Published August 15, 2019

RESEARCH

# Association Mapping Considering Allele Dosage: An Example of Forage Traits in an Interspecific Segmental Allotetraploid *Urochloa* spp. Panel

Filipe Inácio Matias,* Miriam Suzane Vidotti, Karem Guimarães Xavier Meireles, Sanzio Carvalho Lima Barrios, Cacilda Borges do Valle, Cari A. Schmitz Carley, and Roberto Fritsche-Neto

## ABSTRACT

The breeding process in tropical segmental allopolyploid forage *Urochloa* is challenging due to the complex genetic control of the traits. Knowledge about genes associated with forage traits, expressed in the different cutting seasons, are extremely useful to support breeding programs and development of new cultivars. Thus, the aims of our study were (i) to identify genomic regions related to forage traits through genome-wide association studies (GWAS), and (ii) to verify the influence of allele dosage on these results. A panel of 272 genotypes of *Urochloa* spp. [*U. brizantha* (Hoscht. ex A. Rich.) R. Webster × *U. ruziziensis* (Hoscht. ex A. Rich.) R. Webster] was evaluated in both the wet and dry seasons. The GWAS analyses were performed with 26,535 single nucleotide polymorphisms (SNPs) obtained by genotyping-by-sequencing (GBS) using diploid and tetraploid allele dosage configurations. Furthermore, we evaluated scenarios including additive, dominance, and epistatic effects. Seven candidate genomic regions associated with the main forage traits of *Urochloa* spp. were identified. The importance of the diploid and tetraploid molecular configuration in GWAS analyses for segmental allopolyploid species was demonstrated to identify the genomic behavior of important regions. Results demonstrated that it is possible to identify the same regions using both ploidy configurations; however, in some cases, the allele substitution effect can be biased mainly for regions with dominance and epistatic effects. Finally, this study contributes to the understanding of genetic control of tropical forage traits and genomics to accelerate the selection and reduce the cost to release new cultivars.

F.I. Matias, M.S. Vidotti, and R. Fritsche-Neto, Genetics Dep., "Luiz de Queiroz" College of Agriculture, Univ. of São Paulo, Av. Pádua Dias, 11, C. P. 9, 13.418-900, Piracicaba, São Paulo, Brazil; K.G.X. Meireles, S.C.L. Barrios, and C.B. do Valle, Embrapa Beef Cattle, Av. Rádio Maia, 830, Zona Rural, 79.106-550, Campo Grande, Mato Grosso do Sul, Brazil; C.A.S. Carley, Aardevo B.V, 5369 W Irving St. Boise, ID 83706, USA. Received 23 Mar. 2019. Accepted 21 June 2019. *Corresponding author (filipematias23@gmail.com). Assigned to Associate Editor Steve Larson.

**Abbreviations:** CP, crude protein; FGW, field green weight; GATK, Genome Analysis Toolkit; GBS, genotyping-by-sequencing; GWAS, genome-wide association study; IVD, in vitro organic matter digestibility; LIG, lignin in sulfuric acid; MAF, minor allele frequency; NDF, neutral detergent fiber; NIRS, near-infrared reflectance spectroscopy; PCA, principal component analysis; REG, regrowth capacity score; SNP, single nucleotide polymorphism.

Climate change and the decreasing availability of land for livestock are significant challenges to overcome the increasing demand for animal protein (Tilman and Clark, 2014; Grandin, 2015; Ramankutty et al., 2018; Springmann et al., 2018). In tropical regions, native or cultivated pastures constitute the most cost and environmentally effective form to feed cattle (Euclides et al., 2016; Henchion et al., 2017). However, for better animal performance in pastures, improvements should be made in pasture management, animal genetics, and forage genetics (better quality and adaptation to soils, climate, and pests). For instance, forage breeders have selected for typical plant traits such as biomass production, canopy size and structure, disease and insect resistances, forage quality and plant regrowth capacity, and ease of consumption (Hayes et al., 2013; Jank et al., 2014). In addition, forage breeding programs deal with indirect targets since the final product is not the plant performance but an animal product, such as milk or meat derived from the forage.

Published in Crop Sci. 59:2062–2076 (2019).
doi: 10.2135/cropsci2019.03.0185

*Urochloa brizantha* (Hoscht. ex A. Rich.) R. Webster, *U. decumbens* (Stapf) R. Webster, and *U. ruziziensis* (Hoscht. ex A. Rich.) R. Webster are extensively cultivated as pasture in tropical regions (Jank et al., 2014). The available cultivars of *U. brizantha* and *U. decumbens* are allotetraploid segmental and apomictic, whereas cultivars of *U. ruziziensis* are diploid and with sexual reproduction. In 1981, *U. ruziziensis* was artificially tetraploidized using colchicine and afterward used as a sexual genitor (pollen receptor or female genitor) in crosses with *U. brizantha* and *U. decumbens* (pollen donor or male genitor) to develop interspecific hybrids (Swenne et al., 1981; Lutts et al., 1991; do Valle and Pagliarini, 2009). Their purpose was mostly to identify hybrids with the nutritional quality of *U. ruziziensis* and the agronomic performance of *U. brizantha* and *U. decumbens*.

Interspecific hybridization in *Urochloa* spp. is the main strategy to develop new cultivars. Apomictic plants are used as a pollen donor in crosses with sexual tetraploidized plants (Fig. 1a). Hybrids are evaluated as single plants under field trials to select ∼10% of the best-performing individuals, which are subsequently evaluated with at least four replicates (Fig. 1a and 1b). The plots inside the replicate are normally composed of five vegetative clones of each selected hybrid (Fig. 1b). At the end of the pipeline, one or two hybrids are selected and evaluated for animal performance in several locations (Fig. 1c and 1d) to develop a new apomictic cultivar (Fig. 1e). For thorough descriptions of breeding schemes, see Barrios et al. (2013) and Worthington and Miles (2015). Each step described above usually is evaluated for at least 2 yr in the Cerrado biome of Brazil (3–22° S latitude, 39–65° W longitude); hence, the time to develop an *Urochloa* spp. cultivar is approximately 10 to 15 yr.

The breeding process could be accelerated through the development of genomic tools to improve selection efficiency. New genotyping approaches have generated a massive volume of genomic information for different species of animals and plants (Uitdewilligen et al., 2013; Zargar et al., 2015). These tools also are not restrictive, and species without a reference genome can be evaluated and the available genome of related species can be used to discover variant nucleotides in the target population (He et al., 2014). The molecular markers discovered can be used in genomic prediction, genome-wide association studies (GWAS), and marker-assisted selection approaches to select the most promising hybrids at the seedling stage and also to understand the genetic control of important traits (He et al., 2014; Crossa et al., 2017; Bourke et al., 2018). In tropical forages, markers associated with agronomical and nutritional traits could be used to identify genes underlying phenotypes.

In polyploid species, GWAS is more challenging than diploid species, mainly due to the number of genotype classes and gene action (additive, dominance, and epistasis) which, until recently, lacked appropriate analysis methods (Rosyara et al., 2016). Consequently, GWAS in polyploids is a relatively new subject and is predominantly applied by disregarding the allele dosage and using diploid models and software (Sun et al., 2016; Mourad et al., 2018). However, little is known about the consequences of using these models on the GWAS results compared with the use of adequate allele dosage, especially considering *Urochloa* spp. Ferrão et al. (2018) performing GWAS analysis in autotetraploid blueberries (*Vaccinium corymbosum* L.) verified that the importance of tetraploid models varies with traits and the use of diploid models has hindered the detection of single nucleotide polymorphism (SNP)–trait associations in this autotetraploid species. In turn, Inostroza et al. (2018) only detected significant SNP–trait association in the allotetraploid clover (*Trifolium repens* L.) when GWAS analyses were performed with tetraploid genetic models, showing the importance of considering the correct allele dosage. Additionally, improvements in genomic studies in polyploids were provided by the use of new methods to estimate the relationship matrix in tetraploids as described by Endelman et al. (2018) and Amadeu et al. (2016), followed by the development of new software to estimate the genotype call in tetraploids such as ClusterCall (Schmitz Carley et al., 2017) and updog (Gerard et al., 2018).

Genome-wide association studies for *U. brizantha* and *U. decumbens* are even harder due to the species being segmental allotetraploids, one with genome allotetraploid behavior and the other with autotetraploid behavior (Worthington et al., 2016; de Paula et al., 2017). These species have chromosomes with preferential pairing and fully homologous chromosomes pairing at random (Sybenga, 1996). In this case, genomic studies can be performed using markers with diploid and tetraploid allele dosage. The genomic constitution of *U. ruziziensis*, *U. decumbens*, and *U. brizantha* is classified as $B_2B_2$, $B_1B_1B_2B_2$, and $BBB_1B_1$, respectively (de Paula et al., 2017). However, although these genomes have been considered homeologous, less affinity is observed between genomes B and $B_2$, and thus genes segregating as a diploid is possible in hybrids between *U. ruziziensis* and *U. brizantha*.

Despite the noteworthy importance of *Urochloa* spp. for livestock in tropical regions, to our knowledge, there are no GWAS using SNP markers performed within this genus. This study discusses the influence of the allele dosage on the genomic resolution of association mapping analysis in some *Urochloa* species, and thus the aims were (i) to identify genomic regions related to forage traits performance in segmental allotetraploid *Urochloa* hybrids through GWAS using SNP markers from genotyping-by-sequencing (GBS), and (ii) to verify the influence of diploid and tetraploid markers allele dosage configuration on these results.

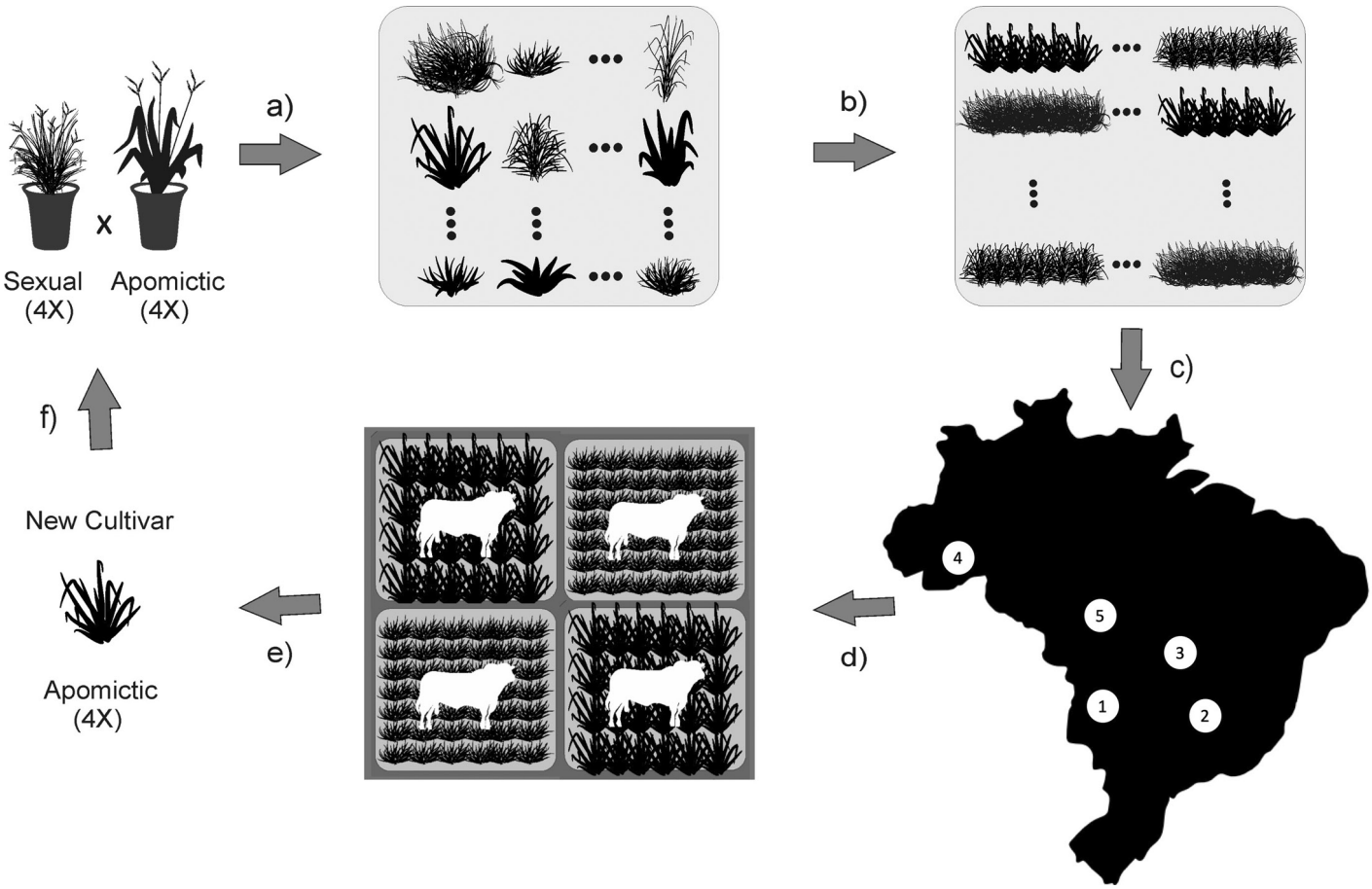


Fig. 1. *Urochloa* spp. breeding program scheme to develop apomictic cultivars. (a) Hybridization: single cross between commercial apomictic cultivars (pollen donor or male genitor) and synthetic sexual parents (pollen receptor or female genitor) in Embrapa Beff Cattle, Campo Grande, Mato Grosso do Sul, Brazil (2 yr = crossing and seed dormancy); (b) Stage 1: progeny evaluation based on one plant per plot (2 yr = two dry and wet seasons); (c) Stage 2: hybrids selected in Stage 1 are evaluated in trials with at least four replicates and five hybrids clones per plot (2 yr = two dry and wet seasons); (d) Stage 3: Regional multi-trial experiments considering the selected hybrids from Stage 2 (at least 2 yr)—this step is conducted with partners of different units of Embrapa in Brazil ([1] Embrapa Beff Cattle, [2] Embrapa Dairy Cattle, [3] Embrapa Cerrados, [4] Embrapa Acre, and [5] Embrapa Agrossilvipastoril); (e) Stage 4: 1 or 2 hybrids selected in Stage 3 are evaluated for animal performance (2 yr = two dry and wet seasons); (f) seed production and release of a new apomictic cultivar, which may also enter the breeding program as a male genitor (2–5 yr).

# MATERIALS AND METHODS

## Plant Material and Phenotyping

A set of 263 tetraploid interspecific hybrids was obtained by crossing *U. brizantha* (pollen donor: apomictic segmental allotetraploid) and hybrids of *U. ruziziensis* × *U. brizantha* (pollen receptor: sexual tetraploids), from the forage breeding program of Brazilian Agricultural Research Corporation (Embrapa Beef Cattle), Campo Grande, Mato Grosso do Sul, Brazil (20°27′ S, 54°57′ W). This population was evaluated in the field using an incomplete block design during seven cuttings, from 2014 to 2015. Nine commercial apomictic cultivars were used as checks: *U. brizantha* cv. 'Marandu', *U. brizantha* cv. 'Paiaguás', *U. brizantha* cv. 'B140', *U. decumbens* cv. 'Basilisk', and the interspecific commercial hybrid 'Mulato II'. The sexual elite tetraploid hybrids from Embrapa genetic bank (BS9, BS15, 336-T1, and 336-T2) were also used as checks. Cuts 1, 4, 5, and 6 were performed in the wet season, whereas Cuts 2, 3, and 7 in the dry season. The full experimental design and biological materials (hybrids and checks) were previously described by Matias et al. (2018). This population was at the breeding stage highlighted at Fig. 1a, and thus each hybrid was available as a single plant, and the environmental influence was evaluated by the check replicates. Each hybrid was individually evaluated for the two groups of traits below.

### *Agronomic Traits*

Field green weight (FGW, kg ha$^{-1}$) was evaluated by cutting the plant ~10 cm above the soil surface and weighing the green matter in the field with a field scale. Final plant regrowth capacity score (REG) was estimated according to the methodology described by de Figueiredo et al. (2012) 7 d after cutting, obtained by the combination between scores for the density of regrown tillers and regrowth speed.

### *Nutritional Traits*

Approximately 200 g of green forage were dried, ground, and analyzed with near-infrared reflectance spectroscopy (NIRS) (Marten et al., 1989). The NIRS calibration was previously performed by comparing the results obtained in the chemical analyzes vs. the spectrum read obtained from the same sample evaluated using NIRS (data not shown). This process was used

to estimate the sample percentage of crude protein (CP), in vitro organic matter digestibility (IVD), neutral detergent fiber (NDF), and lignin in sulfuric acid (LIG) for the samples collected during the Cuttings 3 and 4.

## Statistical Analyses

### Season Effect Estimation

The significance of season effect was estimated by a fixed model approach to verify the difference between dry and rainy seasons. Fixed effects were tested by Wald *F* test supported by the ASreml-R package (Butler et al., 2009). For that, a joint analysis using incomplete block design was performed including all genotypes (9 checks and 263 hybrids) by the following equation:

$$\mathbf{y}_{acdh} = \boldsymbol{\mu} + \mathbf{p}_g + \mathbf{r}_h + \mathbf{q}_{h(c)} + \mathbf{s}_{a\times h} + \mathbf{t}_b + \boldsymbol{\varepsilon}_{acbh} \qquad [1]$$

where $\mathbf{y}$ is the vector for phenotypic data; $\boldsymbol{\mu}$ is the vector for the overall mean; $\mathbf{p}$ is the vector of genotype, with $g = \{1, 2, \ldots, 272\}$; $\mathbf{r}$ is the vector indicating the season effect, with $h = \{$wet or dry$\}$; $\mathbf{q}$ is the vector of cut nested into season effect, with $c = \{1, 4, 5, 6\}$ for wet season, and $c = \{2, 3, 7\}$ for dry season; $\mathbf{s}$ is the vector of interaction between genotypes and season effects; $\mathbf{t}$ is the vector of the block effect, with $b = \{1, 2, \ldots, 10\}$; and $\boldsymbol{\varepsilon}$ is the residual, with $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma^2_{\varepsilon})$ where $\mathbf{I}$ is the identity matrix and $\sigma^2_{\varepsilon}$ is the residual variance component.

### Genetic Effects Estimation

Once the significance of season was identified by the Eq. [1], the phenotypic record was adjusted to be used in the GWAS. The genetic effects were estimated for individual hybrids using fixed model with annual, wet, and dry season considered separately, following the incomplete block design equation (Eq. [2]):

$$\mathbf{y}_{bcd} = \boldsymbol{\mu} + \mathbf{p}_g + \mathbf{q}_c + \mathbf{s}_b + \mathbf{u}_{g\times c} + \boldsymbol{\varepsilon}_{gcb} \qquad [2]$$

where $\mathbf{y}$ is the vector for phenotypic data; $\boldsymbol{\mu}$ is the vector for the overall mean; $\mathbf{p}$ is the vector of genotypes (9 checks and 263 hybrids), with $g = \{1, 2, \ldots, 272\}$; $\mathbf{q}$ is the vector of cut effect, with $c = \{1, 2, \ldots, 7\}$ for annual, $c = \{1, 4, 5, 6\}$ for wet season, and $c = \{1, 3, 7\}$ for dry season; $\mathbf{s}$ is the vector of block effect, with $b = \{1, 2, \ldots, 10\}$; $\mathbf{u}$ is the vector of genotypes × cut interaction effect; and $\boldsymbol{\varepsilon}$ is the residual effect with $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma^2_{\varepsilon})$ where $\sigma^2_{\varepsilon}$ is the residual variance component. Equation [2] was also run with random effect of genotypes, with $\mathbf{p} \sim N(0, \mathbf{I}\sigma^2_g)$, to estimate the heritability following the equation $H^2 = \sigma^2_g / [\sigma^2_g + (\sigma^2_{g\times c}/c) + (\sigma^2_{\varepsilon}/cb)]$, where $\sigma^2_g$ is the genetic variance of genotypes, $\sigma^2_{g\times c}$ in the interaction between genotypes and cuttings, $c$ is the number of cuttings, and $b$ is the number of replicates per blocks (Fristche-Neto et al., 2018).

## Genotyping

Fresh leaf of all 272 genotypes (9 checks and 263 hybrids) was used for DNA extraction with the Qiagen DNeasy kit, and samples were genotyped by sequencing using the ApeKI enzyme (Elshire et al., 2011) and Illumina Hi-Seq 2500 platform with 1 × 100 bp reads. The Genome Analysis Toolkit (GATK) pipeline (McKenna et al., 2010; DePristo et al., 2011) was used to discover and call SNP markers. The genomic data were evaluated by GATK using ploidy = 4 for genotype calling at the tetraploid level. The diploid genotype calls were based on the genotype likelihoods to evaluate the data at the diploid level. For instance, the posterior probability to call Genotypes 1, 2, and 3 from tetraploid configuration were summed to estimate the posterior probability to call Genotype 1 in the diploid configuration [i.e., $p(Aa) = p(Aaaa) + p(AAaa) + p(AAAa)$]. Additionally, $p(aa) = p(aaaa)$ is the posterior probability of Genotype 0 in the tetraploid configuration to be the Genotype 0 in the diploid configuration, and $p(AA) = p(AAAA)$ is the posterior probability of Genotype 4 (tetraploid) to be Genotype 2 (diploid). *Urochloa* does not have a complete genome available; in this case, sequencing reads were aligned with five different genomes: *Setaria virides* (L.) P. Beauv. (*Sv*) (DOE-JGI, 2018a), *Setaria italic* (L.) P. Beauv. (*Si*) (Bennetzen et al., 2012), *Sorghum bicolor* (L.) Moench. (*Sb*) (DOE-JGI, 2018b), *Oryza sativa* L. (*Os*) (Ouyang et al., 2006), *Zea mays* L. (*Zm*) (Schnable et al., 2009), and the *Urochloa* mock reference (*Um*) (Matias et al., 2019). Default alignment parameters were used in the software Burrows-Wheeler Alignment tool (Li and Durbin, 2009), SAMtools (Li et al., 2009; Li, 2011) and Picard (http://broadinstitute.github.io/picard/).

*Urochloa* spp. as described above are segmental allotetraploid species with part of the genome with a tetraploid behavior and part of the genome with a diploid behavior. For this reason, high-quality filtering was applied assuming the diploid level during the quality control process to increase the probability to select markers segregating as diploid species with disomic inheritance. The SNP markers were filtered by minor allele frequency (MAF) ≥1%, missing data per marker ≤50%, minimum read depth per sample ≥8, and general genotype quality score ≥10 on the Phred scale. The average depth was 18 reads per marker by sample. Remaining missing data were imputed using the Random Forest package (Liaw and Wiener, 2002), where all markers with $r^2 \geq 0.1$ were used as predictors, and 300 trees were considered, following the recommendation of Matias et al. (2019). Consequently, we selected 26,535 SNPs with diploid and tetraploid dosage configurations. Three possible genotypes were assigned as diploid (*aa*, *aA*, and *AA*), whereas five possible genotypes were used for tetraploid (*aaaa*, *aaaA*, *aaAA*, *aAAA*, or *AAAA*). The SNP matrix for both ploidies was used to build the additive relationship matrix described by VanRaden (2008) at diploid level ($\mathbf{G}_{\text{Dip}}$), and Vitezica et al. (2013) at tetraploid level ($\mathbf{G}_{\text{tetra}}$), according to the equations

$$\mathbf{W}_{\text{Dip}} = (\mathbf{X}_{\text{Dip}} - 2p_i)$$

$$\mathbf{G}_{\text{Dip}} = \frac{\mathbf{W}_{\text{Dip}}\mathbf{W}'_{\text{Dip}}}{\sum 2p_i(1-p_i)}$$

$$\mathbf{W}_{\text{Tetra}} = (\mathbf{X}_{\text{Tetra}} - 4p_i)$$

$$\mathbf{G}_{\text{Tetra}} = \frac{\mathbf{W}_{\text{Tetra}}\mathbf{W}'_{\text{Tetra}}}{\sum 4p_i(1-p_i)}$$

where $p_i$ is the reference allele frequency, and $\mathbf{X}$ is the allele dosage matrix with genotypes denoted 0 to 2 for diploids and 0 to 4 for tetraploids. A graphic representation of the kinship matrix was built using the R package superheat (Barter and Yu, 2018).

## Genome-Wide Association Studies

Adjusted means of hybrids from Eq. [2] were used to perform the GWAS analyses of traits under annual, wet, and dry seasons. These analyses were inspired by Pajerowska-Mukhtar et al. (2009) with tetraploid potato (*Solanum tuberosum* L.) and association studies with field resistance to diseases. Hybrid genotypes were parameterized by the dosage of reference allele as nuliplex (0 = *aa*), simplex (1 = *Aa*) or duplex (2 = *AA*) for the diploid data, and the nuliplex (0= *aaaa*), simplex (1 = *Aaaa*), duplex (2 = *AAaa*), triplex (3 = *AAAa*), and quadruplex (4 = *AAAA*) for the tetraploid data. The GWAS linear mixed model **Q** + **K** described below was proposed by Yu et al. (2006) and adapted to support each ploidy based on general, additive, simplex dominant, and duplex dominant gene actions, available in the R package GWASpoly (Rosyara et al., 2016).

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{ZSt} + \mathbf{Zu} + \varepsilon \qquad [3]$$

where $\mathbf{y}$ is the vector of the adjusted phenotypes; $\beta$ is the first two principal components from molecular data used as fixed effect covariates (population structure); $\mathbf{t}$ is the vector of SNP effects; $\mathbf{u}$ is the vector of polygenic effects with $\mathbf{u} \sim N(0, \mathbf{K}\sigma^2_g)$ where $\sigma^2_g$ is the variance component of genotypes and $\mathbf{K}$ is the relationship matrix; $\mathbf{G}$ is described above ($\mathbf{G}_{Dip}$ and $\mathbf{G}_{Tetra}$); and $\varepsilon$ is the residual effect vector with $\varepsilon \sim N(0, \mathbf{I}\sigma^2_\varepsilon)$ where $\mathbf{I}$ is the identity matrix and $\sigma^2_\varepsilon$ is the variance component of error. The incidence matrix $\mathbf{X}$ is the incidence matrix accounting for the fixed effects; $\mathbf{Z}$ is the incidence matrix that maps genotypes to observations; $\mathbf{S}$ is the incidence matrix of genetic models: epistatic (general), additive, and dominance. As described by Rosyara et al. (2016), the general model uses each genotype class independently, and the SNP effect in each one is arbitrary. For the additive model, the effect was evaluated by dosage of the reference allele. Dominance models were simplex and duplex dominance. Simplex dominance model evaluates the hypothesis that homozygote genotype of the alternative allele (*aaaa*) has a different performance than others genotypes with at least one copy of the reference allele (*A*----). Duplex dominance model is specific for tetraploid configuration and evaluates if the duplex genotype (*AAaa*) has a similar performance as the nuliplex (*aaaa*) and the simplex genotype (*Aaaa*) or similar performance to triplex (*AAAa*) and tetraplex (*AAAA*) genotypes. False discovery rate was controlled by setting the significance threshold with Bonferroni's multiple testing correction methods with initial $p = 0.05$. The software *MapChart* (Voorrips, 2002) was used to illustrate chromosome regions with at least one significant SNP on GWAS analysis according to the reference genome.

## Gene Annotation

For significant SNPs found on the *Urochloa* mock reference (Matias et al., 2019), a sequence of 50 nucleotide positions from both sides of significant SNPs was selected from the respective reference and compared with the information available in genomic data banks using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990). For the significant SNPs found on the other references (annotated genomes), a window of 1 kb was evaluated to determined candidate genes. Gene function and homologies were evaluated using the NCBI platform (NCBI Resource Coordinators, 2017).

# RESULTS

## Agronomical and Nutritional Phenotypic Variation

There were significant effects of genotype and season (Supplemental Table S1) and genotype effect nested into each environmental condition (Table 1) for all traits. This indicates that at least one genotype had a differential performance between and within wet and dry seasons. The broad-sense heritability considering the annual data ranged from 0.46 to 0.81 for LIG and FGW, respectively. Field green weight showed a different trend when compared with REG, CP, IVD, NDF, and LIG. Field green weight presented a higher heritability in the wet season (77%) than in the dry season (53%), whereas all other traits followed the opposite trend—for instance, NDF showed a moderate heritability in the dry season (45%) and very low heritability in the wet season (27%).

As expected, average values for all traits (across growing seasons) were intermediate between those observed in the wet and dry season. For example, the average FGW yield was 1468.26 kg ha$^{-1}$, whereas its performance in the dry and wet season were 1062.26 and 1773.83 kg ha$^{-1}$, respectively (Table 1). Thus, genotypes showed a better performance for FGW, REG, and IVD during the wet season than during the dry season. In contrast, for CP, NDF, and LIG, there were no significant differences between seasons.

Correlations between traits changed as a function of season (dry and wet). For instance, a positive and significant correlation between FGW and REG was observed, ranging from 0.40 during the dry season to 0.21 during the wet season (Fig. 2, Supplemental Table S2). The same reduction of magnitude value between NDF and CP was observed, ranging from −0.48 (dry) to −0.25 (wet). Also, FGW and NDF, FGW and LIG, NDF and LIG, and IVD and NDF reduced the correlation during the wet season, though with low intensity. On the other hand, IVD and CP and IVD and LIG increased from 0.48 and −0.23 (dry season) to 0.51 and −0.33 (wet season), respectively.

## Population Structure and Diversity Analysis

For the diploid and tetraploid configurations, population structure was evaluated in terms of genotype frequencies, visual representation of the additive relationship matrix, and a biplot of the first two principal components of the marker data. The proportion of homozygous to heterozygous genotypes was slightly different between diploid and tetraploid configurations (Fig. 3a). In homozygote Genotypes 0 for both ploidies, the difference between diploid and tetraploid configurations was 0.75%. On the other hand, using the reference allele homozygote Genotype 2 for diploid (*AA*) and 4 for tetraploid (*AAAA*), the difference was 0.16%. The diploid heterozygote Genotype 1 (22.75%) was distributed among the three

Table 1. Wald test for fixed effects of genotype, broad-sense heritability ($H^2$), and general average ($\bar{x}$) of field green weight (FGW), regrowth capacity (REG), crude protein (CP), in vitro organic matter digestibility (IVD), neutral detergent fiber (NDF), and lignin in sulfuric acid (LIG)

| Season | Parameters | FGW | REG | CP | IVD | NDF | LIG |
|---|---|---|---|---|---|---|---|
| Annual | Genotype | 8181.20** | 5035.10** | 3374.10** | 2122.20** | 2267.70** | 1882.40** |
| | $H^2$ | 0.81 | 0.75 | 0.68 | 0.52 | 0.52 | 0.46 |
| | $\bar{x}$ | 1468.26 | 3.23 | 15.80 | 71.74 | 65.91 | 2.21 |
| Dry | Genotype | 4653.70** | 3999.70** | 2823.00** | 2702.00** | 1632.40** | 1765.20** |
| | $H^2$ | 0.53 | 0.68 | 0.55 | 0.53 | 0.45 | 0.49 |
| | $\bar{x}$ | 1062.26 | 3.56 | 15.72 | 74.69 | 65.06 | 2.20 |
| Wet | Genotype | 5683.30** | 2519.95** | 1641.00** | 1557.30** | 977.20** | 1287.95** |
| | $H^2$ | 0.77 | 0.52 | 0.47 | 0.39 | 0.27 | 0.32 |
| | $\bar{x}$ | 1773.83 | 2.98 | 15.85 | 68.80 | 66.80 | 2.21 |

** Significant at the 0.05 and 0.01 probability levels, respectively.

possible heterozygote tetraploid Genotypes 1 (12.41%), 2 (6.74%), and 3 (2.7%), totaling 21.85%. These differences of frequencies could represent genotyping errors on the edge of genotype probability, mainly caused by the read sampling during the genotyping step. For example, markers with Genotype 3 in tetraploid configuration has $p(AAAa) > p(AAAA)$; however, it can be Genotype 2 ($AA$) instead Genotype 1 ($Aa$) in diploid configuration due the low depth of allele $a$ being accounted as a mistake during the genotype call; in this case, $p(AA) > p(Aa)$.

The heat map of the kinship additive matrix showed differences between the diploid and tetraploid levels (Fig. 3b), and the clusters' shape and size were different according to the ploidy. In the diploid configuration, the number of clusters is easier to identify than in the tetraploid configuration, due to building smaller and more distinctive clusters. However, the use of tetraploid data organized the population in three greater groups. The biplot from the first two principal components in principal component analysis (PCA) explained 13.7% at the diploid level and 12.8% using tetraploid marker data (Fig. 3c). The cloud of points for both ploidies had a triangular shape, but with a different orientation. The genomic structure from the original diallelic cross of this population (Matias et al., 2018) can be identified by the PCA using diploid or tetraploid data configuration.

## GWAS Analyses for the Different Seasons

A complex interaction of genotype and environment was observed when different heritabilities and hybrids performance were observed for all traits across seasons (Table 1, Supplemental Table S1). Overall, additive and dominant models of GWAS were evaluated using markers at the diploid and tetraploid level and only the significant results were described. It was verified that each trait showed different responses regarding the combination between season, ploidy, and GWAS model.

Significant markers across all traits and all GWAS models are summarized in Table 2. We found a marker associated with REG, for the annual and dry season,

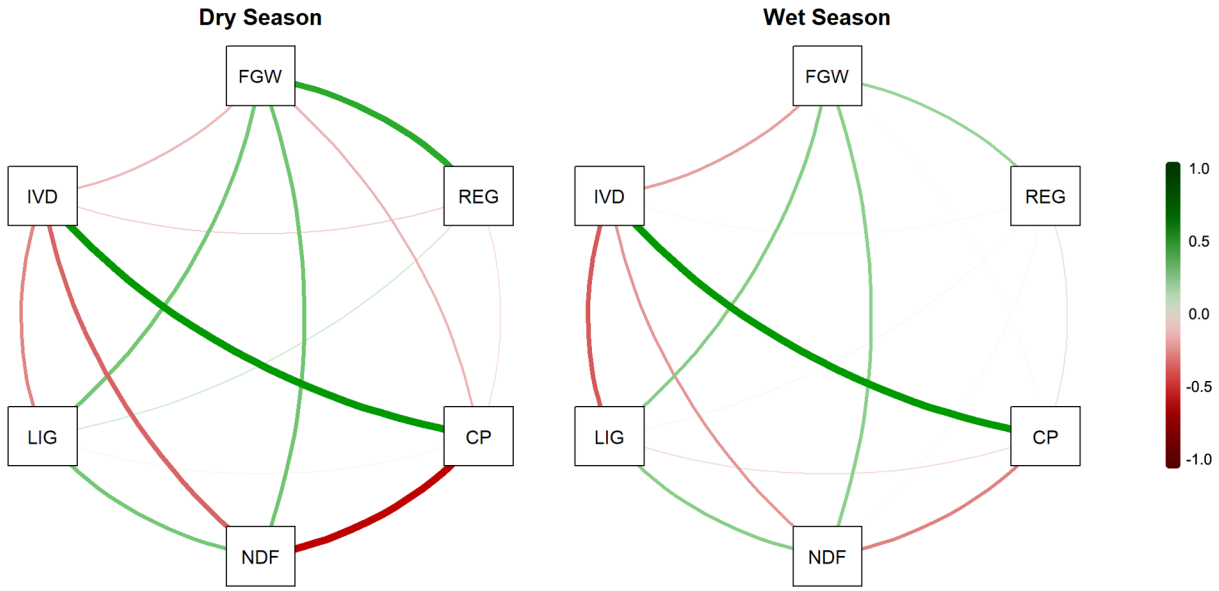


Fig. 2. Correlation network between field green weight (FGW), regrowth capacity (REG), crude protein (CP), in vitro organic matter digestibility (IVD), neutral detergent fiber (NDF), and lignin in sulfuric acid (LIG).

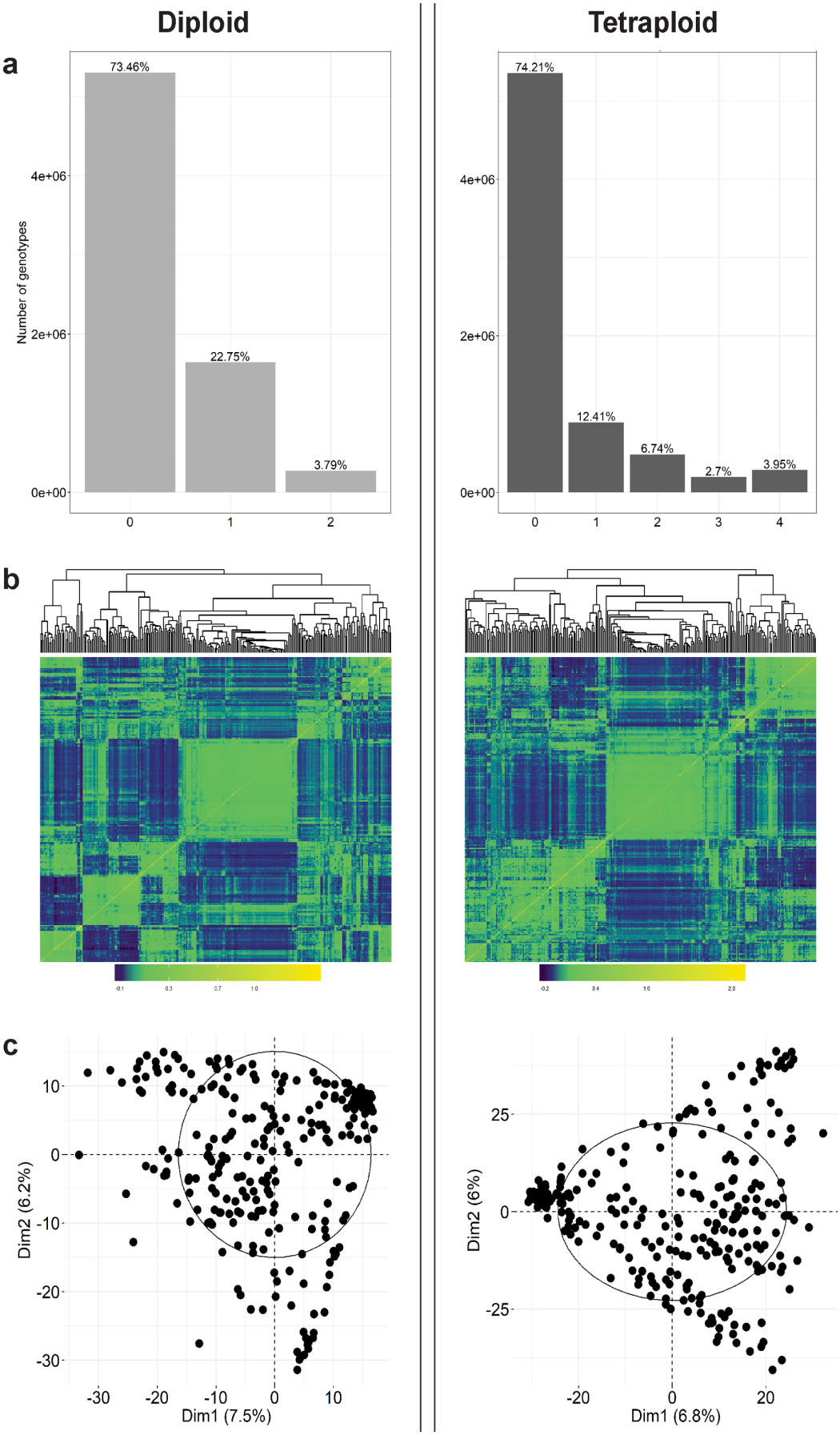


Fig. 3. Population structure and diversity analysis using diploid and tetraploid marker configuration. (A) The proportion of each class of genotype, (B) the kinship matrix heat map, and (C) the biplot from the first two principal components. Dim1, first principal component; Dim 2, second principal component.

**Table 2. Annotated genes from significant single nucleotide polymorphism (SNP) markers associated with *Urochloa* spp. forage traits identified by genome-wide association study (GWAS) analysis.**

| Season | Trait† | Ploidy | Model | PC‡ | Threshold | Reference§ | Chromosome | Position | Score | MAF¶ | Effect | Gene | Protein | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $-\log_{10}(p)$ | | | bp | $-\log_{10}(p)$ | | | | | |
| Annual | REG | Diploid | 1-dom-alt | 0 | 5.18 | Sv | Chr08 | 7,908,449 | 5.83 | 0.25 | 0.45 | *trnD-GUC* | tRNA-Asp | 0.14 |
| | | Tetraploid | 2-dom-alt | | 5.02 | | | | 5.27 | 0.25 | 0.42 | | | 0.15 |
| | NDF | Diploid | 1-dom-alt | 0 | 5.18 | Si | Scaffold_5 | 15,551,397 | 5.21 | 0.07 | −2.54 | *LOC101778276* | Exocyst complex component SEC15B | 0.07 |
| | | Tetraploid | 2-dom-ref | | 4.80 | Um | | 8,160,655 | 4.94 | 0.46 | 1.98 | | Aligned with *Triticum aestivum* chromosome 3B | 0.07 |
| | IVD | Tetraploid | general | 2 | 5.21 | Um | | 128,132 | 5.47 | 0.17 | NA# | | Aligned with *Triticum aestivum* chromosome 3B | 0.09 |
| Dry | FGW | Diploid | 1-dom-alt | 2 | 5.18 | Sb | Chr02 | 1,588,978 | 5.50 | 0.18 | −167.87 | *LOC8084285* | Uncharacterized protein | 0.05 |
| | | Tetraploid | | | 5.03 | | | | 5.86 | 0.17 | −184.60 | | | 0.06 |
| | REG | Diploid | 1-dom-alt | 0 | 5.18 | Sv | Chr08 | 7,908,449 | 5.80 | 0.25 | 0.50 | *trnD-GUC* | tRNA-Asp | 0.11 |
| | | Tetraploid | 2-dom-alt | | 5.02 | | | | 5.70 | 0.25 | 0.49 | | | 0.13 |
| | NDF | Tetraploid | general | 0 | 5.21 | Um | | 128,132 | 5.80 | 0.17 | NA | | Aligned with *Triticum aestivum* chromosome 3B | 0.11 |
| | IVD | Tetraploid | general | 2 | 5.44 | Um | | 91,613 | 7.09 | 0.02 | NA | | No significant similarity found | 0.14 |
| Wet | LIG | Diploid | general | 0 | 5.26 | Um | | 8,160,655 | 5.91 | 0.49 | NA | | Aligned with *Triticum aestivum* chromosome 3B | 0.11 |
| | | Tetraploid | | | 5.21 | | | | 5.64 | 0.46 | NA | | | 0.13 |
| | IVD | Tetraploid | 2-dom-ref | 2 | 4.80 | Um | | 3,259,930 | 4.99 | 0.41 | −1.96 | *LOC101780209* | Uncharacterized protein At3g52155 | 0.08 |
| | | | | | | Sv | Chr_08 | 7,908,449 | 5.97 | 0.25 | −3.11 | *trnD-GUC* | tRNA-Asp | 0.06 |

† REG, regrowth capacity; NDF, neutral detergent d=fiber; FGW, field green weight; IVD, in vitro organic matter digestibility; LIG, lignin in sulfuric acid.

‡ PC, principal component.

§ Sv, *Setaria virides*; Si, *Setaria italica*; Um, *Urochloa mock*; Sb, *Sorghum bicolor*.

¶ MAF, minor allele frequency.

# NA, no specific allele per effect.

mapped on chromosome 8 of *Setaria virides* at the position 7,908,449 bp. This SNP was annotated inside the gene *trnD-GUC*, corresponding to a synthesis of tRNA-Asp. For NDF, two markers for the annual performance and one for the dry season were identified. The first SNP for annual data was aligned with *Setaria italica* scaffold_5 at the position 15,551,397 bp, with a dominant negative effect (−2.54) for the alternative allele. Furthermore, it is near to the gene *LOC101778276*, which is related to the Exocyst Complex Component SEC15B. The second marker for annual and dry season was aligned with *Urochloa mock* (*Um*) reference at the positions 8,160,655 and 128,132 bp, respectively (Supplemental Fig. S1). These markers came from centroids in the *Um* genome reference (Matias et al., 2019) originated from the GBS-SNP-CROP approach (Melo et al., 2016). Although there were no candidate genes on these regions, the last part of both sequences showed the same final nucleotide sequence. In particular, this coincident part had homology with *Triticum aestivum* L. chromosome 3B, but no function was assigned.

Significant markers were found for IVD in all environmental conditions (Table 2). The marker *Um_128132*, mentioned above for NDF, was also significant for IVD in the annual period of evaluation. The *Um_91613* marker was significant for the dry season, but its MAF was very low (0.02), and no significant similarity was found. During the wet season, the dominant model revealed two significant markers. The first was aligned with *Urochloa mock* reference at the 3,259,930-bp position (MAF = 0.41). The region near this marker is similar to the gene *LOC101780209* (uncharacterized protein At3g52155). The latter was previously reported for regrowth capacity, *Sv_Chr08_7908449*, with a negative dominance effect, which reduces the IVD to about −3.38.

Only one marker was significant during dry season for FGW. It was noted in the *Sorghum bicolor* genome, chromosome 02, at position 1,588,978. The alternative allele of this SNP when present in the hybrid is dominant and associated with a reduction of 172.5 kg ha$^{-1}$. Furthermore, the gene *LOC8084285* was annotated in this region, which corresponds to an uncharacterized protein.

In general, three regions from genomes with annotation had significant SNPs aligned: chromosome 2 of *Sorghum bicolor*, chromosome 5 of *Setaria italica*, and chromosome 9 of *Setaria virides* (Supplemental Fig. S2). *Setaria italica* allowed more alignments for *Urochloa* spp., which were relatively well distributed across the genome, following what was observed by Matias et al. (2019).

## GWAS Analysis Using Diploid and Tetraploid Genomic Configuration

Concerning the GWAS models, no pattern was observed for ploidy level within seasons (Table 2). Only the simplex-dominance and general models allowed identification of significant markers associations using diploid configuration. Besides these two models, significant markers were also found using duplex-dominance model for tetraploid configuration.

As already mentioned, the marker *Sv_Chr08_7908449* was significant for REG in the annual data and dry season and also was significant using both diploid and tetraploid data configurations (Table 2, Fig. 4a). The model had an excellent fit, as can be observed in the quantile-quantile plot (Fig. 4b). Complete dominance effect was observed for the marker *Sv_Chr08_7908449* at diploid level, where at least one copy of allele (*A*-) was necessary to improve the REG from −1 to approximately −0.5 (Fig. 4c). However, when the ploidy was expanded to the tetraploid level, we found that only genotypes with three (*AAAa*) or four copies (*AAAA*) of this allele were responsible for improving this trait. This fact indicates that the allele substitution effect was biased when estimated with the diploidized marker data.

For LIG, the marker *Um_8160655* was identified by diploid and tetraploid configuration (Table 2, Fig. 5). The alternative allele exerts a dominant effect on the trait. In this case, the homozygous genotype for the alternative allele (*aa* and *aaaa*) and all the heterozygous (*Aa* and *A--a*) reduced the average value for this trait in approximately 0.5 to 0.6 in comparison with genotypes with homozygous genotype for reference allele (i.e., *AA* = 2 for diploid configuration and *AAAA* = 4 for tetraploid configuration). Thus, only the latter genotype showed the undesirable high levels of LIG for *Urochloa* spp. cultivars.

Markers with statistical significance should be evaluated carefully in segmental allotetraploids, once the real inheritance is not known for all genome regions. One example is the fact that some SNPs were significant using both marker configurations and others only using one specific configuration (Table 2). For example, all significant SNPs found for IVD were identified by the tetraploid configuration (dry, wet season, and annual), and no marker was significant using diploid configuration. Evaluating the marker *Um_91613*, no conclusions should be made, as this marker did not have enough representability of genotype classes in tetraploid configuration (classes "*AAaa*" and "*AAAA*" had only one observation and class "*AAAa*" had no observations). However, an opposite situation was observed for the marker *Um_3259930*. This marker showed many individuals in each class, highlighting that both *Urochloa* spp. genomes (BB and $B_2B_2$) are segregating (Supplemental Table S3).

# DISCUSSION

## Variability of Forage Traits during the Dry and Wet Seasons

Tropical forages are generally subjected to seasonal differences in environmental conditions favoring growth during

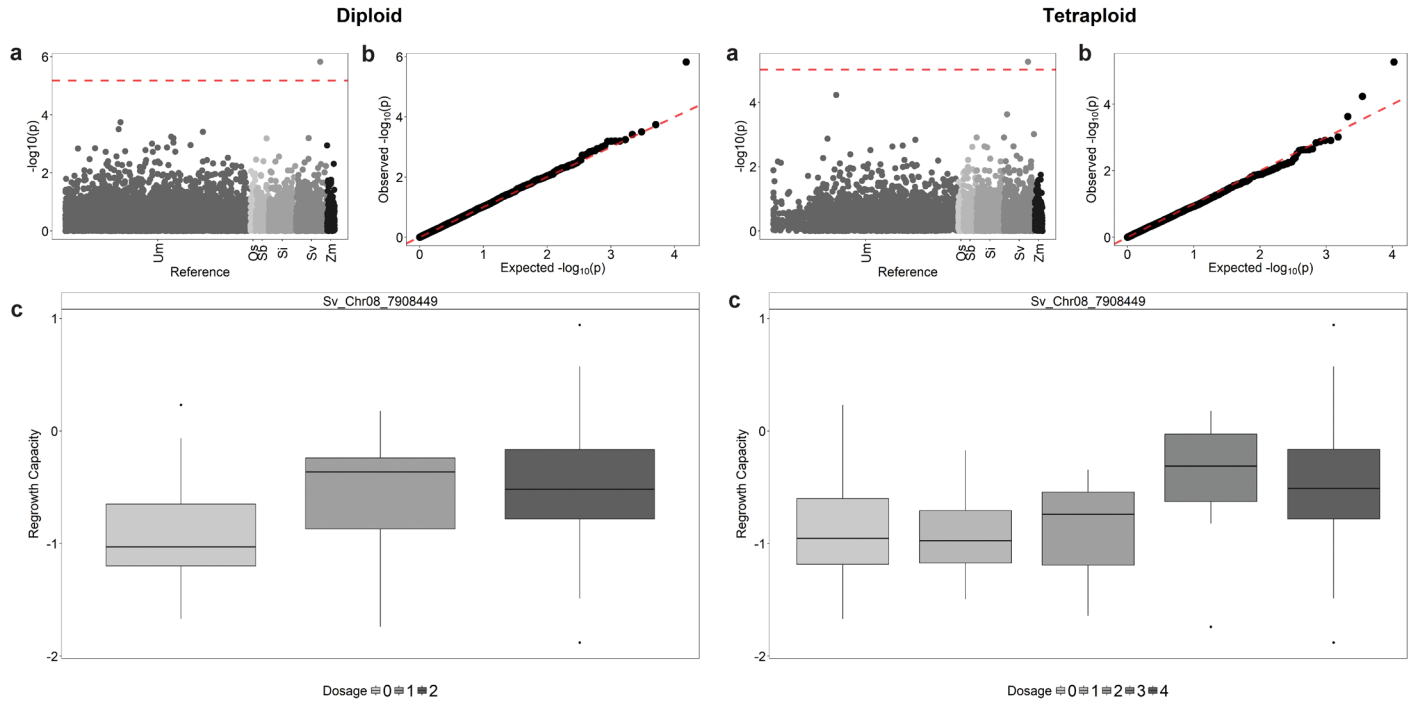


Fig. 4. Genome-wide association study (GWAS) results of *Urochloa* spp. for regrowth capacity (REG) considering the annual data using diploid and tetraploid configuration markers. (A) The Manhattan plot, (B) the quantile-quantile (QQ) plot, and (C) the boxplot showing the trait average by genotype for significant single nucleotide polymorphisms (SNPs). Reference genomes: Si, *Setaria italica*; Sb, *Sorghum bicolor*; Sv, *Setaria virides*; Os, *Oryza sativa*; Zm, *Zea mays*; At, *Arabidopsis thaliana*; Um, *Urochloa mock*.

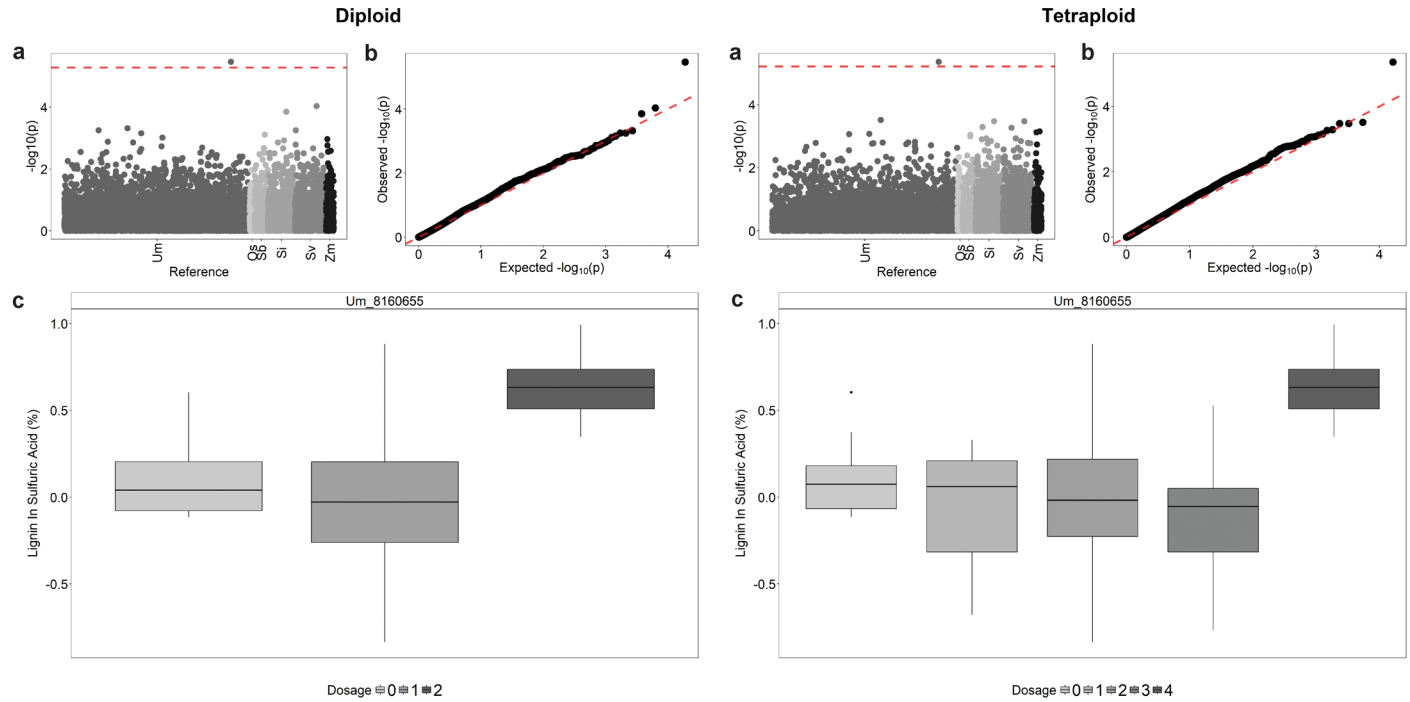


Fig. 5. Genome-wide association study (GWAS) results of *Urochloa* spp. for lignin in sulfuric acid (LIG) on the wet season using diploid and tetraploid configuration markers. (A) The Manhattan plot, (B) the quantile-quantile (QQ) plot, and (C) the boxplot showing the trait average by genotype for significant single nucleotide polymorphisms (SNPs). Reference genomes: Si, *Setaria italica*; Sb, *Sorghum bicolor*; Sv, *Setaria virides*; Os, *Oryza sativa*; Zm, *Zea mays*; At, *Arabidopsis thaliana*; Um, *Urochloa mock*.

the rainy season and dormancy in the dry one. This results in an irregular supply of fodder across the year. According to Jones, (1979), this seasonal difference in forage growth is the main obstacle to animal production in tropical and subtropical regions. Our results showed that *Urochloa* spp. interspecific population has genetic variability in seasonal production to be explored, which allows the selection of genotypes with combined maximum production for

both dry and wet seasons. One of the species used for this interspecific population is *U. brizantha*, which, despite sensitivity to water deficit, has a deep-rooted system and can contribute alleles for dry season adaptation (Santos et al., 2013). The morphological advantage from this parental species may have improved specific genotypes to better resist the water stress resulting in a fast regrowth in the dry season. These results support the strategy of this breeding program to develop cultivars for the Cerrado biome (Janusckiewicz et al., 2015), the climate of which is classified as tropical continental Köppen type Aw (Alvares et al., 2013).

Seasonal variation modifies the environment and promotes physiological and morphological reactions in the plant. For example, ∼20% of the measured metabolites in potato leaflets were simultaneously affected by drought, $CO_2$ enrichment, and diurnal factors combined (Barnaby et al., 2015). During the dry season, the plant uses physiological and anatomical tools to reduce cell activity and control the osmotic regulation (Zheng et al., 2000). Consequently, there is a reduction in cellular turgor and leaf area expansion, stomata closure, floral abscission, acceleration of tissue senescence, and reduction of growth and photosynthesis (Endres et al., 2010; Xoconostle-Cázares et al., 2010; Varshney et al., 2011). All these changes in the plant during water stress may have invoked the expression of different alleles in the evaluated interspecific hybrids, exposing the variability between the genotypes and, consequently, increasing the heritability for the majority of the traits in the dry season (Table 1).

The correlation between agronomical and nutritional traits follows what has been observed previously for the genus *Urochloa*, as described for *U. humidicola* (Rendle) Morrone & Zuloaga (de Figueiredo et al., 2012), *U. decumbens* (Matias et al., 2016), *U. ruziziensis* (Simeão et al., 2016) and *U. brizantha* (Mauri et al., 2015). For forage growth, a considerable content of lignin and fiber is needed for the structural development and thickening of the cell wall, but this is an undesirable trait, as it decreases digestibility. These nutritional traits will be present in tillers after the plant senesce. Furthermore, the accumulation of old tillers increases the proportion of epidermis, bundle sheath cells, and xylem that is not digested. In turn, these morphological structures are heavy and increase the correlation with plant weight. Even though leaves are lighter, this is the most important component of the forage for animal production on pastures, and thus leaf dry matter production should be the target in any forage breeding program (Van Soest, 1995).

## Importance of the Annotated Genes for Forage Yield

This new genomic information can be used for many biological studies and applications in breeding such as genomic selection and GWAS analysis. Here, the genomes of five grasses were used to discover SNPs in an interspecific *Urochloa spp.* population. These markers were evaluated in GWAS analyses to find markers in linkage disequilibrium with genomic regions for forage traits, following the descriptions of Collard and Mackill (2008). Furthermore, to our knowledge, this is the first study reporting the application of GBS in a panel of *Urochloa* spp. for GWAS analysis. We found seven SNPs in candidate regions related to forage yield. The marker *Sv_Chr08_79084* tags a pleiotropic gene that was significant for distinct agronomical (REG) and nutritional (IVD) traits (Table 2). Reports on the gene function annotated for this marker describes the tRNA(Asp) as the acceptor of aspartyl-tRNA synthetase; this recognition is highly specific and essential for cell viability (Choi et al., 2003). Aspartyl-tRNA synthetase (AspRS) is encoded by the impaired in baba-induced immunity 1 (IBI1) gene that, in turn, is activated by b-aminobutyric acid (BABA) to control plant immunity and growth pathways (Luna et al., 2014). Hence, this marker is correlated with aspartate (Asp) metabolism, one of the prominent amino acids in leaf tissues which is usually decreased in response to abiotic stress such as drought, as described for potato plants (Barnaby et al., 2015) and barley (*Hordeum vulgare* L.; Singh et al., 1973). In addition, Asp is a reserve of organic N, so its decrease during water stress suggests that rates of N uptake and assimilation can be diminished (Sicher and Barnaby, 2012). In this study, this marker was significant for the annual and dry seasons and thus implicates the influence of Asp on the plant growth pathways, which is directly related to REG. The *Si_Scaffold5_15551397* marker showed significant association with NDF, annotated with gene *LOC101778276*, which synthesizes the exocyst complex component SEC15B. In turn, this is involved in cell growth and organ morphogenesis, part of the cell plate development on the new primary cell wall. Also, it is involved in the docking of exocytic vesicles with fusion sites on the plasma membrane during secretion (Fendrych et al., 2010). Furthermore, previous findings indicate the role of this macromolecule in cooperation with other proteins for the secretion of cellulose synthase complexes (Zhu et al., 2018), as the cellulose directly related to the fiber content in the plant.

On the other hand, for some markers, there were no annotated genes, or uncharacterized proteins found. However, these *Urochloa* spp. genomic regions have genomic variability associated with fundamental forage traits. For example, FGW is the most reported trait in forage studies, directly related to forage production. Therefore, further studies to characterize this region could help breeders understand the genetic base of forage development.

Different markers were identified for the same trait in different environmental conditions (annual, dry, and wet season), which corroborates that forage yield is associated with the hybrid's performance under abiotic and biotic

stresses (Pabón et al., 2007; Mendonça et al., 2013; Matias et al., 2016). Furthermore, it may indicate a pleiotropic action among many forage traits. For instance, the marker *Um_8160655* was significant for NDF (annual period) and LIG (wet season), and earlier phenotypic studies have shown the high correlation among fiber and lignin content in *Urochloa* species (de Figueiredo et al., 2012; Matias et al., 2016). Although no significant markers for the nutritional trait CP were found, it showed a high correlation with digestibility (Supplemental Table S2). Thus, IVD performance could indirectly evaluate CP.

Specifically, the significant SNPs related to NDF, for the annual and dry season, aligned with the *Urochloa mock* (*Um*) reference genome, showed a similar final sequence (Supplemental Fig. S1). Probably, these markers are in linkage disequilibrium with different copies of the same gene scattered in the polyploid genome. Another possibility is classifying this sequence as a repetitive DNA sequence, as previously reported by Matias et al. (2019). Transposons and retrotransposons are common in the genome of several polyploid species playing an essential role in genome and gene evolution (Vicient and Casacuberta, 2017). In this sense, further studies are necessary to verify its distribution and frequency within the genome. Once confirmed, this sequence could be used as a marker for phylogenetic analysis and could be a benchmark towards unraveling the origin of polyploid species of *Urochloa* spp. and their relationship with closely related species.

The alignment of significant SNPs with reference genomes revealed a considerable consensus of genomic regions between *Urochloa* and other important grasses (Matias et al., 2019). It highlights that these species share genes and genomic regions with *Urochloa* spp. Among them, *Setaria* spp. genomes allowed more alignments and coverage. In accordance with our results, Ferreira et al. (2019) found *Setaria viridis* as the better alternative pseudo-genome in a *U. decumbens* panel. It indicates that this reference genome may be an option to develop SNP primers while the *Urochloa* complete genome is not available. Furthermore, these SNPs might help breeders improve forage yield in other Panicoideae grasses, if used as a novel model plant for understanding genetic and biological processes in the tribe Poaceae (Tang et al., 2017). On the other hand, just the terminal and central regions of chromosome 2 of *Sorghum bicolor* had common alignments with *Urochloa*. This result corroborates with phylogeny and genome evolution studies in grasses where *Urochloa* spp. and *Setaria* spp. belong together in the same evolutionary clade whereas *Sorghum bicolor* belongs to a different clade (Gale and Devos, 1998; Paterson et al., 2009; Schnable et al., 2009; The International Brachypodium Initiative, 2010).

### GWAS for Segmental Allotetraploids Species

Most of the genetic studies in polyploid species simplify the data to use diploid models inducing errors as under- or overestimating the real genetic control of important traits (Dufresne et al., 2014), as observed in the autotetraploid blueberry (Ferrão et al., 2018) and in the allotetraploid clover (Inostroza et al., 2018). However, *Urochloa* spp. are segmental allopolyploid species, and both ploidies diploid and tetraploid genotyping configuration should be accounted for during the GWAS analysis. Our results showed the importance of using tetraploid and diploid configuration for the same markers to identify significant regions and gene action. This approach should be used until the genome becomes available and indicates which region follows a disomic inheritance and which follows tetrasomic inheritance. Our results indicate that it is possible to evaluate tetraploid regions of the *Urochloa* spp. genome using diploid configuration; however, the genetic effect of alleles may be masked in regions with dominance or epistatic control (Fig. 4). Higher population size and greater read depth (genotyping step) can improve the statistical power to estimate the allele dosage and the trustworthy ploidy for each genome region. For example, assuming one region as a diploid, $p(Aa) = p(Aaaa) + p(AAaa) + p(AAAa)$, where the highest number of individuals of the population in class *AAaa* compared with *Aaaa* and *AAAa* is expected, could cause an error of genotyping call.

## CONCLUSION

This study presents the first GWAS analysis in interspecific segmental allotetraploid *Urochloa*, the most important forage genus in the tropical regions. The genetic variability of this panel allowed the identification of SNP markers significantly associated with forage yield traits in different cutting seasons. We found seven different regions related to the main forage traits, which can be a specific region related only with one trait, the same significant region conserved between different genomes, and a pleiotropic region between two or more traits. The season (dry or wet) may influence the genomic regions that are controlling the trait variability, as observed for digestibility. Unfortunately, *Urochloa* does not have a reference genome yet, but the region around these markers can be further investigated and yield improved knowledge about the genomic control of tropical forage traits. This study contributed to better understanding the genome of segmental allopolyploid species, showing the necessity to evaluate the molecular data using both ploidies (diploid and tetraploid configurations) to account for all regions of the genome. Finally, the significant SNPs can be useful to the breeding program to accelerate the selection of future cultivars by reducing the cost and time of evaluation.

## Supplemental Material

Supplemental material is available online for this article.

## Conflict of Interest

The authors declare that there is no conflict of interest.

## Author Contributions

F.I. Matias and R. Fritsche-Neto designed the study. S.C.L. Barrios and C.B. do Valle conducted the field experiment and collected the phenotypic data. K.G.X. Meireles performed the DNA extraction. F.I. Matias performed the SNP calling and filtering. F.I. Matias and M.S. Vidotti performed the data analyses and interpretation. F.I. Matias, M.S. Vidotti, C.A.S. Carley, and R. Fritsche-Neto wrote the paper C.A.S. Carley provided analytical expertise and edited the manuscript. R. Fritsche-Neto supervised the whole study. All authors read and approved the final version of the manuscript for publication.

## Data Availability

All original data are available on request (http://www.genetica.esalq.usp.br/alogamas/data.html). Please check the data in the link above by the title "Data *Urochloa* spp.—Embrapa (Beef Cattle)."


## Acknowledgments

We thank the National Council for Scientific and Technological Development (CNPq), Brazilian Agricultural Research Corporation (EMBRAPA) for financial support. We thank the 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.


## References


Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410. doi:10.1016/S0022-2836(05)80360-2

Alvares, C.A., J.L. Stape, P.C. Sentelhas, J.L. De Moraes Gonçalves, and G. Sparovek. 2013. Köppen's climate classification map for Brazil. Meteorol. Zeitschrift 22:711–728. doi:10.1127/0941-2948/2013/0507

Amadeu, R.R., C. Cellon, J.W. Olmstead, A.A.F. Garcia, M.F.R. Resende, and P.R. Muñoz. 2016. AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. Plant Genome 9(3). doi:10.3835/plantgenome2016.01.0009

Barnaby, J.Y., D. Fleisher, V. Reddy, and R. Sicher. 2015. Combined effects of $CO_2$ enrichment, diurnal light levels and water stress on foliar metabolites of potato plants grown in naturally sunlit controlled environment chambers. Physiol. Plant. 153:243–252. doi:10.1111/ppl.12238

Barrios, S.C.L., C.B. do Valle, G.F. Alves, R.M. Simeão, and L. Jank. 2013. Reciprocal recurrent selection in the breeding of *Brachiaria decumbens*. Trop. Grasslands-Forrajes Trop. 1:52–54. doi:10.17138/TGFT(1)52-54

Barter, R.L., and B. Yu. 2018. Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data. J. Comput. Graph. Stat. 27:910–922. doi:10.1080/10618600.2018.1473780

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

Bourke, P.M., R.E. Voorrips, R.G.F. Visser, and C. Maliepaard. 2018. Tools for genetic studies in experimental populations of polyploids. Front. Plant Sci. 9:513. doi:10.3389/fpls.2018.00513

Butler, D.G., B.R. Cullis, A.R. Gilmour, and B.J. Gogel. 2009. ASReml-R reference manual mixed models for S language environments. Train. Ser. QE02001. Queensland Dep. Prim. Ind. Fish., Brisbane.

Choi, H., K. Gabriel, J. Schneider, S. Otten, and W.H. McClain. 2003. Recognition of acceptor-stem structure of tRNA(Asp) by *Escherichia coli* aspartyl-tRNA synthetase. RNA 9:386–393. doi:10.1261/rna.2139703

Collard, B.C., and D.J. Mackill. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc., B 363:557–572. doi:10.1098/rstb.2007.2170

Crossa, J., P. Pérez-Rodriguez, J. Cuevas, O. Montesinos-López, D. Jarquin, G. de los Campos, et al. 2017. Genomic selection in plant breeding: Methods, models, and perspectives. Trends Plant Sci. 22:961–975. doi:10.1016/j.tplants.2017.08.011

de Figueiredo, U.J., J.A.R. Nunes, and C.B. do Valle. 2012. Estimation of genetic parameters and selection of *Brachiaria humidicola* progenies using a selection index. Crop Breed. Appl. Biotechnol. 12:237–244. doi:10.1590/S1984-70332012000400002

de Paula, C.M.P., F. Souza Sobrinho, and V.H. Techio. 2017. Genomic constitution and relationship in *Urochloa* (Poaceae) species and hybrids. Crop Sci. 57:2605–2616. doi:10.2135/cropsci2017.05.0307

DePristo, M.A., E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498. doi:10.1038/ng.806

do Valle, C.B., and M.S. Pagliarini. 2009. Biology, cytogenetics, and breeding of *Brachiaria*. In: R.J. Singh, editor, Genetic resources, chromosome engineering, and crop improvement. CRC Press, Boca Raton, FL. p. 107–155

DOE-JGI. 2018a. *Setaria viridis* v1.1. USDOE Joint Genome Inst., Walnut Creek, CA. http://phytozome.jgi.doe.gov/ (accessed 3 Mar. 2018).

DOE-JGI. 2018b. *Sorghum bicolor* v3.1. USDOE Joint Genome Inst., Walnut Creek, CA. http://phytozome.jgi.doe.gov/ (accessed 3 Mar. 2018).

Dufresne, F., M. Stift, R. Vergilino, and B.K. Mable. 2014. Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. Mol. Ecol. 23:40–69. doi:10.1111/mec.12581

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. doi:10.1371/journal.pone.0019379

Endelman, J.B., C.A.S. Carley, P.C. Bethke, J.J. Coombs, M.E. Clough, W.L. da Silva, et al. 2018. Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. Genetics 208:77–87. doi:10.1534/genetics.118.300685

Endres, L., J.L. de Souza, I. Teodoro, P.M.G. Marroquim, C.M. dos Santos, and J.E.D. de Brito. 2010. Gas exchange alteration caused by water deficit during the bean reproductive stage. Rev. Bras. Eng. Agric. Ambient. 14:11–16. doi:10.1590/S1415-43662010000100002

Euclides, V.P.B., D.B. Montagner, R.A. Barbosa, C.B. do Valle, and N.N. Nantes. 2016. Animal performance and sward characteristics of two cultivars of *Brachiaria brizantha* (BRS Paiaguás and BRS Piatã). Rev. Bras. Zootec. 45:85–92. doi:10.1590/S1806-92902016000300001

Fendrych, M., L. Synek, T. Pečenková, H. Toupalová, R. Cole, E. Drdová, et al. 2010. The *Arabidopsis* exocyst complex is involved in cytokinesis and cell plate maturation. Plant Cell 22:3053–3065. doi:10.1105/tpc.110.074351

Ferrão, L.F. V., J. Benevenuto, I.d.B. Oliveira, C. Cellon, J. Olmstead, M. Kirst, 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. doi:10.3389/fevo.2018.00107

Ferreira, R.C.U., L.A.d.C. Lara, L. Chiari, S.C.L. Barrios, C.B. do Valle, J.R. Valério, et al. 2019. Genetic mapping with allele dosage information in tetraploid *Urochloa decumbens* (Stapf) R. D. Webster reveals insights into spittlebug (*Notozulia entreriana* Berg) resistance. Front. Plant Sci. 10:92. doi:10.3389/fpls.2019.00092

Fristche-Neto, R., D. Akdemir, and J.L. Jannink. 2018. Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. Theor. Appl. Genet. doi:10.1007/s00122-018-3068-8 [erratum: 131:1603].

Gale, M.D., and K.M. Devos. 1998. Comparative genetics in the grasses. Proc. Natl. Acad. Sci. USA 95:1971–1974. doi:10.1073/pnas.95.5.1971

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

Grandin, T., editor. 2015. Improving animal welfare: A practical approach. 2nd ed. CAB Int., Boston, MA. doi:10.1079/9781780644677.0000

Hayes, B.J., N.O.I. Cogan, L.W. Pembleton, M.E. Goddard, J. Wang, G.C. Spangenberg, and J.W. Forster. 2013. Prospects for genomic selection in forage plant species. Plant Breed. 132:133–143. doi:10.1111/pbr.12037

He, J., X. Zhao, A. Laroche, Z.-X. Lu, H. Liu, and Z. Li. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci. 5:484. doi:10.3389/fpls.2014.00484

Henchion, M., M. Hayes, A. Mullen, M. Fenelon, and B. Tiwari. 2017. Future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. Foods 6:53. doi:10.3390/foods6070053

Inostroza, L., M. Bhakta, H. Acuña, C. Vásquez, J. Ibáñez, G. Tapia, et al. 2018. Understanding the complexity of cold tolerance in white clover using temperature gradient locations and a GWAS approach. Plant Genome 11:170096. doi:10.3835/plantgenome2017.11.0096.

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

Janusckiewicz, E.R., C.B. Chiarelli, D.C.C. Neto, E. Raposo, and A.C. Ruggieri. 2015. How the intercropping between corn and palisade grass cultivars affects forage production and pastures characteristics under grazing. Am. J. Plant Sci. 6:1475. doi:10.4236/ajps.2015.69146

Jones, C.A. 1979. The potential of *Andropogon gayanus* Kunth in the Oxisol and Ultisol savannas of tropical America. Herb. Abstr. 49:1–8.

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. doi:10.1093/bioinformatics/btr509

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2/3:18–22.

Luna, E., M. Van Hulten, Y. Zhang, O. Berkowitz, A. López, P. Pétriacq, et al. 2014. Plant perception of β-aminobutyric acid is mediated by an aspartyl-tRNA synthetase. Nat. Chem. Biol. 10:450–456. doi:10.1038/nchembio.1520

Lutts, S., J. Ndikumana, and B.P. Louant. 1991. Fertility of *Brachiaria ruziziensis* in interspecific crosses with *Brachiaria decumbens* and *Brachiaria brizantha*: Meiotic behavior, pollen viability and seed set. Euphytica 57:267–274. doi:10.1007/BF00039673

Marten, G.C., J.S. Shenk, and F.E. Barton, editors. 1989. Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality. USDA-ARS Agric. Handb. 643. U. S. Gov. Print. Office, Washington, DC.

Matias, F.I., S.C.L. Barrios, L.M. Bearari, K.G.X. Meireles, R.G. Mateus, P.N.C. do Amaral, et al. 2018. Contribution of additive and dominance effects on agronomical and nutritional traits, and multivariate selection on *Urochloa* spp. hybrids. Crop Sci. doi:10.2135/cropsci2018.04.0261

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

Matias, F.I., K. Meireles, S. Nagamatsu, S. Barrios, C. do Valle, M. Carazzolle, et al. 2019. Expected genotype quality and diploidized marker data from genotyping-by-sequencing of *Urochloa* spp. tetraploids. Plant Genome. doi:10.3835/plantgenome2019.01.0002 (in press).

Mauri, J., V.H. Techio, L.C. Davide, D.L. Pereira, F.S. Sobrinho, F. Souza, and F.J. Pereira. 2015. Forage quality in cultivars of *Brachiaria* spp.: Association of lignin and fibers with anatomical characteristics. Aust. J. Crop Sci. 9:1148–1153.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. doi:10.1101/gr.107524.110

Melo, A.T.O., R. Bartaula, and I. Hale. 2016. GBS-SNP-CROP: A reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinformatics 17:29. doi:10.1186/s12859-016-0879-y

Mendonça, S.A., S.C.L. Barrios, U.J. Figueiredo, G.F. Alves, and C.B. Valle. 2013. Agronomic and nutritional evaluation of intraspecific crosses in *Brachiaria decumbens*. Trop. Grassl. 1:103–105. doi:10.17138/TGFT(1)103-105

Mourad, A.M.I., A. Sallam, V. Belamkar, S. Wegulo, R. Bowden, Y. Jin, et al. 2018. Genome-wide association study for Identification and validation of novel SNP markers for Sr6 stem rust resistance gene in bread wheat. Front. Plant Sci. 9:380. doi:10.3389/fpls.2018.00380

NCBI Resource Coordinators. 2017. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45:D12–D17. doi:10.1093/nar/gkw1071

Ouyang, S., W. Zhu, J. Hamilton, H. Lin, M. Campbell, K. Childs, et al. 2006. The TIGR rice genome annotation resource: Improvements and new features. Nucleic Acids Res. 35:D883–D887. doi:10.1093/nar/gkl976

Pabón, A., C. Cardona, J.W. Miles, and G. Sotelo. 2007. Response of resistant and susceptible *Brachiaria* spp. genotypes to simultaneous infestation with multiple species of spittlebugs (Hemiptera: Cercopidae). J. Econ. Entomol. 100:1896–1903. doi:10.1603/0022-0493(2007)100[1896:RORASB]2.0.CO;2

Pajerowska-Mukhtar, K., B. Stich, U. Achenbach, A. Ballvora, J. Lübeck, J. Strahwald, et al. 2009. Single nucleotide polymorphisms in the *Allene Oxide Synthase 2* gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. Genetics 181:1115–1127. doi:10.1534/genetics.108.094268

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, et al. 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556. doi:10.1038/nature07723

Ramankutty, N., Z. Mehrabi, K. Waha, L. Jarvis, C. Kremen, M. Herrero, and L.H. Rieseberg. 2018. Trends in global agricultural land use: Implications for environmental health and food security. Annu. Rev. Plant Biol. 69:789–815.doi:10.1146/annurev-arplant-042817-040256

Rosyara, U.R., W.S. De Jong, D.S. Douches, and J.B. Endelman. 2016. Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome 9(2). doi:10.3835/plantgenome2015.08.0073

Santos, P.M., P.G. da Cruz, L.C. de Araujo, J.R.M. Pezzopane, C.B. do Valle, and C.d.G. Pezzopane. 2013. Response mechanisms of *Brachiaria brizantha* cultivars to water deficit stress. Rev. Bras. Zootec. 42:767–773. doi:10.1590/S1516-35982013001100001

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

Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, et al. 2009. The B73 maize genome: Complexity, diversity, and dynamics. Science 326:1112–1115. doi:10.1126/science.1178534

Sicher, R.C., and J.Y. Barnaby. 2012. Impact of carbon dioxide enrichment on the responses of maize leaf transcripts and metabolites to water stress. Physiol. Plant. 144:238–253. doi:10.1111/j.1399-3054.2011.01555.x

Simeão, R.M., A.S. Silva, and C.B. do Valle. 2016. Flowering traits in tetraploid *Brachiaria ruziziensis* breeding. Crop Breed. Appl. Biotechnol. 16:95–101. doi:10.1590/1984-70332016v16n2a15

Singh, B.T.N., L.G. Paleg, and D. Aspinall. 1973. Nitrogen metabolism and growth in the barley plant during water stress. Aust. J. Biol. Sci. 26:45–56.

Springmann, M., M. Clark, D. Mason-D'Croz, K. Wiebe, B.L. Bodirsky, L. Lassaletta et al. 2018. Options for keeping the food system within environmental limits. Nature. doi:10.1038/s41586-018-0594-0

Sun, C., B. Wang, X. Wang, K. Hu, K. Li, L. Yan, et al. 2016. Genome-wide association study dissecting the genetic architecture underlying the branch angle trait in rapeseed (*Brassica napus* L.). Sci. Rep. 6:33673. doi:10.1038/srep33673

Swenne, A., B.P. Louant, and M. Dujardin. 1981. Induction par la colchicine de formes autotétraploïdes chez *Brachiaria ruziziensis* Germain et Evrard (Graminée). Agron. Trop. 36:134–141.

Sybenga, J. 1996. Chromosome pairing affinity and quadrivalent formation in polyploids: Do segmental allopolyploids exist? Genome 39:1176–1184. doi:10.1139/g96-148

Tang, S., L. Li, Y. Wang, Q. Chen, W. Zhang, G. Jia, et al. 2017. Genotype-specific physiological and transcriptomic responses to drought stress in *Setaria italica* (an emerging model for Panicoideae grasses). Sci. Rep. 7:10009. doi:10.1038/s41598-017-08854-6

The International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768. doi:10.1038/nature08747

Tilman, D., and M. Clark. 2014. Global diets link environmental sustainability and human health. Nature 515:518–522. doi:10.1038/nature13959

Uitdewilligen, J.G.A.M.L., A.M.A. Wolters, B.B. D'hoop, T.J.A. Borm, R.G.F. Visser, and H.J. van Eck. 2013. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One 8:e0141940. doi:10.1371/journal.pone.0062355 [erratum: 10:e011940].

Van Soest, P.J. 1995. Nutritional ecology of the ruminant. Cornell Univ. Press, Ithaca, NY.

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

Varshney, R.K., L. Pazhamala, J. Kashiwagi, P.M. Gaur, L. Krishnamurthy, and D. Hoisington. 2011. Genomics and physiological approaches for root trait breeding to improve drought tolerance in chickpea (*Cicer arietinum* L.). In: A.C. de Oliveira and R.K. Varshney, editors, Root genomics. Springer, Berlin, Heidelberg. p. 233–250 doi:10.1007/978-3-540-85546-0_10

Vicient, C.M., and J.M. Casacuberta. 2017. Impact of transposable elements on polyploid plant genomes. Ann. Bot. 120:195–207. doi:10.1093/aob/mcx078

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

Voorrips, R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78. doi:10.1093/jhered/93.1.77

Worthington, M., C. Heffelfinger, D. Bernal, C. Quintero, Y.P. Zapata, J.G. Perez, et al. 2016. A parthenogenesis gene candidate and evidence for segmental allopolyploidy in apomictic *Brachiaria decumbens*. Genetics 203:1117–1132. doi:10.1534/genetics.116.190314

Worthington, M.L., and J.W. Miles. 2015. Reciprocal full-sib recurrent selection and tools for accelerating genetic gain in apomictic *Brachiaria*. In: A. Hopkins, et al., editors, Molecular breeding of forage and turf. Springer Int. Publ., Cham, Switzerland. p. 19–30. doi:10.1007/978-3-319-08714-6_3

Xoconostle-Cázares, B., F.A. Ramírez-Ortega, L. Flores-Elenes, and R. Ruiz-Medrano. 2010. Drought tolerance in crop plants. Am. J. Plant Physiol. 5:241–256. doi:10.3923/ajpp.2010.241.256

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208. doi:10.1038/ng1702

Zargar, S.M., B. Raatz, H. Sonah, MuslimaNazir, J.A. Bhat, and Z.A. Dar. 2015. Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. J. Crop Sci. Biotechnol. 18:293–308. doi:10.1007/s12892-015-0037-5

Zheng, H., R.C. Babu, M.S. Pathan, L. Ali, N. Huang, B. Courtois, and H.T. Nguyen. 2000. Quantitative trait loci for root-penetration ability and root thickness in rice: Comparison of genetic backgrounds. Genome 43:53–61. doi:10.1139/g99-065

Zhu, X., S. Li, S. Pan, X. Xin, and Y. Gu. 2018. CSI1, PATROL1, and exocyst complex cooperate in delivery of cellulose synthase complexes to the plasma membrane. Proc. Natl. Acad. Sci. USA 115:E3578–E3587. doi:10.1073/pnas.1800182115