, Telismart H, Denis M, Thong-Chane A, Glaszmann JC, Hoarau JY, Nibouche S, Costet L (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

Guo Z, Tucker DM, Lu J, Kishore V, Gay G (2012) Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor Appl Genet 124:261–275. https://doi.org/10.1007/s00122-011-1702-9

Guo G, Zhao F, Wang Y, Zhang Y, du L, Su G (2014) Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet 15:30. https://doi.org/10.1186/1471-2156-15-30

Hackett CA, McLean K, Bryan GJ (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

Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res (Camb) 91:47–60. https://doi.org/10.1017/S0016672308009981

Hayes BJ, Cogan NOI, Pembleton LW, Goddard ME, Wang J, Spangenberg GC, Forster JW (2013) Prospects for genomic selection in forage plant species. Plant Breed 132:133–143. https://doi.org/10.1111/pbr.12037

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49(1):12

Jank L, Barrios SC, do Valle CB, Simeão RM, Alves GF (2014) The value of improved pastures to Brazilian beef production. Crop Pasture Sci 65:1132–1137. https://doi.org/10.1071/CP13319

Jia Y, Jannink J-L (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522. https://doi.org/10.1534/genetics.112.144246

Jonas E, De Koning DJ (2013) Does genomic selection have a future in plant breeding? Trends Biotechnol 31:497–504

Keller-Grein G, Maass BL, Hanson J (1996) Natural variation in brachiaria and existing germplasm. In: Brachiaria: biology, agronomy and improvement. pp 16–42

Kempthorne O (1957) An introduction to genetic statistics

Lehermeier C, Schön CC, de los Campos G (2015) Assessment of genetic heterogeneity in structured plant populations using multivariate whole-genome regression models. Genetics 201:323–337. https://doi.org/10.1534/genetics.115.177394

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li X, Wei Y, Acharya A, Hansen JL, Crawford JL, Viands DR, Michaud R, Claessens A, Brummer EC (2015) Genomic prediction of biomass yield in two selection cycles of a tetraploid alfalfa breeding population. Plant Genome 8:0. https://doi.org/10.3835/plantgenome2014.12.0090

Liaw A, Wiener M (2002) Classification and Regression by randomForest R news 2:18–22. doi: https://doi.org/10.1177/154405910408300516

Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased prediction accuracy in wheat breeding trials using a marker x environment interaction genomic selection model. G3: Genes|Genomes|Genetics 5:569–82. doi: https://doi.org/10.1534/g3.114.016097

Lyra DH, de Freitas Mendonça L, Galli G, Alves FC, Granato ÍSC, Fritsche-Neto R (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. Mol Breed 37. https://doi.org/10.1007/s11032-017-0681-1

Marten GC, Brink GE, Buxton DR, Halgerson JL, Hornstein JS (1984) Near infrared reflectance spectroscopy analysis of forage quality in four legume species1.Crop Sci 24(6):1179. https://doi.org/10.2135/cropsci1984.0011183X002400060040x

Martin M (2011) Cutadapt removes adapter sequence from high-throughput sequencing reads. EMBnet.journal 17:10–12. https://doi.org/10.14806/ej.17.1.200

Matias FI, Barrios SCL, do Valle CB 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

Matias F, Barrios S, Lucas B, Do Valle C et al (2018) Contribution of additive and dominance effects on agronomical and nutritional traits, and multivariate selection on Urochloa spp. hybrids. Crop Sci. https://doi.org/10.2135/cropsci2018.04.0261

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

Mendes-Bonato AB, Pagliarini MS, Forli F, Borges do Valle C, de Oliveira Penteado MI (2002) Chromosome numbers and microsporogenesis in Brachiaria brizantha (Gramineae). Euphytica 125:419–425. https://doi.org/10.1023/A:1016026027724

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. doi: 11290733

Meuwissen T, Hayes B, Goddard M (2016) Genomic selection: a paradigm shift in animal breeding. Anim Front 6:6–14. https://doi.org/10.2527/af.2016-0002

Montagner DB, do Nascimento Júnior D, Sousa BM de L et al (2012) Morphogenesis in guinea grass pastures under rotational grazing strategies. Rev Bras Zootec 41:883–888

Nyine M, Uwimana B, Blavet N, Hřibová E, Vanrespaille H, Batte M, Akech V, Brown A, Lorenzen J, Swennen R, Doležel J (2018) Genomic prediction in a multiploid crop: genotype by environment interaction and allele dosage effects on predictive ability in banana. Plant Genome 11:0. https://doi.org/10.3835/plantgenome2017.10.0090

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2006) The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res 35(Database):D883-D887. https://doi.org/10.1093/nar/gkl976

Peng Z, Fan W, Wang L, Paudel D, Leventini D, Tillman BL, Wang J (2017) Target enrichment sequencing in cultivated peanut (Arachis hypogaea L.) using probes designed from transcript sequences. Mol Gen Genomics 292:955–965. https://doi.org/10.1007/s00438-017-1327-z

Pérez-Rodríguez P, Gianola D, González-Camacho JM, Crossa J, Manès Y, Dreisigacker S (2012) Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3: Genes, Genomes, Genetics 2:1595–1605. https://doi.org/10.1534/g3.112.003665

Pérez P, de los Campos G (2013) BGLR: a statistical package for whole genome regression and prediction. R package version 1:2. http://genomics.cimmyt.org/BGLR-extdoc.pdf

Resende RMS, Casler MD, de Resende MDV et al (2014) Genomic selection in forage breeding: accuracy and methods. Crop Sci 54:143–156. https://doi.org/10.2135/cropsci2013.05.0353

Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome 9:1–10. https://doi.org/10.3835/plantgenome2015.08.0073

Schmitz Carley CA, Coombs JJ, Douches DS, Bethke PC, Palta JP, Novy RG, Endelman JB (2017) Automated tetraploid genotype calling by hierarchical clustering. Theor Appl Genet 130:717–726. https://doi.org/10.1007/s00122-016-2845-5

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112-1115. https://doi.org/10.1126/science.1178534

Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y (2016) Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theor Appl Genet 129:273–287. https://doi.org/10.1007/s00122-015-2626-6

Serang O, Mollinari M, Garcia AAF (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

Sharma SK, MacKenzie K, McLean K, Dale F, Daniels S, Bryan GJ (2018) Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato G3 (Bethesda) g3.200377.2018. doi: https://doi.org/10.1534/g3.118.200377

Song J, Yang X, Resende MFR, Neves LG, Todd J, Zhang J, Comstock JC, Wang J (2016) Natural allelic variations in highly polyploidy Saccharum Complex. Front Plant Sci 7. https://doi.org/10.3389/fpls.2016.00804

Souza MB e, Cuevas J, Couto EG de O, et al (2017) Genomic-enabled prediction in maize using Kernel models with genotype × environment interaction. G3: Genes|Genomes|Genetics g3.117.042341. doi: https://doi.org/10.1534/g3.117.042341

Sverrisdóttir E, Byrne S, Sundmark EHR, Johnsen HØ, Kirk HG, Asp T, Janss L, Nielsen KL (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

Tan B, Grattapaglia D, Wu HX, Ingvarsson PK (2018) Genomic relationships reveal significant dominance effects for growth in hybrid Eucalyptus. Plant Sci 267:84–93. https://doi.org/10.1016/j.plantsci.2017.11.011

Uitdewilligen JGAML, Wolters AMA, D'hoop BB et al (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One. https://doi.org/10.1371/journal.pone.0062355

Vitezica ZG, Varona L, Legarra A (2013) On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics 195:1223–1230. https://doi.org/10.1534/genetics.113.155176

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

Worthington M, Heffelfinger C, Bernal D, Quintero C, Zapata YP, Perez JG, de Vega J, Miles J, Dellaporta S, Tohme J (2016) A parthenogenesis gene candidate

and evidence for segmental allopolyploidy in apomictic Brachiaria decumbens. Genetics 203:1117–1132. https://doi.org/10.1534/genetics.116.190314

You Q, Yang X, Peng Z, Xu L, Wang J (2018) Development and applications of a high throughput genotyping tool for polyploid crops: single nucleotide polymorphism (SNP) array. Front Plant Sci. https://doi.org/10.3389/fpls.2018.00104

Zingaretti L, Monfort A, Pérez-Enciso M (2018) pSBVB: a versatile simulation tool to evaluate genomic selection in polyploid species. G3: Genes|Genomes|Genetics g3.200942.2018. doi: https://doi.org/10.1534/g3.118.200942