University of São Paulo
"Luiz de Queiroz" College of Agriculture

Tropical forage breeding from classic to new genomic tools: an example with interspecific tetraploid *Urochloa* spp. hybrids

Filipe Inácio Matias

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2018

Filipe Inácio Matias
Agronomist

Tropical forage breeding from classic to new genomic tools: an example with interspecific tetraploid *Urochloa* spp. hybrids

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **ROBERTO FRITSCHE-NETO**

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2018

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Matias, Filipe Inácio

Tropical forage breeding from classic to new genomic tools: an example with interspecific tetraploid *Urochloa* spp. hybrids / Filipe Inácio Matias. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2018.

135 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Seleção genômica 2. Associação genômica 3. Modelos mistos 4. Genotipagem por sequenciamento 5. Dosagem alélica 6. *Brachiaria* spp. 7. Poliploidia I. Título

“Não é o mais forte que sobrevive, nem o mais inteligente,
mas o que melhor se adapta às mudanças.”

Leon C. Megginson

DEDICO

Aos meus pais e irmã, pelos ensinamentos, pela paciência, pelo exemplo de pessoas nas quais sempre me espelho e por sempre acreditarem na minha capacidade.

Aos meus amigos, pelo companheirismo e apoio.

AGRADECIMENTOS

Nuca é tarde para realizar nossos sonhos, pois a melhor maneira de prever o futuro é inventa-lo. Hoje, ao concluir esse trabalho, procuro entre palavras àquelas que, talvez, conseguissem expressar meus sentimentos às pessoas tão queridas. E só encontro um simples e sincero: Obrigado!

Obrigado primeiramente a Deus que tão generosamente distribuíra a dádiva dos talentos entre nós.

Obrigado aos meus pais e minha irmã, pela força que me deram, pelo apoio incondicional, pelo incentivo, pela alegria nas minhas conquistas, e mais, por fazerem seus, o meu sonho.

Obrigado aos meus familiares queridos, aos meus primos e tios que tanto me apoiaram e me ajudaram nos momentos que precisei.

Obrigado aos meus amigos pelos conselhos, segredos compartilhados, aventuras, diversões e por serem tão presentes em minha vida e por terem sido minha família nesse tempo longe de casa.

Obrigado ao Professor Dr. Roberto Fritsche-Neto e toda sua equipe, por toda a dedicação, pelos ensinamentos transmitidos, pela orientação, paciência, e por acreditar no meu potencial.

Obrigado a Dra. Karem Guimaraes Xavier Meireles e ao Dr. Sanzio Carvalho Lima Barrios pela oportunidade, pela coorientação e por me darem todo o suporte para a realização deste trabalho. Agradeço também a todos os membros de sua equipe da EMBRAPA Gado de Corte em Campo Grande/MS, pelos esforços na condução do experimento.

Obrigado ao Dr. Jeffrey Endelman da University of Wisconsin por me receber tão gentilmente em seu grupo, por me coorientar no desenvolvimento deste trabalho e por dividir tantos ensinamentos.

Obrigado a Dra. Cacilda Borges do Valle por ser uma inspiração profissional, por ter me influenciado a seguir o caminho do melhoramento de forrageiras tropicais e por todos os ensinamentos.

Obrigado a Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pelo auxílio financeiro concedido em forma de bolsa.

Obrigado a Associação para o Fomento à Pesquisa de Melhoramento de Forrageiras Tropicais - UNIPASTO e a Fundação para o Desenvolvimento da Educação, Ciência e Tecnologia do Estado de Mato Grosso do Sul - FUNDECT pelo apoio financeiro.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

CONTENTS

RESUMO

Melhoramento de forrageiras tropicais do clássico as modernas ferramentas genômicas: um exemplo em híbridos interespecíficos tetraploides de *Urochloa* spp.

Um programa de melhoramento de forragem tropical contém várias peculiaridades, especialmente quando se trata de espécies poliplóides e de apomixia facultativa. Apesar de sua importância, atualmente, faltam informações sobre estudos genéticos de características forrageiras e sobre o emprego de ferramentas genômicas quando comparadas a outras culturas e forragens de clima temperado. O gênero *Brachiaria* é o mais importante para formação de pastagens nas regiões tropicais, principalmente para produção de carne bovina. As espécies comerciais deste gênero são excelentes forrageiras perenes, e a identificação de genótipos superiores depende da seleção de muitas características sob controle genético complexo, com alto custo e avaliação demorada. Portanto, o conhecimento sobre usos e aplicações de ferramentas clássicas e genômicas em características forrageiras pode ser útil para apoiar programas de melhoramento e o desenvolvimento de novas cultivares. Nesse contexto, objetivou-se avaliar diversas ferramentas clássicas e genômicas a serem empregadas como estratégias de seleção em um programa tradicional de melhoramento de forrageiras tropicais. Um painel de híbridos tetraplóides obtidos do cruzamento *Urochloa brizantha* x *Urochloa ruziziensis* foi fenotipado e genotipado para avaliar parâmetros genéticos e realizar estudos genômicos. Para a análise fenotípica clássica, concluímos que não havia uma tendência clara da importância dos efeitos genéticos aditivos e não-aditivos para características agronômicas e nutricionais. O índice de Mulamba e Mock deve ser usado no nível univariado, devido à promoção de uma resposta mais equilibrada à seleção para todas as características na seleção multivariada. Na extração e nas avaliações genômicas, as leituras que foram alinhadas ao genoma de referência 'simulado', criado a partir dos dados de GBS da cultivar 'Marandu', tiveram a maior porcentagem de descoberta de marcadores SNP comparado aos genomas de referência mais próximos, *Setaria viridis* e *S. italica*. Recomendamos diferentes limiares de profundidade de leitura e qualidade de genótipo (GQ) para eliminar leituras de baixa qualidade sem introduzir viés de chamada de genótipo. A validação cruzada revelou que os genótipos ausentes foram imputados com uma precisão mediana de 0,85 pelo algoritmo *Random Forest* para produzir uma matriz genotípica completa, independentemente da frequência de heterozigotos. A análise de associação genômica ampla (GWAS) revelou genes candidatos associados a muitas características forrageiras tropicais, o que poderia ser o primeiro passo em direção à seleção assistida por marcadores (MAS). Além disso, nossos resultados sugerem que a contabilização da dosagem alélica é essencial, uma vez que o nível tetraploide fornece mais informações sobre o verdadeiro estado biológico. Portanto, nossos achados revelam a complexidade da arquitetura genética de características de *Urochloa* spp. e fornecem informações importantes para a aplicação de GWAS em espécies poliploides. A análise de seleção genômica revela que o GBLUP-A (aditivo) e o GBLUP-AD (aditivo + dominância) mostraram capacidades de predição semelhantes, considerando tanto os modelos simples quanto os multi-característica. Por outro lado, combinando-se

GBLUP-AD e informação tetraploide foi possível melhorar a coincidência de seleção. Além disso, o esquema de validação multi-característica 2 (VS2), onde uma característica não é avaliada para alguns indivíduos, pode fornecer um incremento de até 30% da capacidade de previsão. Portanto, é uma estratégia útil para características com baixa herdabilidade. No geral, todos os modelos de seleção genômica considerados proporcionaram maiores ganhos genéticos do que a seleção fenotípica tradicional. Da mesma forma, a dosagem do alelo associado a fatores aditivos, de dominância e multicaracteres aumentou a acurácia dos modelos genômicos de predição para híbridos poliploides interespecíficos. Finalmente, ferramentas genômicas devem ser utilizadas em programas de melhoramento de forragens para reduzir custos e tempo.

Palavras-chave: Seleção genômica; Associação genômica; Modelos mistos; Genotipagem por sequenciamento; Dosagem alélica; *Brachiaria* spp.

ABSTRACT

Tropical forage breeding from classic to new genomic tools: an example with interspecific tetraploid *Urochloa* spp. hybrids

A tropical forage breeding program contains several peculiarities, especially when it involves polyploid species and facultative apomixis. Despite their importance, there is still a lack of information on genetic studies of critical forage traits and on the employment of genomic tools when compared to other crops and temperate forages. The genus *Brachiaria* is the most important for forage in tropical regions mainly beef production. The commercial species in this genus are excellent perennial forage, and the identification of superior genotypes depends on the selection of many characteristics under complex genetic control, with high cost and time-consuming evaluation. Therefore, the knowledge about uses and applications of classic and genomic tools in forage traits may be useful to support breeding programs and the development of new cultivars. In this context, the aim was to evaluate several different classic and genomic tools to be employed as selection strategies in a traditional tropical forage breeding program. A panel of tetraploid hybrids obtained from crossing *Urochloa brizantha* x *Urochloa ruziziensis* was phenotyped and genotyped to evaluate genetic parameters and perform genomic studies. The classic phenotypic analysis showed no clear trend of the importance of additive and non-additive genetics effects for agronomical and nutritional traits. The Mulamba and Mock index should be used in the univariate level, due to the promotion of a more balanced response to selection for all traits in the multivariate selection. In the genomic extraction and evaluations, the reads that were aligned to a 'mock' reference genome, created from GBS data of the cultivar 'Marandu', had more SNP discovered compared to the closest true reference genomes, *Setaria viridis* and *S. italica*. We recommended different thresholds of sample depth and genotype quality (GQ) to eliminate poor quality reads without introducing genotype bias. Cross-validation revealed that missing genotypes were imputed with a median accuracy of 0.85 using *Random Forest* algorithm to produce a complete genotype matrix, regardless of heterozygote frequency. The genome-wide association analysis (GWAS) revealed candidate genes associated with many tropical forage traits across all cutting seasons, which could be the first step toward marker-assisted selection (MAS). Moreover, our results suggest that accounting for allele dosage is essential, since the tetraploid level provided more information about the true biological state. Therefore, our findings revealed the complexity of the genetic architecture of *Urochloa* spp. traits and provided important insights towards the application of GWAS in polyploids species. The genomic selection analysis revealed that GBLUP-A (additive) and GBLUP-AD (additive + dominance) showed similar prediction abilities considering both single and multi-trait models. Conversely, combining GBLUP-AD and tetraploid information could improve the selection coincidence. Furthermore, the multi-trait validation scheme 2 (VS2), where one trait is not evaluated for some individuals, provided an increment of up to 30% to the prediction ability. Therefore, it is an useful strategy for traits with low heritability.

Overall, all genomic selection models considered provided greater genetic gains than the 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 tools should be used in forages breeding programs in order to reduce cost and time.

Keywords: Genomic selection; Genomic association; Mixed models; Genotyping-by-sequencing; Allele dosage; *Brachiaria* spp.

1. INTRODUCTION

Animal protein is an important nutritional source for humans' health due to the high biological value (World Health Organization and United Nations University, 2007). The consumption of animal protein increased following the population growth, and meat and milk from cattle are undoubtedly the favored and most important protein sources (Tilman and Clark, 2014; Henchion et al., 2017; FAO, 2018). Despite improved animal production technologies and sustainable animal production systems there is still an association between a negative impact on the environment and cattle production (Henchion et al. 2017). Thus, the use of native and cultivated pastures rather than animal confinement has contributed to animal welfare, product quality, environmental protection and costs reduction, resulting in a more natural protein source (Lupo et al., 2013; Picasso et al., 2014; Grandin, 2015; O'Callaghan et al., 2016a; b; Henchion et al., 2017).

Brachiaria is the most cultivated genus as pasture on tropical livestock farms and *U. decumbens* (syn. *Brachiaria decumbens*) and *U. brizantha* (syn. *B. brizantha*) the most used species (Jank et al., 2014; Pessoa-Filho et al., 2017). Cultivars from this genus have many advantages such as tolerance to poor and acid soils, good carrying capacity, tolerance to insects and good nutritional value. The economically important *Brachiaria* cultivars are tetraploid (2n = 4x = 36), but this genus also has diploid (2n = 2x = 18), pentaploid (2n = 5x = 45) and hexaploid commercial genotypes (2n = 6x = 54) (Mendes-Bonato et al., 2002; Jank et al., 2014). Due to the cytogenetic behavior of the *Brachiaria* polyploid species, Mendes-Bonato et al. (2002) and Worthington et al. (2016) suggested that they may be complex segmental allopolyploids. The segmental allopolyploids have partial homology among chromosomes and sets of chromosomes with differentiated behavior, varying between allopolyploid and autopolyploid (Sybenga, 1996). Segmental allopolyploids also are common for other plants as already described in the literature for buffelgrass (Jessup et al., 2003) and *Leucaena benth* (Boff and Schifino-Wittmann, 2003).

Initially, the commercial cultivars used in tropical regions came from introductions from Africa (Jank et al., 2011, 2014; Maass et al., 2015). The hybridization between *U. brizantha* tetraploid cultivars was impossible due the apomictic reproduction and lack of compatible sexual source (Miles, 2007). Then in 1981 the *Urochloa ruziziensis*, that is originally a diploid species, had the genome duplicated by colchicine (Swenne et al., 1981) opening a new possibility for *Brachiaria* breeding programs. Now it is possible to hybridize in a recurrent selection schemes (Miles, 2007; Worthington and Miles, 2015)

between sexual "polyploidized" *U. ruziziensis* with apomictic commercial pollen donor *U. decumbens* or *U. brizantha* (Lutts et al., 1991; Souza-Kaneshima et al., 2010; Monteiro et al., 2016; Worthington et al., 2016). Thus, it is possible to select superior males (apomictic) and females (sexual) genotypes to provide a unique and distinctive diversity interspecific hybrids of *Urochloa* spp to be explored.

An ideal forage plant should produce large quantities of dry matter, especially of leaves, good regrowth ability, which allows for more grazing cycles throughout the year, and high nutritional value regarding the content of protein, less fiber and lignin content, which allows for good digestibility (Resende et al. 2008). Although selection on a single trait provides for higher genetic progress, its use may be detrimental due to the occurrence of unwanted correlations between the traits of interest under improvement (Bauer and Léon 2008). Thus, multivariate methods, such as principal component analysis, allow multivariate patterns of interest to be shown graphically, assisting in the simultaneous selection of agronomic and nutritional characteristics.

In a breeding program, it is necessary to estimate genetic parameters in order to establish the best strategy of hybridization, selection and germplasm organization considering traits of interest. Estimation of genetic parameters and correlations have been described in the literature for *U. decumbens* (Mateus et al. 2015; Matias et al. 2016) and *U. humidicola* (Figueiredo et al. 2012). However, studies about genetic parameters such as additive and dominance variance, as well as general and specific combining ability in interspecific hybrid populations are not yet available in the literature.

The molecular studies in *Brachiaria* populations were restricted before to microsatellites (Jungmann et al., 2009, 2010; Vigna et al., 2011; Silva et al., 2013) and more recently in the use of genotyping-by-sequencing methods (GBS) (Worthington et al., 2016; Ferreira et al., 2018). However, little is known about the influence of filtering parameters criteria on GBS data in polyploids forages. New breeding panels and polyploid species can be evaluated genomically using GBS (Poland et al., 2012; Worthington et al., 2016) which is an interesting approach to reduce the complexity of polyploid genomes by restriction enzyme digestion (Huang et al., 2014). Different methods and softwares were developed to perform genotype calling of polyploid species using GBS. All these methodologies have in common the use of the ratio between the number of reads and alleles (Grandke et al., 2014). However, estimating polyploid genotypes require higher read depth compared to diploids. For instance, different genotyping call requirements can be found in the literature for tetraploid data. Genotype call could require a depth of at least 48x as described for *Poa*

grasses (Griffin et al., 2011) or among 60-80x as described for potato (Uitdewilligen et al., 2013). In addition, the genotype call for the single nucleotide polymorphisms (SNP) discovery is an inference by the allele depth, where alleles with shallow depth could be biased. Then, it is necessary to choose carefully the genotype quality score to prevent bias in genomics studies (Anderson et al., 2010; Laurie et al., 2010; McKenna et al., 2010).

Among many quantitative genetics tools, genome-wide association studies (GWAS) have been used to discover genetic regions in significant association with essential traits (Zargar et al., 2015). However, in polyploid species, GWAS is complicated, mainly due to the number of genotype classes and possible modes of gene action which, until recently, where under supported with appropriate analysis methods (Rosyara et al., 2016). Consequently, GWAS use in polyploids is relatively new (Bourke et al., 2018; Ferrão et al., 2018) and is predominantly applied by disregarding the allele dosage and then applying diploid models and software (Sun et al., 2016; Mourad et al., 2018). However, little is known about the consequences of using diploid models on the GWAS results compared to the use of the adequate allele dosage. Furthermore, despite the noteworthy importance of *Urochloa* spp. for livestock in tropical regions, as far as we know, there are no studies of GWAS using SNP markers performed on this genus.

Usually, the whole selection process, from the generation of segregating populations to the release 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 animal performance apart from plant performance (Jank et al, 2014). For instance, one selection cycle in these species demands on average of two 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 the use of 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 others trait (Bauer and Léon, 2008; Dos Santos et al., 2016; Fernandes et al., 2018). It has been successfully implemented using single by single

trait (Jia and Jannink, 2012; Guo et al., 2014) or indices (Schulthess et al., 2016; Lyra et al., 2017).

In this context, our aims were (i) to estimate the contribution of additive and non-additive effects on agronomical and nutritional traits; (ii) to estimate the accuracy of multivariate index selection efficiency; (iii) the influence of different quality filters to select markers from GBS analysis; (iv) to infer the phylogeny of *Brachiaria* species by GBS reads alignment and SNP discovery using five different grass genomes; (v) to compare the missing data imputation accuracy using the Random Forest and Mode methods; (vi) to perform GWAS analysis for forage traits of different cutting seasons using SNP markers from genotyping by sequencing (GBS); (vii) to verify the influence of allele dosage through diploid and tetraploid configuration markers in the GWAS approach; (viii) 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.

REFERENCES

Anderson, C.A., F.H. Pettersson, G.M. Clarke, L.R. Cardon, A.P. Morris, and K.T. Zondervan. 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5(9): 1564–1573.

Bauer, A.M., and J. Léon. 2008. Multiple-trait breeding values for parental selection in self-pollinating crops. Theor. Appl. Genet. 116(2): 235–242.

Boff, T., and M.T. Schifino-Wittmann. 2003. Segmental allopolyploidy and paleopolyploidy in species of *Leucaena benth*: Evidence from meiotic behaviour analysis. Hereditas 138(1): 27–35.

Bourke, P.M., V.W. Gitonga, R.E. Voorrips, R.G.F. Visser, F.A. Krens, and C. Maliepaard. 2018. Multi-environment QTL analysis of plant and flower morphological traits in tetraploid rose. Theor. Appl. Genet.: 1–15.

FAO. 2018. FAOSTAT. Available at http://faostat3.fao.org/faostat-gateway/go/to/%0Ahome/E (verified 3 March 2018).

Fernandes, S.B., K.O.G. Dias, D.F. Ferreira, and P.J. Brown. 2018. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor. Appl. Genet. 131(3): 747–755.

Ferrão, L.F. V., J. Benevenuto, I. de B. Oliveira, C. Cellon, J. Olmstead, M. Kirst, M.F.R. Resende, and P. Munoz. 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.

Ferreira, R.C.U., L.A. de C. Lara, L. Chari, S.C.L. Barrios, C.B. do Valle, J.R. Valerio, F.Z.V. Torres, A.A.F. Garcia, and A.P. de Souza. 2018. Genetic mapping with allele dosage information in tetraploid *Urochloa decumbens* (Stapf) R.D. Webster reveals insights into spittlebug (*Notozulia entreriana* Berg) resistance. bioRxiv: 360594.

Figueiredo UJ de, Nunes JAR, Valle CB do (2012) Estimation of genetic parameters and selection of *Brachiaria humidicola* progenies using a selection index. Crop Breed Appl Biotechnol 12:237–244.

Grandin, T. 2015. Improving Animal Welfare: A Practical Approach (T Grandin, Ed.). 2nd ed. CAB International, Boston, MA, USA.

Grandke, F., S. Ranganathan, A. Czech, J.R. de Haan, and D. Metzler. 2014. Bioinformatic Tools for Polyploid Crops. J. Agric. Sci. Technol. B 4(8).

Griffin, P.C., C. Robin, and A.A. Hoffmann. 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to Poa grasses. BMC Biol. 9(1): 19.

Guo, G., F. Zhao, Y. Wang, Y. Zhang, L. Du, and G. Su. 2014. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet. 15.

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(7): 53.

Huang, Y.F., J.A. Poland, C.P. Wight, E.W. Jackson, and N.A. Tinker. 2014. Using Genotyping-By-Sequencing (GBS) for genomic discovery in cultivated oat. PLoS One 9(7).

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(11): 1132–1137.

Jank, L., C. Valle, and R. Resende. 2011. Breeding tropical forages. Crop Breed. Appl. Biotechnol. S1(September 2015): 27–34.

Jessup, R.W., B.L. Burson, G. Burow, Y.-W. Wang, C. Chang, Z. Li, A.H. Paterson, and M.A. Hussey. 2003. Segmental allotetraploidy and allelic interactions in buffelgrass (*Pennisetum ciliare* (L.) Link syn. *Cenchrus ciliaris* L.) as revealed by genome mapping. Genome 46(2): 304–313.

Jia, Y., and J.-L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192(4): 1513–22.

Jungmann, L., B.B.Z. Vigna, K.R. Boldrini, a C.B. Sousa, C.B. do Valle, R.M.S. Resende, M.S. Pagliarini, M.I. Zucchi, and a P. de Souza. 2010. Genetic diversity and population structure analysis of the tropical pasture grass *Brachiaria humidicola* based on microsatellites, cytogenetics, morphological traits, and geographical origin. Genome 53(9): 698–709.

Jungmann, L., B.B.Z. Vigna, J. Paiva, A.C.B. Sousa, C.B. do Valle, P.R. Laborda, M.I. Zucchi, and A.P. de Souza. 2009. Development of microsatellite markers for *Brachiaria humidicola* (Rendle) Schweick. Conserv. Genet. Resour. 1(1): 475–479.

Laurie, C.C., K.F. Doheny, D.B. Mirel, E.W. Pugh, L.J. Bierut, T. Bhangale, F. Boehm, N.E. Caporaso, M.C. Cornelis, H.J. Edenberg, S.B. Gabriel, E.L. Harris, F.B. Hu, K.B. Jacobs, P. Kraft, M.T. Landi, T. Lumley, T.A. Manolio, C. McHugh, I. Painter, J. Paschall, J.P. Rice, K.M. Rice, X. Zheng, and B.S. Weir. 2010. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet. Epidemiol. 34(6): 591–602.

de los Campos, G., J.M. Hickey, R. Pong-Wong, H.D. Daetwyler, and M.P.L. Calus. 2013. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics 193(2): 327–345.

Lupo, C.D., D.E. Clay, J.L. Benning, and J.J. Stone. 2013. Life-Cycle Assessment of the Beef Cattle Production System for the Northern Great Plains, USA. J. Environ. Qual. 42(5): 1386.

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(3): 267–274.

Lyra, D.H., L. de Freitas Mendonça, G. Galli, F.C. Alves, Í.S.C. Granato, and R. Fritsche-Neto. 2017. Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. Mol. Breed. 37(6).

Maass, B.L., C. a O. Midega, M. Mutimura, V.B. Rahetlah, P. Salgado, J.M. Kabirizi, Z.R. Khan, S.R. Ghimire, and I.M. Rao. 2015. Homecoming of *Brachiaria*: Improved hybrids prove useful for African animal agriculture. East African Agric. For. J. 81(1): 71–78.

Matias FI, Barrios SCL, Valle CB do, et al. (2016) Estimate of genetic parameters in *Brachiaria decumbens* hybrids. Crop Breed Appl Biotechnol 16:115–122.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9): 1297–1303.

Mendes-Bonato, A.B., M.S. Pagliarini, F. Forli, C. Borges Do Valle, and M.I. De Oliveira Penteado. 2002. Chromosome numbers and microsporogenesis in *Brachiaria brizantha* (Gramineae). Euphytica 125(3): 419–425.

Miles, J.W. 2007. Apomixis for cultivar development in tropical forage grasses. Crop Sci. 47(Supplement_3): S--238.

Monteiro, L.C., J.R. Verzignassi, S.C.L. Barrios, C.B. do Valle, G. de L. Benteo, and C.B. de Libório. 2016. Characterization and selection of interspecific hybrids of *Brachiaria decumbens* for seed production in Campo Grande-MS. Crop Breed. Appl. Biotechnol. 16(3): 174–181.

Mourad, A.M.I., A. Sallam, V. Belamkar, S. Wegulo, R. Bowden, Y. Jin, E. Mahdy, B. Bakheit, A. Abo El-Wafaa, J. Poland, and P.S. Baenziger. 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(March): 380.

O'Callaghan, T.F., H. Faulkner, S. McAuliffe, M.G. O'Sullivan, D. Hennessy, P. Dillon, K.N. Kilcawley, C. Stanton, and R.P. Ross. 2016a. Quality characteristics, chemical composition, and sensory properties of butter from cows on pasture versus indoor feeding systems. J. Dairy Sci. 99(12): 9441–9460.

O'Callaghan, T.F., D. Hennessy, S. McAuliffe, K.N. Kilcawley, M. O'Donovan, P. Dillon, R.P. Ross, and C. Stanton. 2016b. Effect of pasture versus indoor feeding systems on raw milk composition and quality over an entire lactation. J. Dairy Sci. 99(12): 9424–9440.

Pessoa-Filho, M., A.M. Martins, and M.E. Ferreira. 2017. Molecular dating of phylogenetic divergence between *Urochloa* species based on complete chloroplast genomes. BMC Genomics 18(1).

Picasso, V.D., P.D. Modernel, G. Becoña, L. Salvo, L. Gutiérrez, and L. Astigarraga. 2014. Sustainability of meat production beyond carbon footprint: A synthesis of case studies from grazing systems in Uruguay. Meat Sci. 98(3): 346–354.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.-L. Jannink. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome J. 5(3): 103.

Resende, R.M.S., M.D. Casler, and M.D.V. de Resende. 2014. Genomic selection in forage breeding: Accuracy and methods. Crop Sci. 54(1): 143–156.

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: 1–10.

Dos Santos, J.P.R., R.C. De Castro Vasconcellos, L.P.M. Pires, M. Balestre, and R.G. Von Pinho. 2016. Inclusion of dominance effects in the multivariate GBLUP model. PLoS One 11(4).

Schulthess, A.W., Y. Wang, T. Miedaner, P. Wilde, J.C. Reif, and Y. Zhao. 2016. Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theor. Appl. Genet. 129(2): 273–287.

Silva, P.I.T., A.M. Martins, E.G. Gouvea, M. Pessoa-Filho, and M.E. Ferreira. 2013. Development and validation of microsatellite markers for *Brachiaria ruziziensis* obtained by partial genome assembly of Illumina single-end reads. BMC Genomics 14(1).

Souza-Kaneshima, A.M. de, C. Simioni, M.F. Felismino, A.B. Mendes-Bonato, C. Risso-Pascotto, C. Pessim, M.S. Pagliarini, and C.B. do Valle. 2010. Meiotic behaviour in the first interspecific hybrids between *Brachiaria brizantha* and *Brachiaria decumbens*. Plant Breed. 129(2): 186–191.

Sun, C., B. Wang, X. Wang, K. Hu, K. Li, Z. Li, S. Li, L. Yan, C. Guan, J. Zhang, Z. Zhang, S. Chen, J. Wen, J. Tu, J. Shen, T. Fu, and B. Yi. 2016. Genome-Wide Association Study Dissecting the Genetic Architecture Underlying the Branch Angle Trait in Rapeseed (*Brassica napus* L.). Sci. Rep. 6.

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(6): 1176–1184.

Tilman, D., and M. Clark. 2014. Global diets link environmental sustainability and human health. Nature 515(7528): 518–522.

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(5).

Vigna, B.B.Z., L. Jungmann, P.M. Francisco, M.I. Zucchi, C.B. do Valle, and A.P. de Souza. 2011. Genetic Diversity and Population Structure of the *Brachiaria brizantha* Germplasm. Trop. Plant Biol. 4(3–4): 157–169.

World Health Organization and United Nations University. 2007. Protein and amino acid requirements in human nutrition. World Health Organ. Tech. Rep. Ser. (935): 1–265.

Worthington, M., C. Heffelfinger, D. Bernal, C. Quintero, Y.P. Zapata, J.G. Perez, J. De Vega, J. Miles, S. Dellaporta, and J. Tohme. 2016. A parthenogenesis gene candidate and evidence for segmental allopolyploidy in apomictic *Brachiaria decumbens*. Genetics 203(3): 1117–1132.

Worthington, M.L., and J.W. Miles. 2015. Reciprocal full-sib recurrent selection and tools for accelerating genetic gain in apomictic *Brachiaria*. p. 19–30. In Molecular Breeding of Forage and Turf. Springer.

Zargar, S.M., B. Raatz, H. Sonah, Muslimanazir, J.A. Bhat, Z.A. Dar, G.K. Agrawal, and R. Rakwal. 2015. Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. J. Crop Sci. Biotechnol. 18(5): 293–308.

2. CONTRIBUTION OF ADDITIVE AND DOMINANCE EFFECTS ON AGRONOMICAL AND NUTRITIONAL TRAITS, AND MULTIVARIATE SELECTION ON *Urochloa* SPP. HYBRIDS

ABSTRACT

A tropical forage breeding program contains several peculiarities, especially when it involves polyploid species and facultative apomixis. *Urochloa* spp. are an excellent perennial forages, and the identification of superior genotypes depends on the selection of many characteristics under complex genetic control, with high cost and time-consuming evaluation. Therefore, the use of tools such as multivariate analysis and diallel analyzes could contribute to improving the efficiency of breeding programs. Thus, the objectives were to estimate (i) the contribution of additive and non-additive effects on agronomical and nutritional traits in a population of interspecific hybrids of *Brachiaria*, originated from a partial diallel between five apomictic and four sexual parents and (ii) the accuracy of multivariate index selection efficiency. Genetic variability was detected between the parents, crosses, and hybrids for all the traits. There was no clear trend of the importance of the additive and non-additive genetic effects on agronomical and nutritional traits. Furthermore, the predominant component of genetic variance changed descending on the characteristic. Moreover, there was no outstanding parent or cross on all traits simultaneously, showing the high variability generated from these parents. The *Mulamba* and *Mock* index associated with principal components analysis allowed a more significant gain only for agronomic characteristics. However, the per se index, at the univariate level, promoted a more balanced response to selection for all traits.

Keywords: *Brachiaria*; Principal components; Mixed models; Apomixis; Mulamba and Mock index

Published in *Crop Sci.* 58:1–15 (2018). As doi 10.2135/cropsci2018.04.0261

2.1. INTRODUCTION

Grasslands are the most economical source of livestock feed in dairy and beef production (Jank et al. 2014). In tropical regions, forage grasses such as *Brachiaria* spp., and *Panicum maximum* are the most planted pastures (Almeida et al. 2011), covering millions of hectares. Due to reproduction through apomixis and to a shortage of commercial cultivars of certain species, such as *U. decumbens*, with a single cultivar available, pastures constitute extensive monocultures with their associated risks (Sluszz 2012). It is thus necessary to

generate new sources of variability either by introducing exotic genotypes or by hybridization whenever possible. Considering the genus *Brachiaria* in Brazil, Embrapa (Empresa Brasileira de Pesquisa Agropecuparia) maintains a germplasm bank with about 450 accesses of thirteen different species. Four of these species have commercial importance (*U. brizantha*, *U. decumbens*, *U. ruziziensis* and *U. humidicola*). *U. ruziziensis* is a diploid sexual species, while *U. brizantha*, *U. decumbens* and *U. humidicola* are apomictic species with different levels of ploidy. For these last three species, the diploid genotypes found in nature are sexual, whereas the accessions of upper ploidy are obligatorily apomictic (Resende et al., 2008).

To accomplish the hybridization in *Brachiaria*, chromosomal duplication of diploid sexual individuals (Swenne et al. 1981; Simioni and Valle 2009) was used in crosses to overcome the apomixis barrier, thus creating new hybrids in the process of forage cultivar development. Sexual tetraploids of *U. ruziziensis* were crossed to natural tetraploids of *U. brizantha* and *U. decumbens* to obtain interspecific hybrids (*Brachiaria* interspecific breeding program of Embrapa), while sexual tetraploids of *U. decumbens* were crossed to a natural tetraploid commercial cultivar of *U. decumbens* "Basilisk" (*Urochloa decumbens* intraspecific breeding program of Embrapa) (Lutts et al. 1991, Souza-Kaneshima, et al. 2010, Mateus et al. 2015, Matias et al. 2016). Hybridization allowed the introgression of traits of interest such as high forage quality of *U. ruziziensis* with resistance to pasture spittlebugs present in some *U. brizantha* and tolerance to acid soils and to low soil fertility of *U. decumbens* as well (Resende et al. 2008; Jank et al. 2014).

The improvement of forage is usually time-consuming and costly because it involves several steps which differ from other plant species since it requires evaluations with grazing animals. Forages in themselves are not the final product desired, but the substrate for the synthesis of animal protein (Jank et al. 2011). Thus, obtaining new cultivars from a partial diallel scheme involves hybridization between elite apomictic and superior sexual parents, evaluation of full or half-sibs, identification of the mode of reproduction (apomictic or sexual), evaluation of value of cultivation and use under cuts (VCU-cut) and VCU under grazing (VCU-grazing) to establish animal performance (Jank et al., 2014; Jank et al., 2011; Valle et al., 2015; Barrios et al., 2013; Hanna and Bashaw, 1987).

An ideal forage plant should produce large quantities of dry matter, especially of leaves, good regrowth ability, which allows more grazing cycles throughout the year, and high nutritional value regarding the content of protein, less fiber and lignin content, which allows for good digestibility (Resende et al. 2008). Although selection on a single trait

provides for higher genetic progress, its use may be detrimental due to the occurrence of unwanted correlations between the traits of interest under improvement (Bauer and Léon 2008). Thus, multivariate methods, such as principal component analysis, allow multivariate patterns of interest to be shown graphically, assisting in the simultaneous selection of agronomic and nutritional characteristics.

In a breeding program, it is necessary to estimate genetic parameters in order to establish the best strategy of hybridization, selection and germplasm organization considering traits of interest. Estimation of genetic parameters and correlations have been described in the literature for *U. decumbens* (Mateus et al. 2015; Matias et al. 2016) and *U. humidicola* (Figueiredo et al. 2012). However, studies about genetic parameters such as additive and dominance variance, as well as general and specific combining ability in interspecific hybrid populations are not yet available in the literature.

The objectives were to estimate (i) the contribution of additive and non-additive effects on agronomical and nutritional traits in a population of interspecific hybrids of *Urochloa* spp., originated from a partial diallel between five apomictic and four sexual parents and (ii) the accuracy of multivariate index selection efficiency.

2.2. MATERIALS AND METHODS

Plant material

A population of 1,000 interspecific hybrids of *Urochloa* (syn. *Brachiaria*), composed by genetic background of *U. ruziziensis, U. decumbens* and *U. brizantha*, from 20 full-sib progenies, was originated from a partial diallel between five elite tetraploid apomictic male genitors (*U. brizantha* cultivar 'Marandu', *U. brizantha* cultivar 'Paiaguás', *U. decumbens cultivar* 'Basilisk', the interspecific commercial hybrid 'Mulato II' and the accession 'B140' of *U. brizantha)* with four sexual elite tetraploid hybrids used as female genitor, obtained in the interspecific breeding program of Embrapa (BS9, BS15, 336-T1 and 336-T2). The characteristics of each genitor are highlighted in Table 1.

Experimental design

The plant material was evaluated in the experimental field area of Embrapa Beef Cattle in Campo Grande, MS, Brazil (latitude 20° 27' S, longitude 54° 37' W and 530 m altitude). This region has two defined seasons, the dry season (May to October), and the wet season (November to April) characterizing the Aw climate on the Köppen's climate

classification (Alvares et al., 2013). The field had around 3,300 m^2 and was composed to be able to support two experimental sub designs:

1 - The first sub design had the objective to evaluate the 20 progenies of full siblings described above. Then, the progenies were considered in a randomized complete block design (RCBD) in ten blocks. Within each block, twenty plots represented the twenty crosses evaluated, and each plot was represented by five different hybrids of the same cross/progeny (Figure 1). Therefore, each progeny was composed of 50 siblings, and the total of plants in this sub design was represented by 1000 hybrids evaluated individually. Additionally, five clonal cutting of each parent (total of 9 parents, table 1) were placed additional in each of the ten blocks, totalizing 45 additional plots/block and 450 plots considering all trial (Figure 1). However, these genotypes were only accounted as a check on the second sub design described below.

2 - The second experimental sub design was adapted inside the last described sub design above, and it planned to evaluate the hybrid *per se*. Each plant was one different hybrid (please, follow the names inside the image, for example, HB-01, HB-02, …, HB-1000). The hybrids did not have replicates, then, an augmented block design (Federer 1961) was required. For environmental effect, the nine parents were used as checks in each block (Resende et al. 2007), represented in dark gray rectangles. On this second sub design each plot was represented only by one plant, i.e., 1,000 plots of hybrids and 450 plots of parents were added, used as checks. In other words, each hybrid was evaluated individually, since they represent different hybrid combinations from heterozygous parents (Figure 1), therefore, the same number of blocks (10) were maintained, with one plant per plot and no replication for the hybrids, spaced 1.5 m x 1.5 m between plants and area of 2.25 m2 per plot. The parents, added in each block as checks, were used only to estimate the environmental effect of the statistical design.

Phenotypic Data

The experimental plots were subjected to seven cuts from 2013 to 2014, with intervals of 48 days: 1°- 29/04/2013, 2°- 03/06/2013, 3° - 28/10/2013, 4°- 17/12/2013, 5°- 13/02/2014, 6°- 16/04/2014 and 7°- 03/06/2014.

Phenotypic Analysis

1- Agronomic characteristics

The agronomic traits were evaluated for each plant individually, through seven cuts, considering the cuts 1, 4, 5 and six as representing the rainy season and cuts 2, 3 and seven the dry season. The biomass of each plant (greenfield weight - FGW, kg ha^{-1}), cut 10 cm above the soil surface, was weighed in the field using a scale. A sub-sample of around 200 g was taken for plant morphological separation (leaves, stem and dead material) and estimation of the dry matter, leaf dry matter (LDM, kg ha^{-1}), total dry matter (TDM, kg ha^{-1}), the leaf: stem ratio (L:S), and the percentage of leaves (% F).

The final plant regrowth capacity (REG) was estimated seven days after cutting, obtained by the combination between scores for the density of regrown tillers (DEN) and regrowth speed (VEL), according to the methodology described by (Figueiredo et al. 2012). The scores for DEN were 1 = less than 20%; 2 = 20-40%; 3 = 40-60%; 4 = 60-80%; and 5 = more than 80% of regrown tillers and the scores for VEL were: 1 = low, 2 = medium 3 = high.

2- Nutritional characteristics

The characteristics related to the nutritional value of the forage were evaluated in cuts 3 and 4. A sample of green forage was obtained through simulation of grazing. This technique consists of manually removing a mass of approximately 80 g forage, in each plant, simulating the harvesting by the animal. Subsequently, this sample was dried, ground and sent to the laboratory for analysis using infrared reflectance spectroscopy (NIRS) (Marten et al. 1989). The parameters considered were crude protein (CP), *in vitro* organic matter digestibility (IVD), neutral detergent fiber (NDF) and lignin in sulfuric acid (LIG. S).

The calibration of the NIRS was performed previously by comparing the results obtained in the wet chemical analyzes the spectrum read from these same samples in the NIRS for several nutritional characteristics. For this purpose, a regression equation was estimated for each nutritional characteristic, using a set of samples of tropical forage grasses (*Urochloa* spp. and *Panicum maximum*) for this purpose (647 samples for CP, 613 for IVD, 631 for NDF and 147 for LIG.S). Estimates of the coefficient of determination were 0.99 (CP), 0.96 (IVD), 0.95 (NDF) and 0.96 (LIG.S), showing a good fit of the model for the prediction of nutritional characteristics (unpublished data).

Statistical analysis

1- Diallel analysis (progenies of full-sib)

1.1- Fitting the model with all cuttings

For the analysis of phenotypic data equations of mixed models were used testing the random effects by the likelihood ratio test (LRT) and the fixed effects by Wald F test (Paula 2013) using the *software* R with the support of the *ASReml-R* (Butler et al. 2009) package. Thus, to obtain the variance components and estimates of genetic parameters of the parents and crosses, the data were subjected to analysis via the restricted maximum likelihood method and best unbiased linear predictor (REML/BLUP) (Resende 2000), according to the following model:

$$y_{cdfg} = \mu + q_c + t_d + u_f + v_g + x_{f\times g} + z_{f\times c} + w_{g\times c} + k_{f\times g\times c} + \varepsilon_{cdfg}, \quad (1)$$

in which y is the vector of phenotypic data, μ is the vector of the fixed effect of the general mean, q is the vector for the fixed effect of cuts with $c = \{1,2,\dots,7\}$, t is the vector of random block-level effect into cuts with $t \sim N(0, I\sigma_t^2)$ with $d = \{1,2,\dots,10\}$, u is the vector of the random effect of general combining ability of apomictic parents (GCA-APO) with $u \sim N(0, I\sigma_{APO}^2)$ and $f = \{1,2,\dots,5\}$, v is the vector of the random effect of general combining ability of the sexual parents (GCA-SEX) with $v \sim N(0, I\sigma_{SEX}^2)$ and $g = \{1,2,\dots,4\}$, x is the vector of the random effect of specific combining ability (SCA) between the parents with $x \sim N(0, I\sigma_{APO_\times SEX}^2)$, z is the vector of the random effect of the interaction of the apomictic parent with cuts with $z \sim N(0, I\sigma_{APO_\times c}^2)$, w is the vector of the random effect of the interaction of the sexual parent with cuts with $w \sim N(0, I\sigma_{SEX_\times c}^2)$, k is the vector of the random effect of the interaction cross with cuts with $k \sim N(0, I\sigma_{APO_\times SEX_\times c}^2)$ and ε is the error vector with $\varepsilon \sim N(0, I\sigma_{E_1}^2)$. The components of variance were tested by the Deviance analysis (ANADEV) using the likelihood ratio test (LRT) and the significance verified by the Chi-square test with 1 degree of freedom. The LRT replaces the ANOVA and the F test of the analysis of variance in cases of models with unbalanced data (Sturion and Resende 2010).

1.2- Fitting the model for each cutting

For the estimation of variance components considering each cut, model 1 was modified to remove the effect of cut and interactions with cuts, keeping only the random effect of blocks (t), apomictic parent (u), sexual parent (v) and cross (x), as follows:

$$y_{dfg} = \mu + t_d + u_f + v_g + x_{f\times g} + \varepsilon_{dfg}, \quad (1.1)$$

1.3-Genetic components evaluation

The use of genetic variance equations adapted to tetraploid species (Lynch et al. 1998) was used to calculate the genetic components of variance for all cuts together and for each cut (harvest) separately. The approximate additive variance was estimated by the equation $\sigma_A^2 = 4 * \sigma_{APO}^2$ where $\sigma_{APO}^2$is the variance component of the apomictic parent. This additive variance is approximate since this species is a polyploid, so 1/36 of the dominance variance is present in the σ_{APO}^2component, which we, however, assumed as zero. The relative dominance variance was calculated by the interaction component $APO_{\times}SEX$, given by $\sigma_D^2 = \frac{9}{2}\left[\sigma_{APO_{\times}SEX}^2 - 2 * \sigma_{APO}^2\right]$, assuming that the trigenetic and quadrigenetic effects are equal to zero. The proportion of each variance was obtained by the division concerning the phenotypic variance $\sigma_{P.1}^2 = \sigma_A^2 + \sigma_D^2 + \frac{\sigma_{E_1}^2}{c*b}$, where $\sigma_{E_1}^2$ is the residual variance coefficient from model 1 or 1.1, c in the number of cuts (accounted only for model 1) and *b* is the number of blocks. The selection at the progeny level was simulated through the ranking of the effect of the apomictic parents, sexual parents and crosses BLUP from model 1. The narrow-sense heritability was calculated by $h_{GCA}^2 = \frac{\sigma_A^2}{\sigma_{P.1}^2}$.

2- Individual analysis (hybrids)

The variance components and estimates of the genetic parameters of individual hybrids in an augmented block design (Federer 1961) were carried out including type (check and hybrids), assisting in the correction of the blocks and cuts effect according to the following model:

$$y_{abcdg} = \mu + m_a + p_b + q_c + s_d + t_g + u_{g\times c} + \varepsilon_{abcdg}, \quad (2)$$

where y is the vector for phenotypic data, μ is the vector for the general mean fixed effect, m is the vector for the fixed effect of type (check or hybrid), p is the vector the fixed check effect with $b = \{1,2,\ldots,9\}$, q is the vector for the fixed cut effect with $c = \{1,2,\ldots,7\}$, s is the vector or the random block level within cut effect $s \sim N(0, I\sigma_s^2)$ and $d = \{1,2,\ldots,10\}$, t is the vector for the random hybrid effect with $t \sim N(0, I\sigma_g^2)$ and $g = \{1,2,\ldots,1000\}$, u is the vector of the random hybrid by cut interaction effect with $u \sim N(0, I\sigma_{g\times c}^2)$ and ε is the error effect with $\varepsilon \sim N(0, I\sigma_{E_2}^2)$.

The hybrid heritability at the means level, in the second model, was calculated by $h^2 = \frac{\sigma_g^2}{\sigma_{P.2}^2}$, for $\sigma_{P.2}^2 = \sigma_g^2 + \frac{\sigma_{g\times c}^2}{c} + \frac{\sigma_{E_2}^2}{c}$, where $\sigma_{E_2}^2$ is the residual coefficient of the variance of model 2. Selective accuracy ($\hat{r}_{\hat{g}g}$) was estimated by the square root of the hybrid heritability. Similarly described above for progenies selection, the selection of individual hybrids from model 2 was made by the rank of the hybrids BLUP and the checks BLUE.

3- Multivariate analysis

In our concern, green field weight, regrowth ability, crude protein and fiber content are the four traits that together could represent the main traits in forage performance. Then, these for traits are used to verify the multivariate pattern of the hybrid and a principal component analysis was conducted using the hybrids BLUP in model 2, followed by the construction of a *biplot* graph with the first two principal components. Each observation was identified in accordance to the cross to which it belonged. This plot had the purpose of assisting in the selection of hybrids with greater FGW, REG, CP and lower NDF. The *R* packages *prcomp* and *ggfortify* (Ginestet 2011) were used to fit this analysis and graphics.

The Mulamba and Mock (MMI) index (Mulamba and Mock 1978) was used for multivariate selection of hybrids, given by:

$$MMI_{CP_i} = \frac{P_{CP1_i} + P_{CP2_i}}{2}$$

where P_{CP1_i} is the position *i* of hybrid in the rank of the first principal component (*CP1*); P_{CP2_i} is the position *i* of the hybrid in the rank of the second principal component (*CP2*). The response to selection was estimated for a selection intensity of 1% (ten superior hybrids).

For the comparison between the multivariate selection using principal components and verification of the direct and indirect effects of this selection, the MMI index based on the relative positions of the ranking of the hybrids for the same four traits FGW, REG, CP, and NDF was carried out according to:

$$MMI_{Uni_i} = \frac{P_{PVC_i} + P_{REB_i} + P_{PB_i+} + P_{FDN_i}}{4}$$

The correlations between agronomic and nutritional characteristics were estimated by the BLUP of the hybrids in model 2 using the Pearson Method $r_{g_{(x,y)}}$ using the function

correlation from *R* package *Agricolae* (De Mendiburu 2014). Although a formal test of correlation significance has not been established yet, the significance of the correlation was verified using an approximate Student´s t-test (Steel and Torrie 1997), considering n - 2 degrees of freedom, where n is the number of hybrids.

2.3. RESULTS

Genetic variability in the populations

Statistical differences between the apomictic parents (APO) were identified (p < 0.01 and 0.05) regarding general combining ability (GCA) for VEL and LDM (Table 2). On the other hand, the sexual parents (SEX) showed differences in GCA for DEN, REG, CP and NDF (p < 0.01 and 0.05 <). Concerning the effect of crosses (SCA), there was a significant difference (p < 0.01 and 0.05 <) for FGW, TDM, VEL, REG, CP, and NDF, indicating that there are combinations of apomictic and sexual parents that can be more advantageous in the breeding program.

FGW, TDM, and REG are traits with a balance between additive and non-additive variation. FGW presented practically 50% of each genetic effect (additive and dominant) when all the cuts were evaluated together (Table 2). DEN, VEL, LDM, CP, NDF, and IVD are predominantly additive traits, with proportions varying from 20% to 90%, while LIG.S showed only dominance effect of 53% (Table 2). L:S and % F showed no relevant genetic effect when all the cuts were evaluated together. These results are only an inference about the genetic behavior of these traits in this population. The narrow-sense heritability for progeny selection ranged from 0% to 84% depending on the trait (Table 2).

When cuts were considered individually (Table S1), genetic effects varied with the seasons. FGW and TDM showed an increase in the dominance effect between cuts 3, 4 and 5 (October 2013 to February 2014). REG showed greater dominance effect on cuts 1 to 4 and greater additive effect on cuts 5 to 7. NDF and IVD showed more dominance on cut 3 and greater additive effect on cut 4. DEN, %F, L:S and LIG.S showed greater dominance effect while VEL, LDM, CP, had a greater additive effect over all the cuts.

When using model 2, for all traits, except L:S (p > 0.05), genotypic differences were found among hybrids (Table 3) (p> 0.01 and 0.05). When evaluating them individually, disregarding the effect of progenies, estimates of heritability between hybrids means were of greater magnitude: 81%, 77.9%, 62.2%, 74% and 67.5% for FGW, TDM, VEL, REG, and CP, respectively. Also, selective accuracy was higher at the individual level, with values

higher than 70% for most traits, which indicates high experimental precision and good fit of the model 2, according to the limits set by Resende and Duarte (2007). The ideal plot size for progenies of *U. ruziziensis* was approximately 3 m^2 (Dias et al., 2014), or a line of 3 m with six plants (Souza Sobrinho et al. 2010, 2011). In this experiment plots with five plants and 2.25 m^2/plant was used, which is more than commonly reported for evaluation of progenies. This spacing could have influenced positively on the experimental accuracy of model 2, for allowing some isolation of hybrids and less competition between plants.

Correlations among agronomic and nutritional traits

The characteristics more positively related to each other were FGW x TDM (0.99), FGW x REG (0.55), DEN x REG (0.73), VEL x REG (0.69), FGW x NDF (0.29) and FGW x LIG.S (0.33) (Table 4). Although DEN and VEL have high correlations with REG (approximately 0.70) these two characteristics have low correlation to each other (around 0.14) (Table 4).

The correlations of CP and IVD with most of the other traits were negative and non-significant even though they are positively correlated (0.50). The high content of fiber and lignin negatively affected digestibility as expected and, a negative correlation was observed between CP x NDF (-0.42), NDF x IVD (-0.31) and LIG.S x IVD (-0.26) (Table 4).

Selection of parents, crosses, and hybrids

The selection of the best parents and crossings was made by the ranking of the genotypic value of estimates of GCA and SCA (Table 5). Cultivar Basilisk was the apomictic parent which provided greater green weight gains, and at the same time, reduced the fiber content, and may give rise to hybrids with better nutritional value. Regarding the sexual parents, the hybrid 336-T1 was the most frequent among the best crosses and was responsible for gains in protein and less fiber content in the progenies. For FGW, REG, CP, and NDF, there was no clear pattern among the parents or a better combination of crosses for all these traits simultaneously.

Comparing the effects of the parents (Table 5) with the effects of the ten best hybrids in the population (Table 6), it was observed that the effect of the individual is far superior to the effect of the cross. However, although there are hybrids with great agronomic and nutritional potential in the population, there was no hybrid that combined all features simultaneously when using the univariate selection intensity of 1% (Table 6).

The SCA was significant for some of the main traits. However, not the same crosses were superior for all traits at the same time (Table 2). Therefore, a multivariate selection at the individual level for FGW, REG, CP, and NDF was used (Figure 2). The first two principal components explained approximately 75% of the total multivariate variation observed among hybrids.

There was no clear pattern such as the clustering of a superior cross (Figure 2). Although no multivariate pattern defines the best cross for all the traits at the same time, there are some hybrids with high potential, with higher FGW, REG, CP and lower NDF. These are written in red in Figure 2 with lower values for PC1 and PC2, simultaneously. These hybrids were selected with the help of IMM_{CP} and correspond to hybrids: 1357 (336-T1 x B140), 1954 (336- T1 x MulatoII), 2002 (336-T2 x MulatoII), 1908 (BS15 x MulatoII), 1871 and 1897 (BS09 x MulatoII), 1620, 1579 (336-T1 x BRS Paiaguás), 1623 and 1629 (336-T2 x BRS Paiaguás) (Table 7).

The selection of the ten best hybrids by MMI_{Uni} for the traits FGW, REG, CP and NDF (Table 7) are written in blue in the lower part of Figure 2. Those hybrids are 1419 (336-T2 x B140), 1366 (336-T1 x B140), 1764 (336-T1 x Basilisk), 1954, 1964 and 1984 (336-T1 x MulatoII), 2012 (336-T2 x MulatoII), 1863 (BS09 x MulatoII), 1546 (BS15 x BRS Paiaguás) and 1623 (336-T2 x BRS Paiaguás). Some of these hybrids are not among the ten best for each one of the traits at the same time (Table 6), which was expected, because these are the best hybrids in a multivariate context, representing, in general, a desirable hybrid. Thus, only hybrids 1954 and 1623 were identified using both strategies of building indexes.

A gain of 817.5, 411.5 and 921.8 $kg.ha^{-1}$ in FGW is expected for selection of the ten best hybrids by $MMI_{CP,}$ MMI_{Uni}, and Uni, respectively (Table 7). MMI_{CP} and MMI_{Uni} indexes were coincident about selection gain for REG, promoting an increase of approximately 0.4 relatives to the population average, which amounts to approximately 65% of the gain by direct univariate selection. Selection using indexes presented some differences, however: MMI_{Uni} allowed higher relative gain compare to Uni for CP and NDF of 63% and 50%, respectively, while MMI_{CP} 27% had 27% better gain for CP and a decrease of 1% for NDF (Table 7).

The direct univariate selection of the best ten hybrids had gains ranging from 1 to 60% (Table 8). Interestingly, the indirect gains of selection followed the correlations between the traits (Table 4). Direct and indirect gains were similar when selecting the 1% best hybrids for FGW and TDM as well as when selecting for CP and DIV. Univariate

direct selection for FGW promoted a reduction of -0.915 % in the performance of the population for DIV, while the univariate selection for DIV reduced -10,508 % in the performance of the population for FGW.

2.4. DISCUSSION

Diallel analysis and selection methods

Significant genetic variability among the hybrids is a fundamental principle in plant breeding to be able to perform selection for a particular characteristic of agronomic interest (Pandolfi Filho et al. 2016). In reciprocal recurrent selection schemes, the specific combining ability between the parents is used to select the best crosses whereas the general combining ability indicates the additive genetic potential of a parent to produce good hybrids (CoTDMock et al., 1949; Oliboni et al. 2013). This combined information allows the selection of parents to be recombined to produce hybrids for the next cycle of selection and intercrossing, to gradually accumulate favorable alleles with a slow reduction in variability (Bernardo 2010).

Breeding of *Urochloa* spp., particularly of *U. decumbens* and *U. brizantha*, involves recombination solely between sexual parents since apomixis prevails in this genus (Worthington and Miles 2015). Thus, a selection scheme called recurrent selection for specific combining ability has been used (RS-SCA) (Miles 2007), in which an elite apomictic parent is used as a tester on a group of sexual plants. Modifications of this scheme resemble a reciprocal recurrent selection (RRS) scheme since to promote crosses a group of apomictic individuals from a population is used as pollen donors to a sexual group of individuals from another population (Worthington and Miles 2015). Such a scheme is typically associated with a system of partial diallel crosses to obtain hybrids and genetic information on the population in each cycle of RRS. Furthermore, the parents are not taken to homozygosity before the next cycle of crossings since these are polyploids with self-incompatibility issues (Valle et al., 1996; Lapointe and Miles, 1992).

This interspecific tetraploid population presented higher additive variability than that observed in a population of diploid *U. ruziziensis* (Simeão et al. 2016c), or *Panicum virgatum* (Bhandari et al. 2010) or *Panicum maximum* (Resende et al. 2004). Most of the traits evaluated presented more addictive effect, so the selection and use of parents with higher general combining ability are reasonable (Mendes et al. 2015). It must be considered, however, that there was no statistically significant variability between sexual and apomictic

parents for several traits (Table 2). Thus, the indications are of sexual parents BS15 and 336 -T2 for higher REG, 336-T1 for higher CP and 336-T2 for lower NDF.

The green field weight and the total dry matter are volume-related characteristics of forage on pasture, which are used for estimating the potential forage on offer and the definition of carrying capacity (Santos and Corrêa 2009). These traits showed high variability of non-additive origin in the case of this population, and only SCA was significant. In this case, the selection should be based on the performance of the best cross, e.g., the B140_x_336-T1 to improve FGW (Table 5). This result is the opposite of what was observed in *U. humidicola*, in which their significant difference between the sexual parents for GCA and no significant interaction (SCA) between sexual and apomictic parents for to FGW and TDM (Figueiredo 2015). In *U. ruziziensis* significant additive variability was also observed for FGW and TDM with the narrow sense heritability of 0.31 and 0.30 respectively (Simeão et al. 2017). These values of heritability are lower than observed for the same traits in this interspecific population of *Urochloa* spp. with FGW=0.46 and TDM=0.59 (Table 2).

The absence of the effect of crosses (SCA) for some of the traits was expected (Table 2), since this population comes from interspecific crosses between polyploid heterozygous species, so that there are many sources of variation within and among progenies, to the point of not being able to discriminate clearly, parents or ideal combinations. The genetic composition of the hybrids within the progenies is very variable, and some may show high hybrid vigor whereas others have much lower vigor within the same progeny, which can contribute to the non-significance of the effect of crosses for many traits (Figueiredo 2015). Then, the best and worst hybrids are distributed among all progenies, and it is not possible to identify the best progeny. Factorial statistical models usually promote a good fit of data from partial diallel. However, the incorrect specification of parameters in the construction of complex model reduces the absorption of variability and consequently affect GCA and SCA (Ogut et al. 2014). It is possible that the number of parameters used in model 1 could also have influenced the lack of significance of the parents and crosses.

Estimates of the proportions of the additive variance and dominance were considered of moderate to a high magnitude for FGW, TDM, REG, CP and NDF in the progenies evaluated (Table 2). In progenies of half-sibs, it has been reported that for traits with lower heritability estimates greater gain is expected when selecting the parents (PST). However, for more substantial magnitude broad sense heritabilities combined selection

methods are indicated, such as selection between progenies and individual selection (HS-IND), as well as those that are based on the selection of individual phenotypes combined with information about the progeny (CSBLUP) (Resende et al. 2013). In this population, however, the selection of parents (PST) is not indicated, since there are only four parents, which could quickly raise inbreeding of the population (Han and Casler 1999). Studies on the effect of selection on the inbreeding in populations of *Urochloa*, however, have not yet been conducted to confirm this assumption. In this case, for univariate selection, the evaluation and the use of the CSBLUP, HS-IND or individual (IND) methods are indicated.

Traits such as FGW, REG, CP, DIV, NDF, and Lig.S are the most important in forage breeding since they tend to correlate well with animal production indicating a plant with good quantity and quality of forage (Montagner et al., 2012; Sousa et al., 2007). The results did not point to a single parent or crossing that could deliver all these characteristics at the same time (Table 5), but there are hybrids in the population that in a multivariate context fall within this ideotype (Figure 2). The non-standardization of complementarity of the parents, in a multivariate context, supports the idea of maintaining all parents, recombination of the sexual parents associated with the selection at the hybrid level (IND). Selected hybrids should be evaluated as to the reproductive system, and if they are sexual, they should be used as parents in crosses, and if they are apomictic, they can follow the next steps of the breeding program (Jank et al. 2014).

Traits correlations

There were positive correlations between FGW and TDM with structural characteristics such as NDF and Lig.S. On the other hand, FGW and TDM exhibited a negative correlation with CP and DIV thus the higher the weight of a plant, the higher its fiber and lignin content and lower its protein. Similar pattern was observed between REG, FGW and TDM with positive correlation of approximately 50% and low or no correlation between REG with CP, DIV, NDF and Lig.S. These results, unlike observed in *U. decumbens* (Matias et al. 2016) and *U. ruziziensis* (Simeão et al. 2016a), in which the increase in weight and regrowth promotes the increase in fiber and lignin for dealing with structural components of the cell wall (Van Soest 1995; Mauri et al. 2015). It indicates that for this population higher regrowth capacity should not promote changes in protein, fiber, and lignin. Furthermore, since the regrowth potential depends on the density and speed of

growth with a positive correlation of about 70%, the selection for REG by itself is recommended since it should automatically implicate in the increase of VEL and DEN

It is interesting to note that when performing selection for CP, which has higher heritability, one is indirectly selecting for digestibility without drastically reducing the FGW and TDM (Table 8), due to the lower correlation between these traits (Table 4). These observations contrast with those for *U. humidicola* (Figueiredo et al. 2012) and *U. decumbens* (Matias et al. 2016), where a high negative correlation between CP and FGW compromised the selection of one in favor of the other. One possibility is that the presence of alleles from *U. ruziziensis* of high nutritional value (Euclides and de Medeiros 2003), and from *U. brizantha* and *U. decumbens* which involves good agronomic production (Jank et al. 2014), allowed for a balance or independence of these groups of contrasting traits. Therefore, these results indicate that hybrids with high FGW, REG, CP and less NDF and LIG.S can be identified. Thus, through multivariate analysis, such as the principal components or multivariate indices, it should be possible to find desired ideotypes for selection of a potential cultivar (Jain and Patel 2016).

Correlation between agronomic and nutritional characteristics similar to this experiment have been estimated for a population of *U. ruziziensis* (Simeão et al. 2016c), where the concentration of fiber and cellulose were found to be inversely proportional to crude protein concentration. It follows the principle of plant physiology that growth or maturity brings about the deposition of cell wall structures and reduces energy reserves (Van Soest 1995). However, for legumes such as *Arachis* (perennial peanut), there were positive correlations between CP with LIG.S, NDF, and TDM (Simeão et al. 2016b), indicating that biomass is directly related to the concentration of CP and nutritional value in forage legumes (Phelan et al. 2014).

The strong correlation between FGW and TDM allow early selection of forage biomass (Borges et al. 2012). The selection of hybrids with greater digestibility provides high crude protein content in the population (Torres et al. 2016). Understanding the correlations between the traits and determining the forage ideotype, assists in the selection of superior hybrids in breeding programs. Direct selection for a character can also promote indirect and disadvantageous selection of other characteristics, such as direct selection for high and CP and DIV promoted the reduction of FGW and TDM (Table 8). Also, the selection to reduce Lig.S could lead to a reduction in FGW and TDM (Table 8). In this case, the selection in a multivariate context becomes an interesting alternative.

Multivariate selection and choice of potential cultivars

Multivariate selection in forages is intended to select commercial ideotypes or also to characterize germplasm banks, as previously described in *P. maximum* (Martuscello et al. 2015), *Pennisetum glaucum* (Kumari et al. 2017), *U. humidicola* (Ahmed et al. 2014) and *U. brizantha* (Olivera et al. 2014). In a multivariate analysis involving principal components, it is possible to identify groups of crosses that present desirable commercial standards in progenies (Beheshtizadeh et al., 2013; Legesse et al., 2013). In this study, however, there was no pattern to define a better parent or multivariate cross. Nonetheless, it was possible to select hybrids individually that approached the commercially desirable ideotype (Figure 2). The variability observed within the same progeny was expected, since it contained both the best and the worst hybrids at the same time.

In the literature, there are approaches for multivariate selection by different techniques, such as the canonical analysis, used by Martuscello et al. (2015) to select hybrids of *P. maximum* adapted the conditions of "agreste" region in Alagoas/Brazil. The authors selected genotypes with high leaf production, height, high leaf/stem ratio and intermediate values of volumetric density of forage. Multiplicative multivariate indices were used to select tetraploid genotypes *U. ruziziensis* to simultaneously tackle greater regrowth after cutting, dry matter and crude protein concentration (Simeão et al. 2016a). In *Arachis* spp., the Mulamba and Mock index were used to select genotypes with higher dry matter content in the dry season and the wet season, less fiber content and highest concentration of protein in the stem (Simeão et al. 2016b).

In this work, the use of the MMI_{CP} index to accomplish selection in the interspecific population of *Urochloa* spp., the gains compared to univariate selection were higher for agronomic characteristics and lower or negative for the nutritional ones (Table 7). On the other hand, although the use of the MMI_{Uni} index did not result in higher gains for agronomic characteristics (*SG.MM*) as observed for MMI_{CP} (*SG.PC*), for example, 411.52 $kg.ha^{-1}$ and 817.53 $kg.ha^{-1}$ for field green weight respectively, at least it did not promote the reduction of the ones for nutritional value, as observed for digestibility where the SG.PC was -0.1 and SG.MM was 0.82. Thus the MMI_{Uni} is performing as a more balanced index between agronomical and nutritional traits, with gains in both groups of traits. This is due to the fact that principal component analysis reorganizes the contribution of each trait to capture most variability in the first component (Marques and Marques 2005), in our case the first two components are representing mainly the variability associated to agronomic characteristics which are privileged with more significant gains than nutritional traits in

MMI_{CP} (Table 7). The MMI_{Uni} considered each trait individually, thus provided a higher balance of multivariate gain than MMI_{CP}. Therefore, the use of the traits directly in the index of Mulamba and Mock to perform multivariate selection is indicated.

2.5. CONCLUSION

This study sought to obtain a better understanding of the inheritance of main characters and selection criteria for *Urochloa* hybrids, in order to help the plant breeder in choosing the selection method to obtain new commercial hybrids. There was no clear trend of the importance of the additive and non-additive effects on agronomical and nutritional traits. Furthermore, the predominant component changed regarding the characteristic. The understanding of the genetic behavior of the main traits, as well as the correlations between these, should facilitate the choice of the best selection criteria. This information is also expected to contribute to the definition and construction of multivariate indices to both identify superior hybrids. The Mulamba and Mock index, per se, at the univariate level, promoted a more efficient response to selection for all traits of the same time than associated with two main principal components.

REFERENCES

Almeida MC da C, Chiari L, Jank L, Valle CB do (2011) Diversidade genética molecular entre cultivares e híbridos de *Brachiaria* spp. e *Panicum maximum*. Ciência Rural 41:1998-2003.

Alvares CA, Stape JL, Sentelhas PC, de Moraes Gonçalves JL, Sparovek G (2013) Köppen's climate classification map for Brazil. Meteorol 22:711–728.

Barrios SCL, do Valle CB, Alves GF, et al. (2013) Reciprocal recurrent selection in the breeding of *Brachiaria decumbens*. Trop Grasslands-Forrajes Trop 1:52–54.

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

Beheshtizadeh H, Rezaie A, Ghandi A (2013) Principal component analysis and determination of the selection criteria in bread wheat (*Triticum aestivum* L.) genotypes. Int J Agric Crop Sci 5:2024.

Bernardo R (2010) Breeding for quantitative traits in plants. Stemma Press

Bhandari HS, Saha MC, Mascia PN, et al. (2010) Variation among half-sib families and heritability for biomass yield and other traits in lowland switchgrass (L.). Crop Sci 50:2355–2363.

Borges V, Sobrinho FS, Lédo FJ da S, Kopp MM (2012) Associação entre caracteres e análise de trilha na seleção de progênies de meios-irmãos de *Brachiaria ruziziensis*. Rev Ceres 59:765–772. doi: 10.1590/S0034-737X2012000600005

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

CoTDMock RE, Robinson HF, Harvey PH (1949) Breeding procedure designed to make maximum use of both general and specific combining ability. Agron J 41:360–367.

De Mendiburu F (2014) Agricolae: Statistical procedures for agricultural research. R Packag version 1:1–6. doi: 10.1525/california/9780520268326.003.0002

Dias K das GO, Gonçalves FMA, de Souza Sobrinho F, et al (2014) Tamanho de parcela e efeito de bordadura no melhoramento de *Urochloa ruziziensis*. Pesqui Agropecuária Bras 48:1426–1431.

Euclides VCP, de Medeiros SR (2003) Valor nutritivo das principais gramíneas cultivadas no Brasil. Embrapa Gado de Corte (Documentos, 139)

Federer WT (1961) Augmented Designs with One-Way Elimination of Heterogeneity. Biometrics 17:447. doi: 10.2307/2527837

Figueiredo UJ de (2015) Capacidade combinatória e estratégias de seleção em *Brachiaria* ssp. Universidade Federal de Lavras

Figueiredo UJ de, Nunes JAR, Valle CB do (2012) Estimation of genetic parameters and selection of *Brachiaria humidicola* progenies using a selection index. Crop Breed Appl Biotechnol 12:237–244.

Ginestet C (2011) ggplot2: Elegant Graphics for Data Analysis. J R Stat Soc Ser A (Statistics Soc 174:245–246. doi: 10.1111/j.1541-0420.2011.01616.x

Han LX, Casler MD (1999) Theoretical Inbreeding at Selectively Neutral Loci in Unparental Mass Selection and Recurrent Selection with Polycrossing of Selected Plants. Crop Sci 39:1009–1015.

Hanna W, Bashaw EC (1987) Apomixis: its identification and use in plant breeding. Crop Sci 27:1136–1139.

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

Jain SK, Patel PR (2016) Genetic Diversity and Principle Component Analyses for Fodder Yield and their Component Traits in Genotypes of Forage Sorghum (*Sorghum bicolor* L. Moench). Ann Arid Zone 55:17–23.

Jank L, Barrios SC, do Valle CB, et al. (2014) The value of improved pastures to Brazilian beef production. Crop Pasture Sci 65:1132–1137.

Jank L, Valle CB, Resende RMS (2011) Breeding tropical forages. Crop Breed Appl Biotechnol 11:27–34.

Kumari J, Bag MK, Pandey S, et al. (2017) Assessment of phenotypic diversity in pearl millet [*Pennisetum glaucum* (L.) R. Br.] germplasm of Indian origin and identification of trait-specific germplasm. Crop Pasture Sci 67:1223–1234.

Lapointe SL, Miles JW (1992) Germplasm case study: *Brachiaria* species. Pastures Trop Lowl CIAT, Cali, Colomb 43–55.

Legesse H, Dechassa N, Gebeyehu S, et al (2013) Multivariate Analysis as a Tool for Indirect Selection of Common Bean Genotypes (*Phaseolus vulgaris* L) for Soil Acidity Tolerance under Field Conditions. Sci Technol Arts Res J 2:7–15.

Lutts S, Ndikumana J, Louant BP (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.

Lynch M, Walsh B, others (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA, U.S.A.

Marques JM, Marques MAM (2005) As componentes principais no descarte de variáveis em um modelo de regressão múltipla. Rev da FAE 8:93–101.

Marten GC, Shenk JS, Barton FE (1989) Near-infrared reflectance spectroscopy (NIRS): Analysis of forage quality. Agric Handb 95.

Martuscello JA, Braz TG dos S, Jank L, et al. (2015) Identification of ideotypes by canonical analysis in *Panicum maximum*. Ciência e Agrotecnologia 39:147–153.

Mateus RG, Barrios SCL, Valle CB do, et al. (2015) Genetic parameters and selection of *Brachiaria decumbens* hybrids for agronomic traits and resistance to spittlebugs. Crop Breed Appl Biotechnol 15:227–234.

Matias FI, Barrios SCL, Valle CB do, et al. (2016) Estimate of genetic parameters in *Brachiaria decumbens* hybrids. Crop Breed Appl Biotechnol 16:115–122.

Mauri J, Techio VH, Davide LC, et al. (2015) Forage quality in cultivars of *Brachiaria* spp.: Association of lignin and fibers with anatomical characteristics. Aust J Crop Sci 9:1148–1153.

Mendes MHS, Pereira CH, Souza JC de (2015) Diallel analysis of maize hybrids for agronomic and bromatological forage traits. Acta Sci Agron 37:141–146.

Miles JW (2007) Apomixis for cultivar development in tropical forage grasses. Crop Sci 47:S--238.

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

Mulamba NN, Mock JJ (1978) Improvement of yield potential of the ETO blanco maize (*Zea mays* L.) population by breeding for plant traits [Mexico].

Ogut F, Maltecca C, Whetten R, et al. (2014) Genetic analysis of diallel progeny test data using factor analytic linear mixed models. For Sci 60:119–127.

Oliboni R, Faria MV, Neumann M, et al (2013) Análise dialélica na avaliação do potencial de híbridos de milho para a geração de populações-base para obtenção de linhagens Diallelic analysis in assessing the potential of maize hybrids to generate base-populations for obtaining lines. Semin Ciências Agrárias, Londrina 34:7–18.

Olivera Y, Machado R, Ramírez J, et al (2014) Morphological characterization of 19 *Brachiaria brizantha* accessions on an acid soil. Pastures and Forages 7:137–141.

Pandolfi Filho AD, Do Valle CB, Barrios SCL, et al (2016) Avaliação de genitoras sexuais de *Brachiaria* spp. na época de seca. Arch Zootec 65:213–219.

Paula GA (2013) Modelos de regressão: com apoio computacional. IME-USP São Paulo

Phelan P, Moloney a. P, McGeough EJ, et al. (2014) Forage Legumes for Grazing and Conserving in Ruminant Production Systems. CRC Crit Rev Plant Sci 34:37–41. doi: 10.1080/07352689.2014.898455

Resende RMS, Casler MD, de Resende MDV (2013) Selection methods in forage breeding: A quantitative appraisal. Crop Sci 53:1925–1936. doi: 10.2135/cropsci2013.03.0143

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

Resende RMS, de Resende MDV, do Valle CB, et al. (2007) Selection efficiency in *Brachiaria* hybrids using a posteriori blocking. Crop Breed Appl Technol 7:296.

Resende RMS, do Valle CB, Jank L (2008) Melhoramento de forrageiras tropicais. Embrapa Gado de Corte

Resende RMS, Jank L, Do Valle CB, Bonato ALV (2004) Biometrical analysis and selection of tetraploid progenies of *Panicum maximum* using mixed model methods. Pesqui Agropecu Bras 39:335–341. doi: 10.1590/S0100-204X2004000400006

Resende MDV de (2000) Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes. Embrapa Florestas, Colombo

Resende MDV de, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. Pesqui Agropecuária Trop 37:182–194. doi: 10.5216/pat.v37i3.1867

Santos PM, Corrêa L de A (2009) Manejo de pastagens tropicais. Embrapa Pecuária Sudeste

Simeão R, Silva A, Valle C, et al. (2016a) Genetic evaluation and selection index in tetraploid *Brachiaria ruziziensis*. Plant Breed 135:246–253. doi: 10.1111/CPr.12353

Simeão RM, Assis GML, Montagner DB, Ferreira RCU (2016b) Forage peanut (*Arachis* spp.) genetic evaluation and selection. Grass Forage Sci. 72:322-332

Simeão RM, Silva AS, Valle CB do (2016c) Flowering traits in tetraploid *Brachiaria ruziziensis* breeding. Crop Breed Appl Biotechnol 16:95–101.

Simeão RM, Valle CB, Resende MD V (2017) Unravelling the inheritance, QST and reproductive phenology attributes of the tetraploid tropical grass *Brachiaria ruziziensis* (Germain et Evrard). Plant Breed 136:101–110.

Simioni C, Valle CB do (2009) Chromosome duplication in *Brachiaria* (A. Rich.) Stapf allows intraspecific crosses. Crop Breed Appl Biotechnol 9:328–333.

Sluszz T (2012) Monitoramento tecnológico de cultivares de forrageiras tropicais. Cad Prospecção 5:1–13.

Souza-Kaneshima AM de, Simioni C, Felismino MF, et al (2010) Meiotic behaviour in the first interspecific hybrids between *Brachiaria brizantha* and *Brachiaria decumbens*. Plant Breed 129:186–191. doi: 10.1111/j.1439-0523.2009.01674.x

Sousa LF, Mauricio RM, Gonçalves LC, et al (2007) Productivity and nutritional value of *Brachiaria brizantha* cv. Marandu in a silvopastoral system. Arq Bras Med Veterinária e Zootec 59:1029–1037.

Souza Sobrinho F, Borges V, Lédo FJ da S, Kopp MM (2010) Repetibilidade de características agronômicas e número de cortes necessários para seleção de *Urochloa ruziziensis*. Pesqui Agropecuária Bras 45:579–584.

Souza Sobrinho F, Lédo F da SJ, Kopp MM (2011) Estacionalidade e estabilidade de produção de forragem de progênies de *Brachiaraia ruziziensis*. Ciência e Agrotecnologia 35:684–691.

Steel RG, Torrie JH (1997) Principles and procedures of statistics: a biometrical approach, 3rd edn. McGraw-Hill, New York

Sturion JA, Resende MDV de (2010) Avaliação genética e análise de deviance em um teste desbalanceado de procedência e progênie de Ilex paraguariensis. Pesqui Florest Bras 30:157–160. doi: 10.4336/2010.pfb.30.62.157

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

Torres FE, do Valle CB, Lempp B, et al (2016) Contribuição dos caracteres de qualidade da forragem ao teor de proteína bruta em *Urochloa brizantha*. Pesqui Agropecuária Bras 51:284–287.

Valle CB do, Savidan YH, Miles JW, et al (1996) Genetics, cytogenetics, and reproductive biology of *Brachiaria*. Brachiaria Biol Agron Improv 147–163.

Van Soest PJ (1995) Nutritional ecology of the ruminant. Journal of Nutrition, 125:1025-1025.

Worthington ML, Miles JW (2015) Reciprocal Full-sib Recurrent Selection and Tools for Accelerating Genetic Gain in Apomictic *Brachiaria*. In: Molecular Breeding of Forage and Turf. Springer International Publishing, Cham, pp 19–30

FIGURES

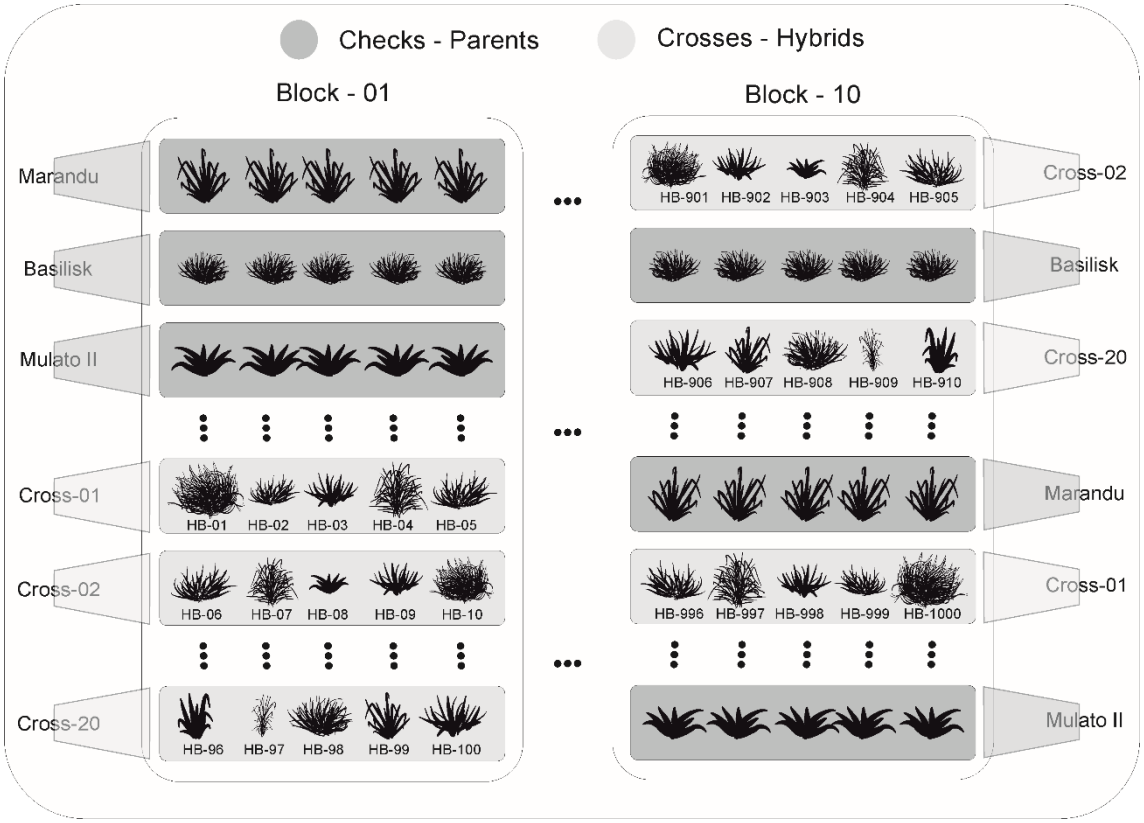

Figure 1. Scheme of two **e**xperimental sub design for the evaluation of interspecific *Brachiaria* hybrids from crosses between apomictic and sexual tetraploid heterozygous parents. The progenies are randomized and repeated in all ten blocks (ex: Cross-01, Cross-02,..., Cross-20, represented in light gray rectangles) defining the random complete block design (sub design 1). Each plot in the block is composed of 5 different hybrids from a single cross thus totaling 50 hybrids per progeny. These 50 full-sib F1 hybrids/ progenies segregate in the plot. Thus there is no hybrid *per se* replication, and there are 1,000 different hybrids in the experiment (HB-01, HB-02,..., HB-1000). To evaluate and select the individual hybrids was added on the sub design 1 the nine parents, used as checks, in each block, characterizing on an augmented block design (sub design 2). Therefore, five clonal cutting of each parent were placed additional in each of the ten blocks, totalizing 45 additional plots/block on the sub design 1 and 450 plots considering sub design 2. Each plant in sub design 2 was represented only by one plant, i.e., 1,000 plots of hybrids and 450 plots of parents (checks). The checks only were accounted for statistical analysis of the sub design 2.

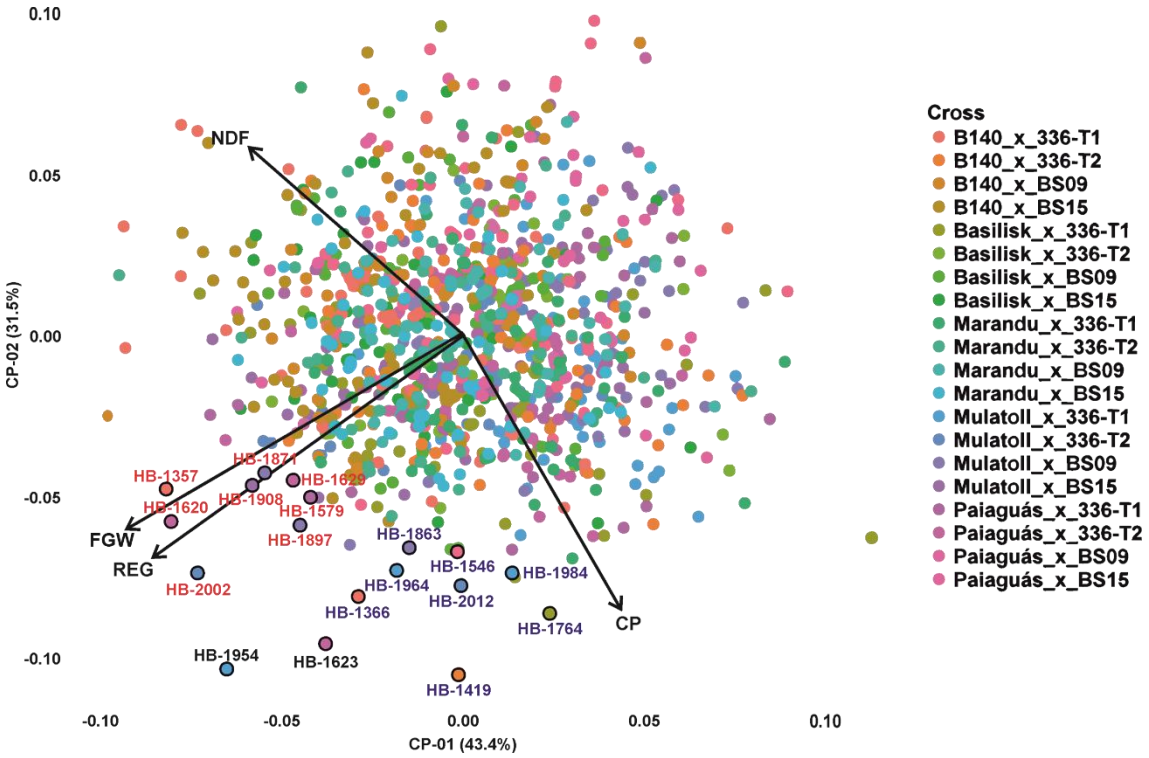

Figure 2. Biplot of the principal component multivariate analysis of 1,000 interspecific hybrids of *Brachiaria* spp., from a partial diallel between five tetraploid apomictic parents (*U. brizantha* 'Marandu', *U. brizantha* accession B140, *U. brizantha* 'BRS Paiaguás', *U. decumbens* 'Basilisk' and *Brachiaria* spp. 'Mulato II') with four tetraploid sexual interspecific hybrids (BS9, BS15, 336-T1 and 336-T2) for field green weight (kg.ha^{-1}); Regrowth scores; % crude protein and % neutral detergent fiber.

TABLES

Table 1 - Genitors that were used to obtain the progenies of full-siblings descriptions:

Apomictic genitors	
Species	**Characteristics**
U. brizantha (Cv. Marandu)	High productivity and intolerant to flood
U. brizantha (B140)	High productivity
U. brizantha (Cv. Paiaguás)	High productivity (especially in the dry season)
U. decumbens (Cv. Basilisk)	Medium-high productivity, tolerant to drought and tolerant to toxic aluminum
Urochloa spp. (Mulato II)	High productivity and intolerant to flood
Sexual genitors	
Species	***Characteristics***
4 hybrids of sexual *Urochloa spp* (BS09, BS15, 336-T1 and 336-T2)	Good seed production and high productivity

Table 2 – Deviance analysis (ANADEV) for apomictic parents (σ^2_{APO}), sexual parents (σ^2_{SEX}), crosses effect ($\sigma^2_{APO_\times SEX}$) and interaction with cuttings ($\sigma^2_{APO_\times SEX_\times c}$). Additive variance ($\sigma^2_A$), dominant variance ($\sigma^2_D$) and general mean ($\bar{X}$) for agronomic and nutritional value traits based on the evaluation of *Brachiaria* spp. hybrids in seven cuts

	FGW		TDM		DEN		VEL		REG		LDW	
σ^2_{APO}	4381.84	ns	5777.55	ns	0.00	ns	0.01	*	0.00	ns	3709.61	*
σ^2_{SEX}	0.11	ns	0.04	ns	0.02	*	0.00	ns	0.05	*	3886.82	ns
$\sigma^2_{APO_\times SEX}$	12264.92	**	13894.79	**	0.00	ns	0.01	**	0.01	**	1145.56	ns
$\sigma^2_{APO_\times SEX_\times c}$	3206.05	ns	5727.73	ns	0.00	*	0.00	ns	0.01	**	0.00	ns
σ^2_A	17527.35		23110.19		0.00		0.03		0.02		14838.43	
σ^2_D	15755.58		10528.64		0.00		0.00		0.02		0.00	
% σ^2_A	46.45		58.88		28.28		83.69		40.98		69.62	
% σ^2_D	41.76		26.82		0.00		0.00		38.03		0.00	
$\bar{X}$	1084.82		1369.41		2.98		1.54		1.60		443.96	
	%L		L:S		CP		NDF		IVD		LIG.S	
σ^2_{APO}	0.00	ns	0.15	ns	0.10	ns	0.13	ns	0.22	ns	0.00	ns
σ^2_{SEX}	14.93	*	1.97	ns	0.21	**	0.21	*	0.37	ns	0.00	ns
$\sigma^2_{APO_\times SEX}$	0.00	ns	0.40	ns	0.11	**	0.18	*	0.20	ns	0.00	ns
$\sigma^2_{APO_\times SEX_\times c}$	19.01	*	0.37	ns	0.00	ns	0.20	*	0.34	ns	0.00	ns
σ^2_A	0.00		0.61		0.41		0.50		0.90		0.00	
σ^2_D	0.00		0.42		0.00		0.00		0.00		0.01	
% σ^2_A	0.00		3.65		78.46		62.56		53.15		0.00	
% σ^2_D	0.00		2.52		0.00		0.00		0.00		53.20	
$\bar{X}$	62.86		2.92		15.44		63.34		72.33		2.07	

FGW: Field Green Weight (kg.ha^{-1}); TDM: Total Dry Matter (kg.ha^{1}); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDW: Leaf dry matter (kg.ha^{-1}); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%). **significant by *LRT test* with one degree of freedom considering 1% of probability by the x^2 *test*. * significant by *LRT test* considering 5% of probability by the x^2 *test*. ^{ns} Non-significant.

Table 3 - Additive variance (σ_A^2), dominant variance (σ_D^2) and general mean ($\bar{X}$) for agronomic and nutritional value traits based on individual cuts of *Brachiaria* spp. hybrids

Trait	Cut	$\bar{X}$	σ_A^2	σ_D^2	% σ_A^2	% σ_D^2
FGW	1	1084.98	49240.15	3785.03	47.99	3.69
	2	675.92	36902.64	0.00	66.72	0.00
	3	591.95	23052.30	4834.71	54.53	11.44
	4	1287.47	51317.24	14237.91	44.96	12.47
	5	1194.32	0.17	88776.60	0.00	67.26
	6	1065.60	109389.06	0.00	74.03	0.00
	7	431.71	10969.74	0.00	66.73	0.00
TDM	1	1369.52	64918.31	25456.30	38.88	15.25
	2	690.32	32502.95	0.00	65.61	0.00
	3	681.20	18934.35	31607.83	27.85	46.49
	4	1321.06	51737.96	16695.75	45.57	14.71
	5	1556.59	0.26	139210.47	0.00	68.15
	6	1212.89	132436.55	0.00	74.47	0.00
	7	583.11	14019.44	0.00	61.84	0.00
DEN	3	2.99	0.00	0.03	0.00	42.29
	4	2.69	0.00	0.06	0.00	61.18
	5	2.47	0.01	0.02	10.79	29.57
	6	2.51	0.00	0.02	6.51	26.21
	7	2.49	0.03	0.00	41.43	0.00
VEL	3	1.55	0.02	0.00	45.22	0.00
	4	1.46	0.05	0.00	64.33	0.00
	5	0.87	0.08	0.00	70.67	0.00
	6	0.76	0.02	0.00	40.99	0.00

	7	0.93	0.01	0.02	17.25	37.36
REG	1	1.60	0.02	0.10	12.41	51.12
	2	2.42	0.00	0.29	0.00	81.89
	3	3.53	0.00	0.11	0.00	59.61
	4	3.15	0.01	0.11	4.84	62.85
	5	2.34	0.13	0.00	65.59	0.00
	6	2.27	0.05	0.00	44.90	0.00
	7	2.42	0.02	0.00	26.32	0.00
LDW	2	443.99	13532.02	0.00	45.88	0.00
	3	407.86	16275.31	0.00	61.72	0.00
%L	2	62.86	0.00	138.96	0.00	76.36
	3	73.80	0.00	32.03	0.00	49.10
L:S	2	2.92	0.05	2.08	1.88	71.75
	3	10.60	2.64	0.00	4.03	0.00
CP	3	15.44	0.59	0.00	68.43	0.00
	4	15.57	0.26	0.00	58.53	0.00
NDF	3	63.34	0.21	0.65	13.66	43.33
	4	66.53	0.71	0.00	56.02	0.00
IVD	3	72.33	0.28	2.33	6.12	51.63
	4	69.30	1.96	0.00	61.21	0.00
LIG.S	3	2.07	0.00	0.01	0.00	42.59
	4	2.18	0.00	0.01	0.64	47.35

FGW: Field Green Weight ($kg.ha^{-1}$); TDM: Total Dry Matter ($kg.ha^{1}$); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDW: Leaf dry matter ($kg.ha^{-1}$); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%).

Table 4 – Deviance analysis (ANADEV) for *Brachiaria* spp. interspecific hybrids effect (σ_g^2). Selective accuracy ($\hat{r}_{\hat{g}g}$), heritability between hybrid means (h^2) and general mean ($\bar{X}$) for agronomic and nutritional value traits based on the evaluation in seven cuts

	FGW		TDM		DEN		VEL		REG		LDW	
σ_g^2	144522.40	**	162654.20	**	0.08	**	0.07	**	0.22	**	33837.46	**
$\hat{r}_{\hat{g}g}$	0.90		0.88		0.71		0.79		0.86		0.75	
h^2	0.81		0.78		0.51		0.62		0.74		0.56	
$\bar{X}$	1540.61		1790.55		3.57		1.81		2.29		767.86	
	%L		L:S		CP		NDF		IVD		LIG.S	
σ_g^2	41.72	**	0.00	ns	1.47	**	2.48	**	5.98	**	0.04	**
$\hat{r}_{\hat{g}g}$	0.56		-		0.82		0.70		0.71		0.65	
h^2	0.31		0.00		0.68		0.49		0.51		0.42	
$\bar{X}$	66.92		2.57		15.75		64.41		73.32		2.17	

FGW: Field Green Weight (kg.ha^{-1}); TDM: Total Dry Matter (kg.ha^{1}); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDW: Leaf dry matter (kg.ha^{-1}); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%).**significant by *LRT test* considering 1% of probability by the x^2 *test*. * significant by *LRT test* considering 5% of probability by the x^2 *test*. ^{ns} Non-significant.

Table 5 – Correlation estimates between agronomic and nutritive value traits in *Brachiaria* interspecific hybrids

	TDM	DEN	VEL	REG	LDW	%L	L:S	CP	NDF	IVD	LIG.S
FGW	0.99 *	0.40 *	0.49 *	0.55 *	0.51 *	0.03 ns	-0.17 *	0.00 ns	0.29 *	-0.19 *	0.33 *
TDM		0.39 *	0.49 *	0.54 *	0.51 *	0.02 ns	-0.16 *	-0.01 ns	0.30 *	-0.19 *	0.33 *
DEN			0.14 *	0.73 *	0.20 *	0.19 *	0.08 *	0.12 *	-0.07 *	-0.05 ns	0.08 *
VEL				0.69 *	0.33 *	0.14 *	-0.05 ns	-0.14 *	0.21 *	-0.11 *	0.24 *
REG					0.39 *	0.29 *	0.04 ns	-0.08 *	0.09 *	-0.13 *	0.20 *
LDW						0.25 *	-0.23 *	-0.15 *	0.20 *	-0.17 *	0.18 *
%L							0.38 *	0.00 ns	-0.02 ns	0.01 ns	0.02 ns
L:S								0.03 ns	-0.14 *	0.04 ns	-0.10 *
CP									-0.42 *	0.50 *	-0.08 *
NDF										-0.31 *	0.37 *
IVD											-0.26 *

FGW: Field Green Weight (kg.ha^{-1}); TDM: Total Dry Matter (kg.ha^{1}); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDW: Leaf dry matter (kg.ha^{-1}); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%).*Correlation estimate significantly different from zero by the *t* test. ^{ns} Non-significant.

Table 6 – General combining ability (GCA) estimates of apomictic parents (APO), sexual parents (SEX) and specific combining ability (SCA) estimates of the five best and five worst *Brachiaria* spp. crosses for different agronomic and nutritional value traits

				APO				
Rank	VEL	GCA	LDW	GCA	-		-	
1	B140	0.13	Basilisk	74.44	-		-	
2	Basilisk	-0.02	B140	45.51	-		-	
3	Marandu	-0.02	Marandu	-20.85	-		-	
4	MulatoII	-0.04	MulatoII	-47.23	-		-	
5	Paiaguás	-0.05	Paiaguás	-51.87	-		-	
$\bar{X}$		1.54		443.96	-		-	
				SEX				
Rank	DEN	GCA	REG	GCA	CP	GCA	NDF	GCA
1	336-T2	0.08	BS15	0.12	336-T1	0.52	336-T2	-0.51
2	BS15	0.06	336-T2	0.11	BS09	0.05	336-T1	-0.11
3	336-T1	0.04	336-T1	0.05	336-T2	-0.04	BS15	0.15
4	BS09	-0.18	BS09	-0.30	BS15	-0.52	BS09	0.46
$\bar{X}$		2.98		1.60		15.44		63.34
				APO:SEX				
Rank	FGW	SCA	REG	SCA	CP	SCA	NDF	SCA
1	B140_x_336-T1	119.38	B140_x_BS15	0.22	MulatoII_x_BS09	0.34	B140_x_336-T2	-0.36
2	Basilisk_x_BS09	93.81	Marandu_x_336-T2	0.08	B140_x_336-T2	0.32	MulatoII_x_336-T1	-0.32
3	MulatoII_x_336-T2	93.12	B140_x_BS09	0.08	Marandu_x_BS15	0.20	MulatoII_x_BS09	-0.32
4	Basilisk_x_BS15	72.85	MulatoII_x_336-T2	0.08	Basilisk_x_BS15	0.19	Basilisk_x_336-T1	-0.25
5	Paiaguás_x_336-T2	68.49	B140_x_336-T1	0.05	MulatoII_x_336-T1	0.18	Paiaguás_x_336-T2	-0.20
16	Marandu_x_336-T1	-92.65	Marandu_x_336-T1	-0.04	Paiaguás_x_BS09	-0.22	MulatoII_x_336-T2	0.14
17	Basilisk_x_336-T2	-103.25	MulatoII_x_BS15	-0.05	Basilisk_x_BS09	-0.22	Marandu_x_336-T2	0.18

18	Paiaguás_x_BS15	-121.45	Paiaguás_x_BS15	-0.14	Marandu_x_336-T2	-0.24	B140_x_BS09	0.21
19	Paiaguás_x_BS09	-122.17	MulatoII_x_BS09	-0.14	MulatoII_x_336-T2	-0.24	Basilisk_x_BS09	0.41
20	MulatoII_x_336-T1	-158.24	B140_x_336-T2	-0.14	B140_x_BS15	-0.63	B140_x_336-T1	0.72
$\bar{X}$		1084.82		1.60		15.44		63.34

FGW: Field Green Weight (kg.ha^{-1}); VEL: Regrowth speed; REG: Regrowth capacity; LDW: Leaf dry matter (kg.ha^{-1}); CP: Crude protein (%); NDF: Neutral detergent fiber (%). $\bar{X}$: General mean.

Table 7 – Genotypic value (GV) of the best ten *Brachiaria* spp. hybrids

Rank	FGW	GV	TDM	GV	DEN	GV	VEL	GV	REG	GV	LDW	GV
1	HB-1620	1753.89	HB-1620	1807.96	HB-1215	0.57	HB-1371	0.56	HB-1566	1.16	HB-1797	637.75
2	HB-1357	1253.53	HB-1897	1181.85	HB-1207	0.55	HB-1372	0.56	HB-2002	1.13	HB-1897	629.27
3	HB-1897	1186.26	HB-1743	1170.68	HB-1552	0.54	HB-1353	0.50	HB-1954	1.13	HB-1885	447.11
4	HB-1255	1128.18	HB-1558	1083.58	HB-1419	0.49	HB-1319	0.44	HB-1301	1.08	HB-1128	400.60
5	HB-1743	1103.18	HB-1255	1070.22	HB-1566	0.49	HB-1366	0.44	HB-1623	0.96	HB-1785	396.80
6	HB-1954	1083.64	HB-1785	1059.89	HB-1427	0.48	HB-1520	0.44	HB-1348	0.96	HB-1747	353.95
7	HB-1785	1006.09	HB-1797	1030.22	HB-1629	0.48	HB-1716	0.44	HB-1366	0.95	HB-1928	327.43
8	HB-1372	960.75	HB-1617	995.75	HB-2028	0.48	HB-1063	0.44	HB-1452	0.92	HB-1428	322.79
9	HB-1908	956.76	HB-1372	991.44	HB-1511	0.47	HB-1311	0.44	HB-1552	0.92	HB-1943	303.97
10	HB-2002	954.71	HB-1885	965.37	HB-1147	0.46	HB-1811	0.44	HB-1111	0.85	HB-1546	301.67
$\bar{X}$		1540.61		1790.55		3.57		1.81		2.29		767.86
Rank	**%L**	**GV**	**L:S**	**GV**	**CP**	**GV**	**NDF**	**GV**	**IVD**	**GV**	**LIG.S**	**GV**
1	HB-1245	7.69	-	0.00	HB-1089	2.23	HB-1773	-4.37	HB-1683	5.15	HB-1563	-0.51
2	HB-2035	7.21	-	0.00	HB-1744	2.20	HB-1173	-3.08	HB-1479	5.03	HB-1773	-0.44
3	HB-1385	7.12	-	0.00	HB-1730	2.20	HB-1524	-2.92	HB-1488	4.28	HB-1685	-0.36
4	HB-1231	6.96	-	0.00	HB-1589	2.17	HB-1512	-2.92	HB-1577	4.22	HB-1429	-0.34
5	HB-1394	6.90	-	0.00	HB-1983	2.14	HB-1416	-2.67	HB-1582	4.10	HB-1997	-0.33
6	HB-1177	6.70	-	0.00	HB-1991	2.12	HB-1441	-2.61	HB-1588	4.01	HB-1798	-0.32
7	HB-1591	6.61	-	0.00	HB-1752	2.09	HB-1230	-2.55	HB-1674	3.99	HB-1191	-0.31
8	HB-1784	6.57	-	0.00	HB-1997	2.06	HB-1407	-2.42	HB-1961	3.93	HB-1981	-0.30
9	HB-1547	6.35	-	0.00	HB-1784	2.02	HB-1419	-2.42	HB-1374	3.88	HB-1416	-0.30
10	HB-1794	6.22	-	0.00	HB-1665	2.00	HB-1421	-2.32	HB-1970	3.83	HB-1094	-0.28
$\bar{X}$		66.92		2.57		15.75		64.41		73.32		2.17

FGW: Field Green Weight (kg.ha^{-1}); TDM: Total Dry Matter (kg.ha^{1}); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDW: Leaf dry matter (kg.ha^{-1}); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%).

Table 8 – Genotypic value (GV) of the ten best *Brachiaria spp*. hybrids selected by Mulamba e Mock index (MMI) using the first two principal components (MMI_{PC}), mean genotypic value of the selected hybrids by principal components (*GV.PC*), mean genotipic value of selected hybrids using MMI based only in phenotypic information (*GV.MM*), mean genotypic value of selected hybrids by univariate analysis (*GV.Uni*), broad sense heritability estimate (h^2), selection gain by MMI_{PC} considering 1% of selection intensity (*SG.PC*), selection gain by phenotypic MMI_{Uni} considering 1% of selection intensity (*SG.MM*), selection gain by univariate analysis considering 1% of selection intensity (*SG.Uni*), relative selection gain given by *SG.CP*/ *SG.Uni* (r*SG.CP*), relative selection gain given by *SG.MM*, *SG.Uni* (*rSG.MM*) and general mean ($\bar{X}$)

MMI_{PC}	FGW*	TDM	DEN	VEL	REG*	LDW	%L	L:S	CP*	NDF*	IVD	LIG.S
HB-2002	954.71	874.99	0.33	0.38	1.13	0.00	0.00	0.00	0.52	0.12	-1.59	0.26
HB-1954	1083.64	924.72	0.43	0.38	1.13	0.00	0.00	0.00	1.70	-0.04	1.44	-0.01
HB-1620	1753.89	1807.96	0.08	-0.05	0.15	0.00	0.00	0.00	0.04	0.22	-0.05	-0.01
HB-1357	1253.53	931.12	0.25	0.40	0.56	0.00	0.00	0.00	0.27	1.00	-1.97	-0.03
HB-1908	956.76	839.45	0.25	0.40	0.51	0.00	0.00	0.00	0.33	0.37	-1.94	0.27
HB-1897	1186.26	1181.85	-0.05	0.19	0.22	629.27	1.44	0.00	0.15	-0.87	-1.54	0.09
HB-1871	937.62	845.44	0.19	0.18	0.38	0.00	0.00	0.00	0.54	0.67	-0.40	0.05
HB-1623	741.01	749.57	0.40	0.31	0.96	0.00	0.00	0.00	1.49	-0.92	2.16	-0.07
HB-1629	667.64	651.08	0.48	0.19	0.70	184.95	2.09	0.00	-0.25	-0.52	1.03	0.06
HB-1579	563.55	601.18	0.28	0.44	0.65	146.62	1.99	0.00	0.98	0.51	0.78	0.12
GV.PC	1009.86	940.74	0.26	0.28	0.64	96.08	0.55	0.00	0.58	0.05	-0.21	0.07
SG.PC	817.53	733.16	0.13	0.18	0.47	53.46	0.17	0.00	0.39	0.03	-0.10	0.03

FGW: Field Green Weight ($kg.ha^{-1}$); TDM: Total Dry Matter ($kg.ha^{1}$); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDW: Leaf dry matter ($kg.ha^{-1}$); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%). *Traits used to direct selection in *MMI*.

Cont. **Table 8 –** Genotypic value (GV) of the ten best *Brachiaria spp*. hybrids selected by Mulamba e Mock index (MMI) using the first two principal components (MMI_{PC}), mean genotypic value of the selected hybrids by principal components (*GV.PC*), mean genotypic value of selected hybrids using MMI based only in phenotypic information (*GV.MM*), mean genotypic value of selected hybrids by univariate analysis (*GV.Uni*), broad sense heritability estimate (h^2), selection gain by MMI_{PC} considering 1% of selection intensity (*SG.PC*), selection gain by phenotypic MMI_{Uni} considering 1% of selection intensity (*SG.MM*), selection gain by univariate analysis considering 1% of selection intensity (*SG.Uni*), relative selection gain given by *SG.CP*/ *SG.Uni* (r*SG.CP*), relative selection gain given by *SG.MM*/ *SG.Uni* (*rSG.MM*) and general mean ($\bar{X}$)

MMI_{Uni}	FGW*	TDM	DEN	VEL	REG*	LDW	%L	L:S	CP*	NDF*	DIV	LIG.S
HB-1419	574.27	466.83	0.49	0.19	0.68	0.00	0.00	0.00	1.61	-2.42	2.85	-0.12
HB-1623	741.01	749.57	0.40	0.31	0.96	0.00	0.00	0.00	1.49	-0.92	2.16	-0.07
HB-2012	511.44	585.50	0.16	0.19	0.38	0.00	0.00	0.00	1.21	-1.85	1.34	0.04
HB-1366	470.19	361.33	0.18	0.44	0.95	0.00	0.00	0.00	1.24	-0.86	1.31	-0.12
HB-1964	523.00	574.90	0.27	0.19	0.59	0.00	0.00	0.00	1.25	-0.96	1.47	-0.04
HB-1863	499.87	442.51	-0.04	0.32	0.54	0.00	0.00	0.00	0.67	-1.44	0.56	-0.03
HB-1764	141.36	129.86	0.27	0.07	0.43	0.00	0.00	0.00	1.82	-2.28	3.16	-0.03
HB-1546	284.19	254.93	0.25	0.06	0.54	301.67	5.21	0.00	0.98	-1.58	1.34	-0.11
HB-1954	1083.64	924.72	0.43	0.38	1.13	0.00	0.00	0.00	1.70	-0.04	1.44	-0.01
HB-1984	254.37	196.12	0.23	0.05	0.33	81.22	1.31	0.00	1.47	-1.84	0.66	-0.08
GV.MM	508.34	468.63	0.26	0.22	0.65	38.29	0.65	0.00	1.34	-1.42	1.63	-0.06
SG.MM	411.52	365.23	0.13	0.14	0.48	21.31	0.20	0.00	0.91	-0.69	0.82	-0.02
GV.Uni	1138.70	1135.70	0.50	0.47	1.01	412.13	6.83	0.00	2.12	-2.83	4.24	-0.35
SG.Uni	921.83	885.11	0.26	0.29	0.75	229.32	2.12	0.00	1.43	-1.38	2.14	-0.15
rSG.PC	0.89	0.83	0.53	0.60	0.64	0.23	0.08	0.00	0.27	-0.02	-0.05	-0.21
rSG.MM	0.45	0.41	0.53	0.47	0.65	0.09	0.10	0.00	0.63	0.50	0.38	0.16
h^2	0.81	0.78	0.51	0.62	0.74	0.56	0.31	0.00	0.68	0.49	0.51	0.42
$\bar{X}$	1540.61	1790.55	3.57	1.81	2.29	767.86	66.92	2.57	15.75	64.41	73.32	2.17

FGW: Field Green Weight (kg.ha^{-1}); TDM: Total Dry Matter (kg.ha^{1}); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDW: Leaf dry matter (kg.ha^{-1}); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%). *Traits used to direct selection in *MMI*.

Table 9 – Direct selection gain ($SG_D\%$), in percentage, considering 1% of selection intensity (10 superior hybrids selected in univariate analysis – Table 7) – estimates placed in the diagonal of the table with values in bold. Indirect selection gain ($SG_I\%$), in percentage, considering 1% of selection intensity – estimates placed above and below the diagonal of the table

	FGW	**TDM**	**DEN**	**VEL**	**REG**	**LDW**	**%L**	**L:S**	**CP**	**NDF**	**IVD**	**LIG.S**
FGW	**59.84**	47.24	2.22	8.97	16.34	9.23	-0.10	0.00	0.63	0.55	-0.92	2.52
TDM	56.29	**49.43**	0.98	6.04	10.01	17.09	-0.32	0.00	-0.79	0.65	-1.01	3.11
DEN	12.81	10.10	**7.15**	3.23	24.25	2.65	0.28	0.00	-0.41	-0.20	-0.03	-0.97
VEL	18.41	15.54	-1.32	**16.19**	15.86	0.00	0.00	0.00	-1.52	0.59	-0.59	1.52
REG	19.51	14.54	5.51	9.79	**32.52**	1.76	0.23	0.00	-0.15	0.27	0.00	-0.08
LDW	37.91	34.39	1.27	6.32	11.38	**29.86**	0.42	0.00	-0.44	0.30	-0.59	2.26
%L	-5.26	-1.40	1.75	0.15	7.85	-2.15	**3.17**	0.00	1.64	-0.58	0.40	-1.03
L:S	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00	0.00
CP	-10.52	-7.54	-0.32	-6.20	-9.69	-7.63	-0.17	0.00	**9.09**	-0.83	1.03	-1.79
NDF	-14.88	-14.57	-0.57	-4.37	-8.85	-1.99	-0.40	0.00	2.37	**-2.15**	0.84	-3.19
IVD	-10.51	-9.86	-0.98	-1.84	-7.77	-3.82	-0.54	0.00	4.48	-0.51	**2.92**	-1.87
LIG.S	-22.66	-19.21	-0.61	-6.50	-8.27	-6.98	-0.61	0.00	0.91	-1.07	0.34	**-6.72**

FGW: Field Green Weight ($kg.ha^{-1}$); TDM: Total Dry Matter ($kg.ha^{1}$); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDW: Leaf dry matter ($kg.ha^{-1}$); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%).

Table S1 - Additive variance (σ_A^2), dominant variance (σ_D^2) and general mean ($\bar{X}$) for agronomic and nutritional value traits based on individual cuts of *Brachiaria* spp. hybrids

Trait	Cut	$\bar{X}$	σ_A^2	σ_D^2	% σ_A^2	% σ_D^2
FGW	1	1084.98	49240.15	3785.03	47.99	3.69
	2	675.92	36902.64	0.00	66.72	0.00
	3	591.95	23052.30	4834.71	54.53	11.44
	4	1287.47	51317.24	14237.91	44.96	12.47
	5	1194.32	0.17	88776.60	0.00	67.26
	6	1065.60	109389.06	0.00	74.03	0.00
	7	431.71	10969.74	0.00	66.73	0.00
TDM	1	1369.52	64918.31	25456.30	38.88	15.25
	2	690.32	32502.95	0.00	65.61	0.00
	3	681.20	18934.35	31607.83	27.85	46.49
	4	1321.06	51737.96	16695.75	45.57	14.71
	5	1556.59	0.26	139210.47	0.00	68.15
	6	1212.89	132436.55	0.00	74.47	0.00
	7	583.11	14019.44	0.00	61.84	0.00
DEN	3	2.99	0.00	0.03	0.00	42.29
	4	2.69	0.00	0.06	0.00	61.18
	5	2.47	0.01	0.02	10.79	29.57
	6	2.51	0.00	0.02	6.51	26.21
	7	2.49	0.03	0.00	41.43	0.00
VEL	3	1.55	0.02	0.00	45.22	0.00
	4	1.46	0.05	0.00	64.33	0.00
	5	0.87	0.08	0.00	70.67	0.00
	6	0.76	0.02	0.00	40.99	0.00

	7	0.93	0.01	0.02	17.25	37.36
REG	1	1.60	0.02	0.10	12.41	51.12
	2	2.42	0.00	0.29	0.00	81.89
	3	3.53	0.00	0.11	0.00	59.61
	4	3.15	0.01	0.11	4.84	62.85
	5	2.34	0.13	0.00	65.59	0.00
	6	2.27	0.05	0.00	44.90	0.00
	7	2.42	0.02	0.00	26.32	0.00
LDM	2	443.99	13532.02	0.00	45.88	0.00
	3	407.86	16275.31	0.00	61.72	0.00
%L	2	62.86	0.00	138.96	0.00	76.36
	3	73.80	0.00	32.03	0.00	49.10
L:S	2	2.92	0.05	2.08	1.88	71.75
	3	10.60	2.64	0.00	4.03	0.00
CP	3	15.44	0.59	0.00	68.43	0.00
	4	15.57	0.26	0.00	58.53	0.00
NDF	3	63.34	0.21	0.65	13.66	43.33
	4	66.53	0.71	0.00	56.02	0.00
IVD	3	72.33	0.28	2.33	6.12	51.63
	4	69.30	1.96	0.00	61.21	0.00
LIG.S	3	2.07	0.00	0.01	0.00	42.59
	4	2.18	0.00	0.01	0.64	47.35

FGW: Field Green Weight ($kg.ha^{-1}$); TDM: Total Dry Matter ($kg.ha^{1}$); DEN: Density of regrown tiller; VEL: Regrowth speed; REG: Regrowth capacity; %L: Percentage of leaves; LDM: Leaf dry matter ($kg.ha^{-1}$); L:S: Leaf:stem ratio; CP: Crude protein (%); NDF: Neutral detergent fiber (%); IVD: in vitro organic matter digestibility (%); LIG.S: Lignin in sulfuric acid (%).

3. DIPLOIDIZED MARKER DATA FROM GENOTYPING-BY-SEQUENCING OF THE TETRAPLOID FORAGE GRASS *Urochloa*

ABSTRACT

Although genotyping-by-sequencing (GBS) is a well-established marker technology in diploids, the development of best practices for tetraploid species is a topic of current research. We determined the theoretical relationship between read depth and expected genotype quality (GQ) for tetraploid vs. diploidized genotype calls. Assuming no error, 11 reads are sufficient to classify tetraploid samples as heterozygous vs. homozygous with 95% accuracy, compared with 61 reads to determine allele dosage. One response to this challenge in previous studies has been the use of diploid models, but their indiscriminate application to tetraploid samples can produce suboptimal genotype calls and incorrect GQ scores. We developed an R script to convert tetraploid GBS data in Variant Call Format (VCF) into diploidized genotype calls and applied it to 267 interspecific hybrids of the tetraploid forage grass *Urochloa*. When reads were aligned to a 'mock' reference genome, created from GBS data of the cultivar 'Marandu', 46,147 bi-allelic SNPs were discovered, compared to less than 6000 SNPs when aligning to the closest true reference genomes, *Setaria viridis* and *S. italica*. We recommend using thresholds for both sample depth and GQ to eliminate poor quality reads without introducing genotype bias, which reduced the number of SNPs with less than 50% missing data to 23,936 in the *Urochloa* panel. Cross-validation revealed that missing genotypes were imputed with a median accuracy of 0.85, regardless of heterozygote frequency, to produce a complete genotype matrix that will be used for genome-wide prediction and association analysis.

Keywords: *Brachiaria*; Genotype quality; SNP; Randon forest algorithm; Missing data

First draft submitted for review in the "Plant Genome" journal.

3.1. INTRODUCTION

Urochloa is the most cultivated genus as pasture on tropical livestock farms due to its tolerance to acidic soils, good carrying capacity, insect resistance, and nutritional value (Jank et al., 2014; Pessoa-Filho et al., 2017). The most economically important species are *U. decumbens* (syn. *Brachiaria decumbens*) and *U. brizantha* (syn. *B. brizantha*), which are both tetraploid (2n = 4x = 36). Apomixis is the normal mode of reproduction in these species, and for many years genetic

improvement in South America was based on screening new introductions from Africa (Miles 2007; Jank et al., 2011). To facilitate breeding by sexual hybridization, Swenne et al. (1981) utilized colchicine-induced tetraploids of the diploid species *U. ruziziensis* (2n = 2x = 18) as female parents to cross with apomictic tetraploids. This interspecific hybridization scheme has become the foundation of the *Urochloa* breeding programs at CIAT and EMBRAPA (Lutts et al., 1991; Miles et al. 2006).

As in other crops, genome-wide markers have the potential to provide significant value for *Urochloa* breeding programs. Several previous studies have utilized microsatellite markers to study population structure in *Urochloa* (Jungmann et al., 2009, 2010; Vigna et al., 2011; Silva et al., 2013), but the ubiquity and cost-effectiveness of SNPs are advantageous for discovering genetic variants and predicting complex traits. Both arrays and genotyping-by-sequencing (GBS) of multiplexed, reduced-representation libraries have been utilized to generate large, bi-allelic SNP datasets in heterozygous tetraploids, including potato (Felcher et al. 2012; Uitdewilligen et al. 2013), alfalfa (Li et al. 2014a; Li et al. 2014b), rose (Koning-Boucoiran et al. 2015), kiwi (Melo et al. 2016), and *Urochloa* (Worthington et al. 2016; Ferreira et al. 2018). Both arrays and GBS generate a signal for each allele that can be used to predict allele dosage, i.e., the tetraploid or other polyploid genotypes. For the SNP array, signal intensity is not necessarily proportional to allele dosage, and therefore different classification algorithms have been explored (Voorrips et al. 2011; Serang et al. 2012; Schmitz Carley et al. 2017).

For GBS data, the allele signal intensity is the read count, which can be analyzed using the aforementioned classifiers, but the focus of this manuscript is genotype calling based on a binomial model. The binomial model is central to the well-established software packages GATK (McKenna et al., 2010; Depristo et al., 2011) and FreeBayes (Garrison and Marth 2012), as well as more recent contributions, such as R package *updog* (Gerard et al. 2018). It is generally recognized that higher read depth is needed to estimate allele dosage in polyploids, but precise guidelines are lacking. Uitdewilligen et al. (2013) developed KASP assays for 270 GBS markers in potato and compared the genotype calls from the two methods; the results under different filtering criteria led the authors to conclude that "~60-80X can be used as a lower boundary for reliable assessment of allele copy number…." Bastien et al. (2018) used a threshold of 53 reads for determining allele dosage because it was deemed "sufficient to distinguish between the five expected genotypic classes based on a chi-square distribution."

Our first objective was to use probability theory to clarify the relationship between read depth and genotype quality (GQ) in tetraploids. GQ is a standard metric in the FORMAT field of the VCF

file and defined as $-10\log_{10}(q)$, where q is the probability that the genotype call is erroneous (Danecek et al., 2011). Theoretical results were used to guide the analysis of GBS data for a panel of 267 interspecific *Urochloa* hybrids. Because very few markers had sufficient read depth to determine allele dosage with reasonable accuracy, genotype calls were made using a diploid approximation, in which the three heterozygotes were not distinguished. This approximation is common for GBS in heterozygous tetraploids, and typically a threshold of 11 reads is used to ensure the probability of misclassifying a heterozygote as homozygous (due to sampling error) is less than 5% (Li et al. 2014b). However, this threshold is based on the assumption of no error, and our theoretical treatment elucidates how the threshold increases with sequencing error.

Even with a diploid approximation, the *Urochloa* GBS dataset contained missing data. Imputation of missing genotypes in GBS datasets has been studied extensively in inbred lines and heterozygous diploids, with Hidden Markov Models being the preferred method when a genetic or physical map for the markers is available (Hickey et al. 2012; Swarts et al. 2014; Fragoso et al. 2015). When a map is not available, as was the case for our *Urochloa* dataset, the Random Forest algorithm (Breiman 2001) can still be used and has performed well in other species (Poland et al. 2012; Rutkoski et al. 2013; Money et al. 2015). Our objective was to evaluate the genotyping-by-sequencing data using different filtering thresholds and references genomes, also to use cross-validation to determine the imputation accuracy of Random Forest in the *Urochloa* dataset.

3.2. MATERIALS AND METHODS

3.2.1. Expected Genotype Quality

A binomial model was used to determine the statistical relationship between read depth and expected genotype quality (EGQ) for a particular genotypic class, such as 'simplex' for tetraploid genotypes or 'heterozygous' for diploidized genotypes. Let $f(k, N, \rho)$ denote the probability mass function for the binomial distribution with k successes out of N trials and success probability ρ. The likelihood of observing k reads of the alternate allele given N total reads for tetraploid genotype $i \in \{0,1,2,3,4\}$ was modeled as $f_i \equiv f\left(k, N, \rho_i = \frac{i}{4}[1-\varepsilon] + \left[1-\frac{i}{4}\right]\varepsilon\right)$, where the error rate ε is the probability that a read is generated by one allele but counted toward the other (due to errors in PCR or sequencing). Under a uniform prior, the maximum a posteriori (MAP) tetraploid genotype call for the observed result (k,N) is the value of i that maximizes f_i. For some values of k, the MAP solution

does not equal the true value. Summing *f* over these values of *k*, and expressing the result on the phred scale, is EGQ for tetraploid genotype *x* at read depth *N*:

$$\mathrm{EGQ}_{\mathrm{tet}}(x,N) = -10\log_{10}\sum_{k=0}^{N} f(k,N,\rho_x)[1-\delta(x,\mathrm{MAP}_{\mathrm{tet}})] \qquad [1]$$

The symbol δ in Eq. 1 is the Kronecker delta function, which equals 1 when its two arguments are equal and 0 when they are unequal.

For diploidized genotype calls, the three possible genotypic states are denoted $\{A,H,B\}$, where the heterozygous state *H* = dosages 1, 2, or 3, and the homozygous states *A* = dosage 0 and *B* = dosage 4. The corresponding 3-vector of posterior probabilities is proportional to $(p_A,p_H,p_B) \equiv (f_0, f_1+f_2+f_3, f_4)$, and the MAP solution for the observed result (*k*,*N*) is the value of *j* that maximizes p_j. For some values of *k*, the MAP solution does not equal the diploidized genotype *y* corresponding to the true tetraploid state *x*. Summing *f* over these values of *k*, and expressing the result on the phred scale, is $\mathrm{EGQ}_{\mathrm{dip}}$ for tetraploid genotype *x* at read depth *N*:

$$\mathrm{EGQ}_{\mathrm{dip}}(x,N) = -10\log_{10}\sum_{k=0}^{N} f(k,N,\rho_x)\left[1-\delta(y,\mathrm{MAP}_{\mathrm{dip}})\right] \qquad [2]$$

Although Eq. 1 and 2 tend to increase with read depth, they are not monotone functions of *N* (Figure S1). The results we present for EGQ actually correspond to the monotone extension $\phi_{i,N} = \min_{M\geq N} \mathrm{EGQ}_{i,M}$, which has the property that $\phi_{i,N} \geq \phi_{i,M}$ when $N > M$.

3.2.2. *Urochloa* GBS

Genomic DNA was extracted using the Qiagen DNeasy kit for 267 tetraploid *U. ruziziensis* x *U. brizantha* hybrids from Embrapa Beef Cattle, as well as for the *Urochloa* cultivar 'Marandu'. GBS libraries were prepared according to Elshire et al. (2011), using the ApeKI enzyme and sequenced on five lanes of the Illumina Hi-Seq 2500 platform with 1x100 bp reads. Reads were demultiplexed and trimmed using *Cutadapt* (Marin 2011) and then aligned to five different *Poaceae* genomes with *bwa-mem* (Li and Durbin 2009): *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), and *Zea mays* (Zm) (Schnable et al., 2009). Reads were also aligned to a *Urochloa* "mock" reference genome (available at *dryad*), generated from the reads for 'Marandu' with the GBS-SNP-CROP pipeline (Melo et al., 2016). *SAMtools* (Li et al., 2009; Li, 2011) and *Picard* (http://broadinstitute.github.io/picard/) were used to mark duplicate reads, organize and combine files, respectively. The *Genome Analysis Toolkit,* or GATK

(McKenna et al., 2010; Depristo et al., 2011), was used for variant discovery with the recommended hard filters and tetraploid genotype calling. The alignment percentage for each reference was evaluated with *Bowtie2* (Langmead and Salzberg, 2012).

The VCF file from GATK (available at *dryad*) was processed using R scripts (R Development Core Team 2017) to analyze the results and perform additional filtering. Only bi-allelic SNPs were retained. The VCF file includes variants relative to the reference genome, regardless of whether they are polymorphic in the genotyped population. To identify polymorphic markers, the total number of reads for the minor allele, or minor allele depth (MAD), was calculated for each marker based on the AD field, and variants with MAD < 2 were removed. GATK calculates allele frequency based on the dosage of called genotypes, which was deemed unreliable due to low read depth. A suitable proxy for filtering that does not require allele dosage information is the frequency of genotypes homozygous for the major allele, which was capped at 0.99. For each sample, GATK provides the phred-scaled likelihood (PL) for each of the 5 tetraploid genotypes, which was converted into a posterior probability p_i for genotype $i \in \{0,1,2,3,4\}$ (assuming a uniform prior) by

$$p_i = \frac{10^{-\mathrm{PL}_i/10}}{\sum_{i=0}^{4} 10^{-\mathrm{PL}_i/10}} .$$

The tetraploid genotype call corresponds to the largest probability, and $\mathrm{GQ}_{\mathrm{tet}} = -10\log_{10}\left(1 - \max_i p_i\right)$.

Due to the low read depth per sample in the *Urochloa* dataset, diploidized genotype calls were made in which the three heterozygous genotypes were not differentiated. This corresponds to defining a new vector of posterior probabilities, $\tilde{\mathbf{p}} = (p_0,\ p_1 + p_2 + p_3,\ p_4)$, in which the probability of the heterozygous state is the sum of the probabilities for the simplex, duplex, and triplex genotypes. The diploidized genotype call corresponds to the largest probability, and $\mathrm{GQ}_{\mathrm{dip}} = -10\log_{10}\left(1 - \max_i \tilde{p}_i\right)$.

Missing genotypes were imputed with the R package *randomForest* (Liaw and Wiener, 2002), which is based on the algorithms in Breiman (2001). For each marker, a training set of 100 clones was randomly selected from the clones with genotypes, and all other clones with genotype data were used for validation. Because each marker had no more than 50% missing data, this ensured at least 33 clones were available for validation. 300 classification trees were used for prediction, and all markers with $r^2 \geq 0.1$ were used as m potential predictors. We used the default setting of randomly

sampling $\sqrt{m}$ preditors at each split. Classification accuracy is the proportion of clones in the validation set for which the predicted genotype is correct.

3.3. RESULTS

Expected Genotype Quality

A binomial model was used to determine the statistical relationship between read depth and expected genotype quality (EGQ) for a particular genotypic class, such as 'simplex' for tetraploid genotypes or 'heterozygous' for diploidized genotypes. EGQ involves the expectation over all possible allele counts at a particular depth, whereas GQ corresponds to a particular allele count. In addition to read depth, the other key parameter affecting EGQ is the error rate, defined as the probability that a read is generated by one allele but counted toward the other, due to errors during PCR or sequencing. Since EGQ is reported on the phred scale, a score of 13 corresponds to 95% accuracy, and a score of 20 corresponds to 99% accuracy.

Figure 1 shows how EGQ differs for simplex vs. duplex genotypes, as well as when allelic dosage is estimated (blue) vs. diploidized calls (green). Higher accuracy is achieved for simplex compared to duplex samples when allelic dosage is determined, but under diploidized genotype calling the reverse is true. The intuitive reason for this result is that a duplex genotype can appear as either simplex or triplex due to sampling variation, but comparable uncertainty for the simplex genotype exists only in the direction of higher dosage (i.e., with the duplex). If dosage is not determined, however, then simplex genotypes are more readily confounded with nulliplex homozygotes than duplex samples are with either homozygote. In the absence of error (solid lines), 11 reads are needed to make diploidized genotype calls with 95% accuracy, compared with 61 reads for tetraploid genotypes. As Figure 1 shows, allelic errors have a greater effect on EGQ_{dip} than EGQ_{tet}. With 1% error (dashed lines), the minimum depth needed to achieve 95% accuracy for diploidized genotypes increases to 17 reads, while the minimum depth for tetraploid genotypes increases to 63 reads.

GBS of *Urochloa* hybrids

As no reference genome for *Urochloa* was available, the reference genomes of five other *Poaceae* species were evaluated for alignment. Figure 1 shows the number and percentage of aligned reads from the ApeKI-reduced representation of the *Urochloa brizantha* cultivar 'Marandu.' The percentage

of reads aligned was low for all genomes, ranging from 1.92% for *Oryza sativa* to 7.88% for *Setaria italica*. For both *Setaria* species and *Sorghum bicolor*, over 3/4 of the aligned reads mapped to a unique location. For *Oryza sativa* and *Zea mays*, this proportion decreased to 1/2. The same five genomes were compared with respect to variant discovery in a panel of 267 interspecific tetraploid *Urochloa* hybrids. After removing variants with median depth < 8, the two *Setaria* species generated the most bi-allelic SNPs, in the range 5000–6000 (Table 1).

To better utilize the GBS reads, a 'mock' reference genome was built by clustering the (trimmed) reads from cv. Marandu into 1,309,910 non-redundant, consensus sequences, or "centroids" (Melo et al. 2016). When the GBS reads for the 267 interspecific hybrids were aligned to the centroids, the number of bi-allelic SNPs with median depth ≥ 8 increased to 46,147 (Table 1). A highly repetitive sequence was detected in the centroids, for which the first 50 bp. The entire 50 bp sequence was present in 3.3% of the centroids, and when truncated to the first 40 or 30 bp, the frequency increased to 8.5% and 14.9%, respectively. The repetitive sequence was also detected in all 267 *Urochloa* hybrids. A nucleotide BLAST search of the 50 bp sequence against the NCBI database returned highly significant matches to a diverse set of species, including *Larimichthys crocea* and *Cyprinus carpio* (100% identity across 49 bp), *Triticum aestivum* and *Solanum pennellii* (98% identity across 50 bp).

Our initial depth threshold of 8 reads corresponds to $EGQ_{dip} \geq 10$, assuming 0% allelic error. Of the 46,147 bi-allelic SNPs with median depth above this threshold (using the mock reference for alignment), 23% were classified as "rare" alleles based on the presence of less than 5% heterozygosity (distribution in Figure 3). According to the binomial model, a depth threshold of 42 is needed for $EGQ_{tet} \geq 10$ (assuming 0% error). As only 1895 SNPs had median depth ≥ 42, tetraploid genotype calls were not pursued.

Figure 4 shows the distribution of GQ_{dip} scores for all 343,977 genotypes with depth = 8 in the filtered dataset. The sharp peak at $GQ_{dip} = 10$ for homozygous genotypes is consistent with our theoretical result that, for 8 reads, simplex genotypes constrain EGQ_{dip} at 10. By contrast, the distribution of GQ_{dip} for heterozygous genotypes is bimodal, with just under half of the samples over 20 and the other half below 7. The difference between these groups is that, in the former (high GQ) the minor allele count is ≥ 1, while for the latter (low GQ) the minor allele count is 0. Although this seems inconsistent with a heterozygous genotype call, the output from the GATK Haplotype Caller is not based solely on read counts (e.g., base and mapping quality scores are also used). To remove

these problematic calls, genotypes with either $GQ_{dip} < 10$ or depth < 8 were set to NA, leaving 23,936 SNPs with less than 50% missing data (Table 1).

LD and Genotype Imputation

The success of genotype imputation depends on the amount of linkage disequilibrium (LD) between markers, which is often quantified by the physical distance at which r^2 (the squared correlation) drops below some threshold. However, since a physical reference genome is unavailable for *Urochloa*, LD was quantified based on the maximum r^2 for each SNP. Figure 5 (left panel) shows the distribution of r^2_{max} for the 3,154 SNPs from the filtered dataset that are 25–75% heterozygous, to capture a range of difficulty for imputation. The median value of r^2_{max} was approximately 0.5 for heterozygote frequencies below 0.5 but gradually decreased as the proportion of heterozygotes increased toward 0.75.

Cross-validation accuracy was determined with a training set of 100 clones, selected at random from all clones with genotype data for a particular marker. The accuracy shown in Figure 5 (right panel) is the proportion of predicted values equal to the masked value. The results are binned by heterozygote frequency, with the median accuracy shown by a solid line and the first and third quartiles by dashed lines. Imputation with the population mode is a simple baseline method that, by definition, has lower accuracy as the frequency of the modal genotype declines (Figure 5). By contrast, the Random Forest method was largely unaffected by heterozygote frequency, with a median accuracy of approximately 0.85.

3.4. DISCUSSION

GBS in tetraploids

As mentioned in the introduction, there has been variation in the filtering criteria used in previous studies involving GBS of tetraploids. One of the most cited is Uitdewilligen et al. (2013), who recommended 60–80X to determine allele dosage. Our theoretical calculations indicate this range corresponds to 95–98% genotype accuracy for allelic error rates below 1%. For diploidized genotype calling, the threshold of 11 reads in Li et al. (2014b) has been commonly used by others, which corresponds to 95% accuracy in the absence of error. Allowing for some error, and recognizing that 98% accuracy is a better goal for genotyping, our theoretical calculations indicate 15–20 reads is a more appropriate target for diploidized genotypes.

The difference in expected genotype quality for simplex (or triplex) vs. duplex genotypes has important implications for filtering GBS data. Setting a minimum GQ value will create bias against duplex samples when calling tetraploid genotypes, and against simplex/triplex genotypes with diploidized genotypes. Using a depth threshold corresponding to the desired minimum EGQ does not introduce this bias, but this does not address the issue of reads with low base or mapping quality. A combination of the two approaches seems best, using a depth threshold with EGQ equal to the GQ threshold.

The aforementioned considerations are appropriate for genotype calling based on the posterior mode. An alternative approach is to estimate allele dosage based on the posterior mean, which produces fractional genotype calls (Ashraf et al. 2014; Sverrisdóttir et al. 2017). Such data are suitable when additive models are used in association analysis and genome-wide prediction, but a number of genetic analyses require integral estimates of dosage, including linkage analysis (Hackett et al. 2013; Zheng et al. 2016), dominance effects (Rosyara et al. 2016; Endelman et al. 2018; Chen et al. 2018), and haplotype inference (Su et al. 2008; Neigenfind et al. 2008; Aguiar and Istrail 2013; Berger et al. 2014; Das and Vikalo 2015; Motazedi et al. 2017).

This study used the traditional approach of setting hard thresholds for genotype calling, followed by imputation of the missing data. We used depth and GQ thresholds to achieve 90% genotype accuracy and allowed for up to 50% missing data per marker, which were imputed with accuracy close to this value (75% of the samples had 80–90% accuracy). We did not explore the interplay between GQ threshold and imputation accuracy, but this is an interesting topic for future research. It seems appealing to select thresholds to achieve similar accuracy in the samples called based on allele counts vs. those that are imputed. Ultimately, the traditional two-step approach (threshold then impute) is suboptimal because the read counts for the missing genotypes are not utilized during imputation. For ordered markers, this limitation can be overcome by using Hidden Markov Models (HMMs) with read counts as the emission states. This approach has been used in diploid mapping populations (Fragoso et al. 2015; Bilton et al. 2018; Zheng et al. 2018) and can be extended to the HMMs developed for SNP array data in tetraploids (Hackett et al. 2013; Zheng et al. 2016). For unordered markers, alternative imputation methods need to be explored.

Molecular breeding in *Urochloa*

Few molecular studies with genotyping by sequencing (GBS) in interspecific hybrids of *Urochloa* are available in the literature. The repetitive sequence observed in the *Urochloa* mock

reference can mean repetitive elements on the genome as transposon or retrotransposon. These elements can be located near host genes and can regulate their activities for biotic or abiotic conditions (Bennetzen and Wang, 2014). The intention now is to carefully evaluate these sequences and understand their importance for the *Urochloa* towards their influence on the genomic behavior under stresses.

With the molecular data it will be possible to evaluate the identity of genotypes used as a progenitor in the crosses, as well as the pedigree information of the hybrids from the traditional breeding program in forages. Usually, the crosses are made in the field by placing the sexual genotype within the apomictic genotype plot, and only the seeds from sexual genotype are collected to develop the full sib progeny. This method is widely used, but the wind can bring pollen from different genotypes, which can lead to mistakes when constructing the pedigree. Furthermore, to the best of our knowledge, there are no reports on genomic prediction and genome-wide association studies for this species. Thus, research applying these tools can improve the knowledge about the genetic control of forage traits and the development of new cultivars.

In conclusion, we recommend the reduction in the number of genotypes per plate or on increase in the number of replicates of the same sample during the GBS sequencing. This approach is important for polyploid species to improve the depth of the reads and consequently to reduce the bias on the genotype call step. Also, the genotype quality filter is an interesting piece of information to select markers with high quality that account for the alignment step associated with depth of information.

REFERENCES

Anderson, C.A., F.H. Pettersson, G.M. Clarke, L.R. Cardon, A.P. Morris, and K.T. Zondervan. 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5:1564–1573. doi:10.1038/nprot.2010.116

Bennetzen, J.L., J. Schmutz, H. Wang, R. Percifield, J. Hawkins, A.C. Pontaroli, M. Estep, L. Feng, J.N. Vaughn, J. Grimwood, and others. 2012. Reference genome sequence of the model plant Setaria. Nat. Biotechnol. 30:555

Bennetzen, J.L., and H. Wang. 2014. The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. Annu. Rev. Plant Biol. 65:505–530. doi:10.1146/annurev-arplant-050213-035811

Boff, T., and M.T. Schifino-Wittmann. 2003. Segmental allopolyploidy and paleopolyploidy in species of *Leucaena benth*: Evidence from meiotic behaviour analysis. Hereditas 138:27–35. doi:10.1034/j.1601-5223.2003.01646.x

D'hoop, B.B., M.J. Paulo, K. Kowitwanich, M. Sengers, R.G.F. Visser, H.J. van Eck, and F.A. van Eeuwijk. 2010. Population structure and linkage disequilibrium unravelled in tetraploid potato. Theor. Appl. Genet. 121:1151–1170. doi:10.1007/s00122-010-1379-5

DaCosta, J.M., and M.D. Sorenson. 2014. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS One 9. doi:10.1371/journal.pone.0106713

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Depristo, M.A., E. Banks, R. Poplin, K. V. Garimella, J.R. Maguire, C. Hartl, A.A. Philippakis, G. Del Angel, M.A. Rivas, M. Hanna, A. McKenna, T.J. Fennell, A.M. Kernytsky, A.Y. Sivachenko, K. Cibulskis, S.B. Gabriel, D. Altshuler, and M.J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–501. doi:10.1038/ng.806

DOE-JGI. 2018a. *Setaria viridis* v1.1. http://phytozome.jgi.doe.gov/

DOE-JGI. 2018b. *Sorghum bicolor* v3.1. http://phytozome.jgi.doe.gov/

FAO. 2018. FAOSTAT. http://faostat3.fao.org/faostat-gateway/go/to/%0Ahome/E (accessed March 3, 2018).

Ferreira, R.C.U., L.A. de C. Lara, L. Chari, S.C.L. Barrios, C.B. do Valle, J.R. Valerio, F.Z.V. Torres, A.A.F. Garcia, and A.P. de Souza. 2018. Genetic mapping with allele dosage information in tetraploid *Urochloa decumbens* (Stapf) R.D. Webster reveals insights into spittlebug (*Notozulia entreriana* Berg) resistance. bioRxiv 360594. doi:10.1101/360594

Fu, Y.-B., and G.W. Peterson. 2011. Genetic Diversity Analysis with 454 Pyrosequencing and Genomic Reduction Confirmed the Eastern and Western Division in the Cultivated Barley Gene Pool. Plant Genome J. 4:226. doi:10.3835/plantgenome2011.08.0022

GATK. 2018. Understanding and Adapting the Generic Hard-Filtering Recommendations. https://software.broadinstitute.org/gatk/documentation/article.php?id=6925 (accessed March 3, 2018).

Giussani, L.M., J.H. Cota-Sánchez, F.O. Zuloaga, and E.A. Kellogg. 2001. A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C4 photosynthesis. Am. J. Bot. 88:1993–2012. doi:10.2307/3558427

Gómez-Martinez, R., A. Culham, and others. 2000. Phylogeny of the subfamily Panicoideae with emphasis on the tribe Paniceae: evidence from the trnL-F cpDNA region. Grasses Syst. Evol. 136–140

Grandin, T. 2015. Improving Animal Welfare: A Practical Approach. 2nd ed. T. Grandin, ed. CAB International, Boston, MA, USA.

Grandke, F., S. Ranganathan, A. Czech, J.R. de Haan, and D. Metzler. 2014. Bioinformatic Tools for Polyploid Crops. J. Agric. Sci. Technol. B 4. doi:10.17265/2161-6264/2014.08.001

Griffin, P.C., C. Robin, and A.A. Hoffmann. 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa grasses*. BMC Biol. 9:19. doi:10.1186/1741-7007-9-19

Hackenberg, D., M.R. McKain, S.G. Lee, S. Roy Choudhury, T. McCann, S. Schreier, A. Harkess, J.C. Pires, G.K.-S. Wong, J.M. Jez, and others. 2017. Gα and regulator of G-protein signaling (RGS) protein pairs maintain functional compatibility and conserved interaction interfaces throughout evolution despite frequent loss of RGS proteins in plants. New Phytol. 216:562–575

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

Huang, Y.F., J.A. Poland, C.P. Wight, E.W. Jackson, and N.A. Tinker. 2014. Using Genotyping-By-Sequencing (GBS) for genomic discovery in cultivated oat. PLoS One 9. doi:10.1371/journal.pone.0102448

Hyten, D.L., I.Y. Choi, Q. Song, R.C. Shoemaker, R.L. Nelson, J.M. Costa, J.E. Specht, and P.B. Cregan. 2007. Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937–1944. doi:10.1534/genetics.106.069740

Ishigaki, G., T. Gondo, M. Ebina, K. Suenaga, and R. Akashi. 2010. Estimation of genome size in *Brachiaria* species. Grassl. Sci. 56:240–242. doi:10.1111/j.1744-697X.2010.00200.x

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

Jank, L., C. Valle, and R. Resende. 2011. Breeding tropical forages. Crop Breed. Appl. Biotechnol. S1:27–34. doi:10.1590/S1984-70332011000500005

Jessup, R.W., B.L. Burson, G. Burow, Y.-W. Wang, C. Chang, Z. Li, A.H. Paterson, and M.A. Hussey. 2003. Segmental allotetraploidy and allelic interactions in buffelgrass (*Pennisetum ciliare* (L.) Link syn. *Cenchrus ciliaris* L.) as revealed by genome mapping. Genome 46:304–313. doi:10.1139/g03-005

Jungmann, L., B.B.Z. Vigna, K.R. Boldrini, a C.B. Sousa, C.B. do Valle, R.M.S. Resende, M.S. Pagliarini, M.I. Zucchi, and a P. de Souza. 2010. Genetic diversity and population structure analysis of the tropical pasture grass *Brachiaria humidicola* based on microsatellites, cytogenetics, morphological traits, and geographical origin.. Genome 53:698–709. doi:10.1139/g10-055

Jungmann, L., B.B.Z. Vigna, J. Paiva, A.C.B. Sousa, C.B. do Valle, P.R. Laborda, M.I. Zucchi, and A.P. de Souza. 2009. Development of microsatellite markers for *Brachiaria humidicola* (Rendle) Schweick. Conserv. Genet. Resour. 1:475–479. doi:10.1007/s12686-009-9111-y

Keller-Grein, G., B.L. Maass, and J. Hanson. 1996. Natural variation in *Brachiaria* and existing germplasm

Kim, S., C.S. Kim, J. Lee, I.Y. Lee, Y.J. Chung, M.S. Cho, and S.C. Kim. 2014. Phylogenetic relationships among species of *Setaria* (Paniceae; Panicoideae; Poaceae) in Korea: insights from nuclear (ITS and kn1) and chloroplast DNA sequence data. Plant Syst. Evol. 301:725–736.

Kraberger, S., S. Saumtally, D. Pande, M.H.R. Khoodoo, S. Dhayan, A. Dookun-Saumtally, D.N. Shepherd, P. Hartnady, R. Atkinson, F.M. Lakay, B. Hanson, D. Redhi, A.L. Monjane, O.P. Windram, M. Walters, S. Oluwafemi, J. Michel-Lett, P. Lefeuvre, D.P. Martin, and A. Varsani. 2017. Molecular diversity, geographic distribution and host range of monocot-infecting mastreviruses in Africa and surrounding islands. Virus Res. 238:171–178. doi:10.1016/j.virusres.2017.07.001

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

Laurie, C.C., K.F. Doheny, D.B. Mirel, E.W. Pugh, L.J. Bierut, T. Bhangale, F. Boehm, N.E. Caporaso, M.C. Cornelis, H.J. Edenberg, S.B. Gabriel, E.L. Harris, F.B. Hu, K.B. Jacobs, P. Kraft, M.T. Landi, T. Lumley, T.A. Manolio, C. McHugh, I. Painter, J. Paschall, J.P. Rice, K.M. Rice, X. Zheng, and B.S. Weir. 2010. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet. Epidemiol. 34:591–602. doi:10.1002/gepi.20516

Li, F., G. Fan, C. Lu, G. Xiao, C. Zou, R.J. Kohel, Z. Ma, H. Shang, X. Ma, J. Wu, X. Liang, G. Huang, R.G. Percy, K. Liu, W. Yang, W. Chen, X. Du, C. Shi, Y. Yuan, W. Ye, X. Liu, X. Zhang, W. Liu, H. Wei, S. Wei, G. Huang, X. Zhang, S. Zhu, H. Zhang, F. Sun, X. Wang, J. Liang, J. Wang, Q. He, L. Huang, J. Wang, J. Cui, G. Song, K. Wang, X. Xu, J.Z. Yu, Y. Zhu, and S. Yu. 2015. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat. Biotechnol. 33:524–530. doi:10.1038/nbt.3208

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., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 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 2:18–22. doi:10.1177/154405910408300516

Lupo, C.D., D.E. Clay, J.L. Benning, and J.J. Stone. 2013. Life-Cycle Assessment of the Beef Cattle Production System for the Northern Great Plains, USA. J. Environ. Qual. 42:1386. doi:10.2134/jeq2013.03.0101

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

Maass, B.L., C. a O. Midega, M. Mutimura, V.B. Rahetlah, P. Salgado, J.M. Kabirizi, Z.R. Khan, S.R. Ghimire, and I.M. Rao. 2015. Homecoming of *Brachiaria*: Improved hybrids prove useful for African animal agriculture. East African Agric. For. J. 81:71–78. doi:10.1080/00128325.2015.1041263

Manching, H., S. Sengupta, K.R. Hopper, S.W. Polson, Y. Ji, and R.J. Wisser. 2017. Phased Genotyping-by-Sequencing Enhances Analysis of Genetic Diversity and Reveals Divergent Copy Number Variants in Maize. G3: Genes|Genomes|Genetics 7:2161–2170. doi:10.1534/g3.117.042036

Mather, K.A., A.L. Caicedo, N.R. Polato, K.M. Olsen, S. McCouch, and M.D. Purugganan. 2007. The extent of linkage disequilibrium in rice (*Oryza sativa* L.). Genetics 177:2223–2232. doi:10.1534/genetics.107.079616

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 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

Mendes-Bonato, A.B., M.S. Pagliarini, F. Forli, C. Borges Do Valle, and M.I. De Oliveira Penteado. 2002. Chromosome numbers and microsporogenesis in *Brachiaria brizantha* (Gramineae). Euphytica 125:419–425. doi:10.1023/A:1016026027724

Miles, J.W. 2007. Apomixis for cultivar development in tropical forage grasses. Crop Sci. 47:S--238

Monteiro, L.C., J.R. Verzignassi, S.C.L. Barrios, C.B. do Valle, G. de L. Benteo, and C.B. de Libório. 2016. Characterization and selection of interspecific hybrids of *Brachiaria decumbens* for seed production in Campo Grande-MS. Crop Breed. Appl. Biotechnol. 16:174–181

Moskvina, V., N. Craddock, P. Holmans, M.J. Owen, and M.C. O'Donovan. 2006. Effects of differential genotyping error rate on the type I error probability of case-control studies. Hum Hered 61:55–64. doi:HHE2006061001055 [pii]\r10.1159/000092553

Nitthaisong, P., G. Ishigaki, H. Tanaka, and R. Akashi. 2016. Chromosome number, genomic variation, and molecular markers to assess genetic diversity of *Brachiaria* species. Crop Sci. 56:312–321. doi:10.2135/cropsci2015.04.0203

O'Callaghan, T.F., H. Faulkner, S. McAuliffe, M.G. O'Sullivan, D. Hennessy, P. Dillon, K.N. Kilcawley, C. Stanton, and R.P. Ross. 2016a. Quality characteristics, chemical composition, and sensory properties of butter from cows on pasture versus indoor feeding systems. J. Dairy Sci. 99:9441–9460. doi:10.3168/jds.2016-11271

O'Callaghan, T.F., D. Hennessy, S. McAuliffe, K.N. Kilcawley, M. O'Donovan, P. Dillon, R.P. Ross, and C. Stanton. 2016b. Effect of pasture versus indoor feeding systems on raw milk composition and quality over an entire lactation. J. Dairy Sci. 99:9424–9440. doi:10.3168/jds.2016-10985

Ouyang, S., W. Zhu, J. Hamilton, H. Lin, M. Campbell, K. Childs, F. Thibaud-Nissen, R.L. Malek, Y. Lee, L. Zheng, and others. 2006. The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res. 35:D883--D887

Pessoa-Filho, M., A.M. Martins, and M.E. Ferreira. 2017. Molecular dating of phylogenetic divergence between *Urochloa* species based on complete chloroplast genomes. BMC Genomics 18. doi:10.1186/s12864-017-3904-2

Picasso, V.D., P.D. Modernel, G. Becoña, L. Salvo, L. Gutiérrez, and L. Astigarraga. 2014. Sustainability of meat production beyond carbon footprint: A synthesis of case studies from grazing systems in Uruguay. Meat Sci. 98:346–354. doi:10.1016/j.meatsci.2014.07.005

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012a. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7. doi:10.1371/journal.pone.0032253

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.-L. Jannink. 2012b. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome J. 5:103. doi:10.3835/plantgenome2012.06.0006

Raboin, L.M., J. Pauquet, M. Butterfield, A. D'Hont, and J.C. Glaszmann. 2008. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. Theor. Appl. Genet. 116:701–714. doi:10.1007/s00122-007-0703-1

Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T.A. Graves, and others. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science (80-.). 326:1112–1115

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

Silva, P.I.T., A.M. Martins, E.G. Gouvea, M. Pessoa-Filho, and M.E. Ferreira. 2013. Development and validation of microsatellite markers for *Brachiaria ruziziensis* obtained by partial genome assembly of Illumina single-end reads. BMC Genomics 14. doi:10.1186/1471-2164-14-17

Stich, B., C. Urbany, P. Hoffmann, and C. Gebhardt. 2013. Population structure and linkage disequilibrium in diploid and tetraploid potato revealed by genome-wide high-density genotyping using the SolCAP SNP array. Plant Breed. 132:718–724. doi:10.1111/pbr.12102

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, F., and H. Ishwaran. 2017. Random forest missing data algorithms. Stat. Anal. Data Min. 10:363–377. doi:10.1002/sam.11348

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. doi:10.1371/journal.pone.0062355

Vigna, B.B.Z., L. Jungmann, P.M. Francisco, M.I. Zucchi, C.B. do Valle, and A.P. de Souza. 2011. Genetic Diversity and Population Structure of the *Brachiaria brizantha* Germplasm. Trop. Plant Biol. 4:157–169. doi:10.1007/s12042-011-9078-1

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

Wall, J.D., L.F. Tang, B. Zerbe, M.N. Kvale, P.Y. Kwok, C. Schaefer, and N. Risch. 2014. Estimating genotype error rates from high-coverage next-generation sequence data. Genome Res. 24:1734–1739. doi:10.1101/gr.168393.113

Wang, Y.-H., H.D. Upadhyaya, A.M. Burrell, S.M.E. Sahraeian, R.R. Klein, and P.E. Klein. 2013. Genetic Structure and Linkage Disequilibrium in a Diverse, Representative Collection of the C4 Model Plant, *Sorghum bicolor*. Genes|Genomes|Genetics 3:783–793. doi:10.1534/g3.112.004861

World Health Organization and United Nations University. 2007. Protein and amino acid requirements in human nutrition.. World Health Organ. Tech. Rep. Ser. 1–265. doi:ISBN 92 4 120935 6

Worthington, M., C. Heffelfinger, D. Bernal, C. Quintero, Y.P. Zapata, J.G. Perez, J. De Vega, J. Miles, S. Dellaporta, and J. Tohme. 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*. Springer.

FIGURES

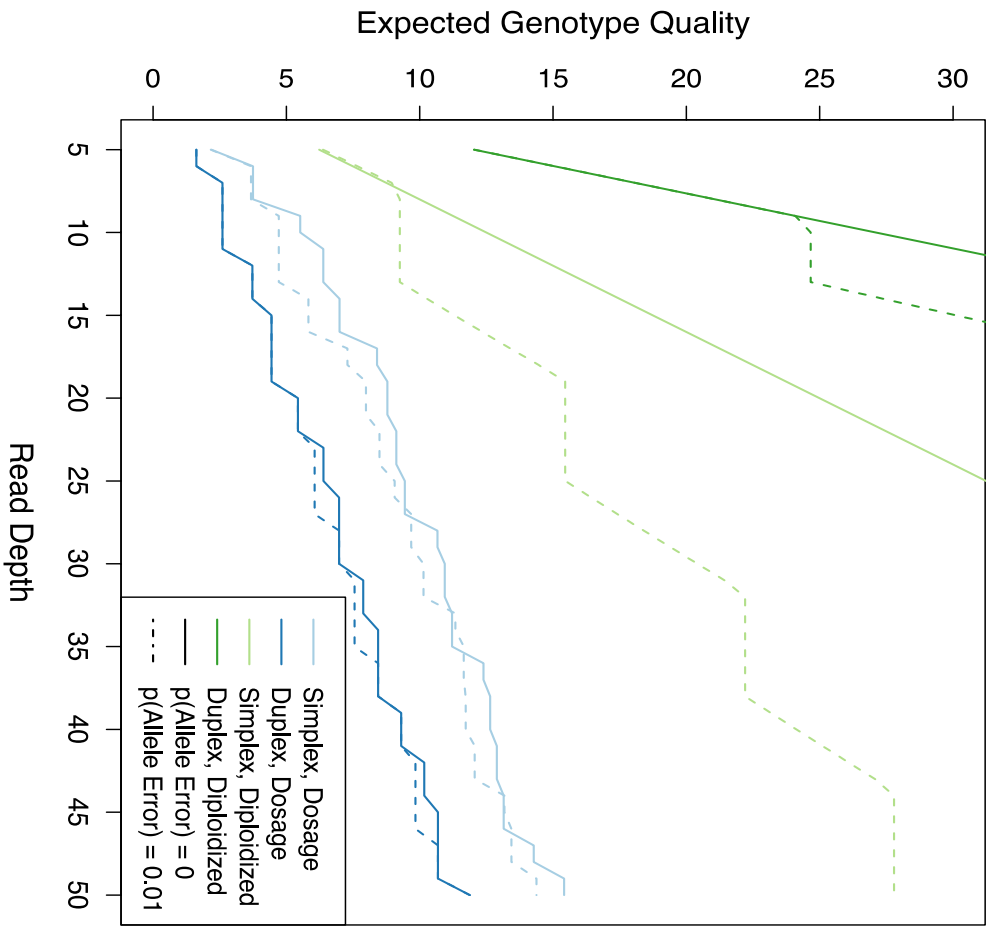

Figure 1. Expected Genotype Quality as a function of read depth, for two different allele error rates.

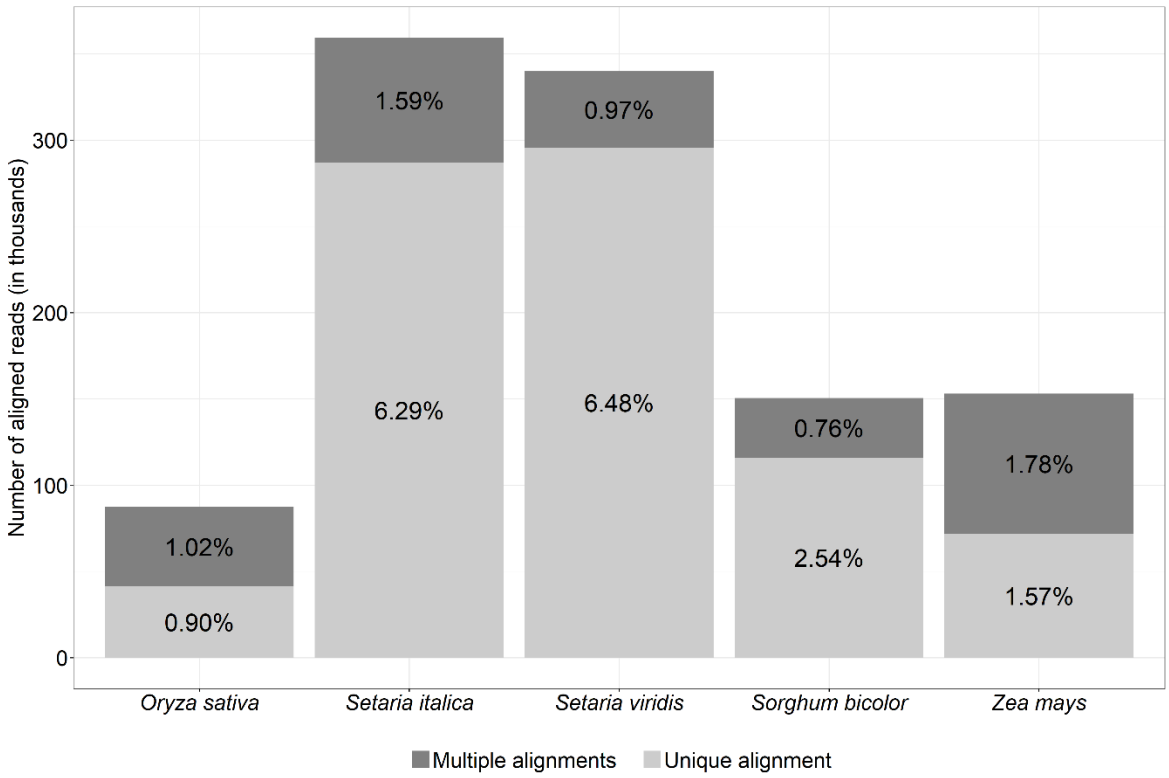

Figure 2. Number and percentage of reads from the *Urochloa brizantha* cultivar 'Marandu' that aligned to five *Poaceae* reference genomes.

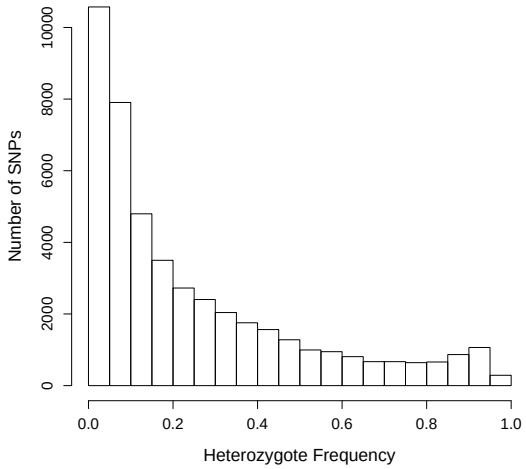

Figure 3. Distribution of heterozygote frequency for 46,147 bi-allelic SNPs discovered using the *Urochloa* mock reference genome for alignment.

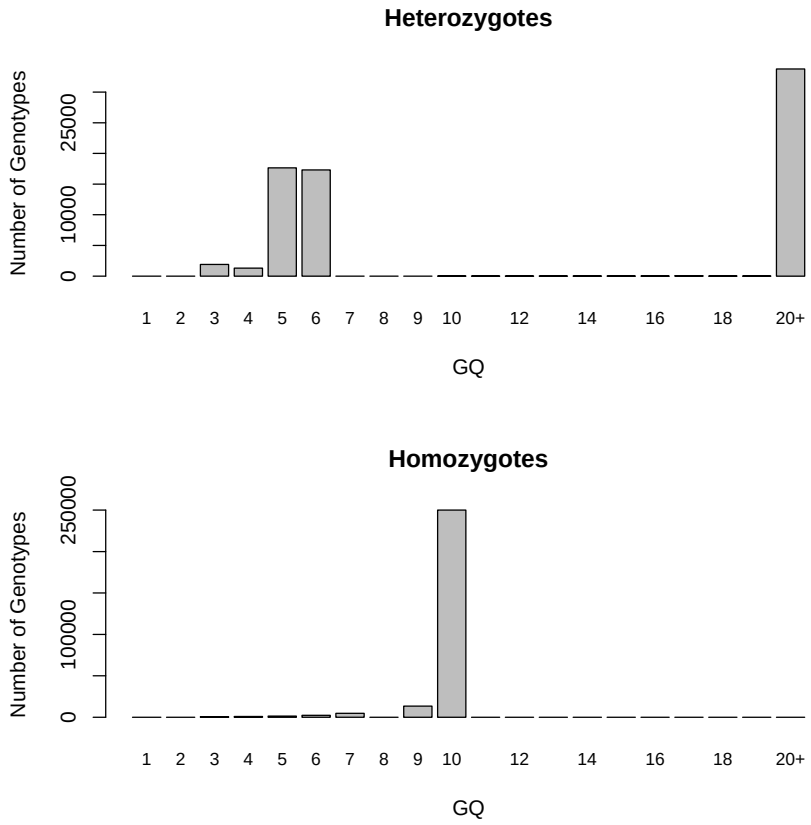

Figure 4. Distribution of genotype quality (GQ_{dip}) scores for diploidized genotypes with sample depth = 8 in a filtered set of 46,147 SNPs, which were discovered using the *Urochloa* mock reference genome for alignment.

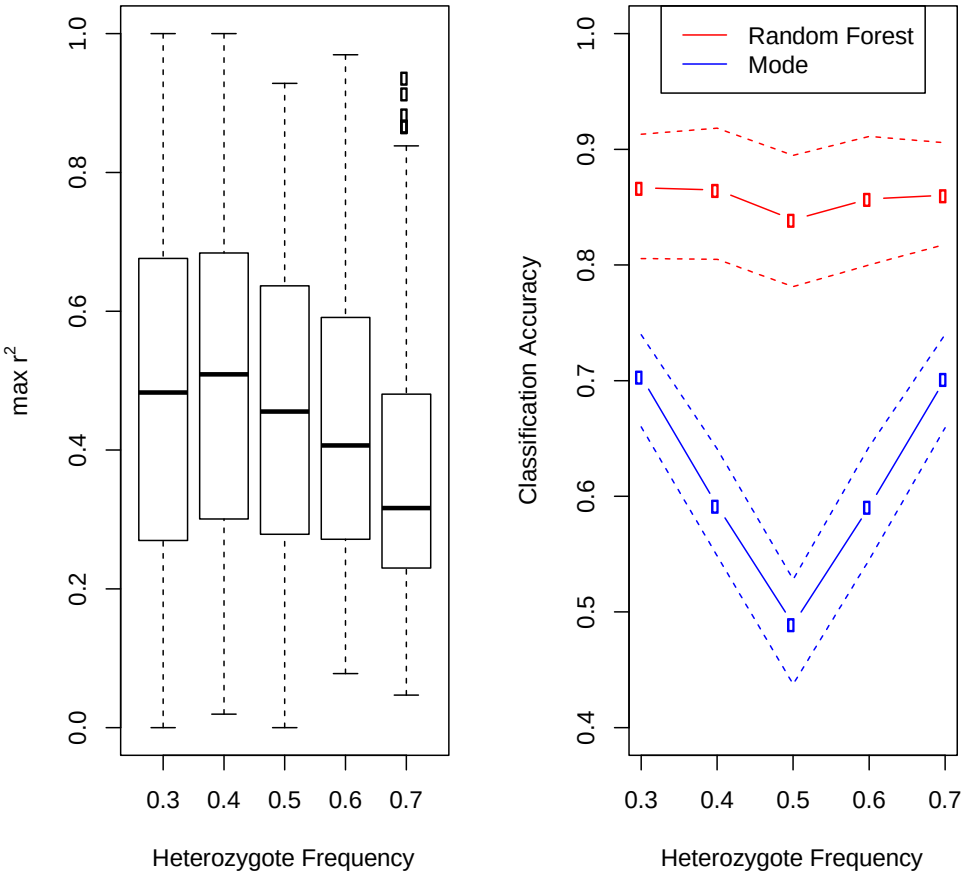

Figure 5. (Left) Distribution of the maximum LD (r^2) for each marker, binned by heterozygote frequency. To exclude SNPs from the same short read, which have the same pattern of missing data and are therefore useless for imputation, markers for which the pattern of missing data was > 90% identical were excluded when calculating max r^2. (Right) Cross-validation accuracy when imputing missing genotypes.

TABLES

Table 1. Number of bi-allelic SNPs with < 50% missing data, for different reference genomes and filtering criteria. Samples that did not meet the thresholds shown in the first row of the table were set to NA.

Reference Genome	Sample Depth ≥ 8	Sample Depth ≥ 8 $GQ_{dip} \geq 10$
Urochloa 'mock'	46,147	23,936
S. viridis	5,948	3,195
S. italica	5,121	2,807
S. bicolor	2,412	1,337
Z. mays	1,333	765
O. sativa	994	570

4. ASSOCIATION MAPPING CONSIDERING ALLELE DOSAGE: AN EXAMPLE OF FORAGE TRAITS IN AN INTERSPECIFIC TETRAPLOID *Urochloa* SPP. HYBRID PANEL

ABSTRACT

The breeding process in tropical polyploid 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 during different seasons, and (ii) to verify the influence of allele dosage on diploid and tetraploid configuration to identify genomic regions using markers from genotyping by sequencing (GBS). A panel of tetraploid hybrids (*Urochloa brizantha* x *Urochloa ruziziensis*) evaluated in wet and dry seasons was used. The GWAS was performed with 26,535 single nucleotide polymorphisms (SNPs) using diploid and tetraploid allele dosage configurations. The GWAS revealed the first seven different candidate genomic regions associated with the main forage traits of *Urochloa* spp. Our results demonstrated that it is possible to identify the same regions using both diploidized and polyploidy configuration, however, it can be misleading for some regions with dominance and epistatic control. Finally, this study contributes to further understand tropical forage genomics in order, to accelerate the selection, and reduce the cost to release new cultivars.

Keywords: GWAS; *Brachiaria*; Polyploid; Genotyping-by-sequencing; Allele dosage

4.1. INTRODUCTION

Climate change, and the decreasing availability of the land for livestock are significant challenges to overcome in order to meet the increasing demand for animal protein (World Health Organization and United Nations University, 2007; Tilman and Clark, 2014; Grandin, 2015). 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). The strategies to improve animal production on pastures rely on good nutrition derived from forage cultivars bred for better adaptation to soils, climate and pests, on good pasture management, on animals of superior genetics and in good health. 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 this sense, l forage breeding programs deals with targets beyond the readily observable crop performance since the final product is not the plant but an animal product, such as milk or meat. Thus, animal performance on the new cultivar needs to be evaluated before its release.

The genus *Urochloa* spp. is the most important forage for tropical regions, primarily the species *U. brizantha*, *U. decumbens* and *U. ruziziensis* (Jank et al., 2014) that are still known as *Brachiaria* in Brazil (Torres González and Morton, 2005). For the first two, genotypes available in the market are tetraploid and apomictic, while for the latter cultivars are diploid and sexual. In 1981 *U. ruziziensis* was artificially tetraploidized using colchicine, and later used as a sexual genitor in crosses with *U. brizantha* and *U. decumbens* to develop interspecific hybrids (Swenne et al., 1981; Lutts et al., 1991; Valle and Pagliarini, 2009). Their purpose was 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. has persisted as the main strategy to develop new cultivars (Figure 1) using superior apomictic cultivars as a pollen donor in crosses with superior selected sexual plants. Hybrids produced (frequently, 2,000 genotypes) are evaluated as single plants under field trials to select about 10% of the best performing individuals (~200 genotypes), which are a subsequently evaluated with a higher number of replicates (Figure 1a and 1b). Following the pipeline, 1 or 2 genotypes are selected and evaluated in terms of animal performance in several locations (Figure 1c and 1d). Finally, new apomictic cultivars are selected for release (Figure 1e). The best sexual genotypes and the new apomictic cultivars reenter the program as parents (Figure 1f). For thorough descriptions of breeding schemes see Barrios and colleagues (Barrios et al., 2013) and Worthington and Miles (Worthington and Miles, 2015).

Each stage of the tropical forage breeding program is usually evaluated for at least two wet and two dry seasons (two years) in the Cerrado biome in South America (Alvares et al., 2013). Hence, the time to develop a *Urochloa* spp. cultivar is approximately 10-15 years. The process could be accelerated through the development of tools to improve selection efficiency. For instance, significant markers identified by GWAS could be used on the first step of forage breeding (Figure 1a) to select the most promising hybrids at the seedling stage and eliminate the incomplete block design experiments. Therefore, we could start directly in the second stage (Figure 1b) and reduce the cycle duration by 4-5 years. Also, markers associated with agronomical and nutritional forage traits could be used to select genotypes either during the dry and/or the wet season for traits expressed in one or both seasons, which has not been reported yet. Furthermore, the results could help breeders

to understand the complex interactions between these environmental conditions and forage yield and quaity, of more pressing concern in the context of climate change.

Among many quantitative genetics tools, genome-wide association studies (GWAS) have been used to discover genetic regions with significant association to essential traits (Zargar et al., 2015). However, in polyploid species, GWAS is complicated, mainly due to the number of genotype classes and possible modes of gene action which, until recently, lacked appropriate analysis methods (Rosyara et al., 2016a). Consequently, GWAS in polyploids is a relatively new subject [16, 17] 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 to the use of the adequate allele dosage ones. Furthermore, despite the noteworthy importance of *Urochloa* spp. for livestock in tropical regions, there are no studies of GWAS using SNP markers performed within this genus.

This paper discusses the hypothesis that the use of the polyploid allele dosage instead of diploidized dosage can provided better genomic resolution for genetic value predictions, due to increasing the proximity of the statistical parameters to the genomic reality of the *Urochloa* genus. This type of result was observed for other species such as hexaploid chrysanthemum (Grandke et al., 2016)4 and autotetraploid blueberries (Ferrão et al., 2018). Moreover, the diploid configuration may mask the discovery of significant SNPs in GWAS in some cases. This could cause error during marker-based or marker-assisted selection, in which the true effect of a given SNP only comes to light by modeling the biologically appropriate ploidy. Therefore, the aims here were: (i) to identify genomic regions related with forage traits performance between the different seasons in tetraploid interspecific *Urochloa spp*. hybrids, and (ii) verify the influence of allele dosage on diploid and tetraploid configuration to identify genomic regions using SNP markers from genotyping by sequencing (GBS).

4.2. MATERIALS AND METHODS

Plant material and phenotyping

A set of 272 tetraploid interspecific hybrids were formed by crossing *Urochloa brizantha* x *Urochloa ruziziensis,* from the forage breeding program of Embrapa Beef Cattle, Brazil (Matias et al., 2018). This panel was evaluated using an incomplete block design during seven cuttings, from 2014

to 2015, at the experimental field area of the same company in Campo Grande-MS, Brazil (20°27'S; 54°57'W). In these experiments, the following commercial apomictic cultivars were used as a checks: *U. brizantha* cultivar 'Marandu', *U. brizantha* cultivar 'Paiaguás', *U. decumbens cultivar* 'Basilisk', the interspecific commercial hybrid 'Mulato II' and the accession 'B140' of *U. brizantha*, and the sexual elite tetraploid hybrids from Embrapa genetic bank (BS9, BS15, 336-T1 and 336-T2). Cuts 1, 4, 5 and 6 represented the wet season whereas cuts 2, 3 and 7 the dry season. The full experimental design and biological material (hybrids and checks) were already previously described by Matias and colleagues (Matias et al., 2018). The population was at the first stage of a forage breeding program. Thus, each hybrid was available as a single plant (Figure1a). Each hybrid was individually evaluated for the following two groups of traits:

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

Nutritional traits: Approximately 80 g of green forage fromcuttings 3 and 4 from each plant were dried, ground, and analyzed with infrared reflectance spectroscopy (NIRS) (Marten et al., 1989). The calibration of the NIRS was performed previously by comparing the results obtained in the wet chemical analyzes versus the spectrum read from these same samples in the NIRS for several nutritional characteristics (unpublished data). This process was used to estimate crude protein (CP), in vitro organic matter digestibility (IVD), neutral detergent fiber (NDF), and lignin in sulfuric acid (LIG).

Season effect estimation: The significance of season effect was estimated previously by a fixed model approach only to verify the difference between dry and rainy season, testing the fixed effects by Wald F test supported by the ASreml-R package (Butler et al., 2009). This model is important to identify that genotypes had different performance under different seasons. For that, a joint analysis using incomplete block design was performed including all genotypes (9 checks and 272 hybrids) by the following equation:

$$y_{acdh} = \mu + p_a + r_h + q_{a(c)} + s_d + \varepsilon_{acdh} \quad (1)$$

where $\boldsymbol{y}$ is the vector for phenotypic data; $\boldsymbol{\mu}$ is the vector for the overall mean; $\boldsymbol{p}$ is the vector of genotype, modeled as a fixed effect, with $a = \{1,2,\dots,281\}$; $\boldsymbol{r}$ is the vector indicating the fixed effect season, with $h = \{Wet\ or\ Dry\}$; $\boldsymbol{q}$ is the vector of cut nested to season effect, modeled as fixed, with $c = \{1,4,5,6\}$ for wet season, and $c = \{2,3,7\}$ for dry season; $\boldsymbol{s}$ is the vector of the block effect, modeled as fixed, with $d = \{1,2,\dots,10\}$; and $\boldsymbol{\varepsilon}$ is the residual, with $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$ where $\boldsymbol{I}$ is the identity matrix and σ_ε^2 is the residual variance component.

Genetic effects estimation: Once the significance of season was identified by the model (1), the phenotypic record was adjusted to be used in the GWAS. The genetic effects were estimated for individual hybrids using fixed model and testing the fixed effects by Wald F test supported by the ASreml-R package (Butler et al., 2009). Annual, rainy and dry season were considered in a two-step model. The second model (2) was evaluated as a complete block design with only the nine check entries to estimate the block effect, by the following equation:

$$y_{bcd} = \mu + p_b + q_c + s_d + u_{b_{\times}c} + \varepsilon_{bcd} \quad (2)$$

where $\boldsymbol{y}$ is the vector for phenotypic data; $\boldsymbol{\mu}$ is the vector for the overall mean; $\boldsymbol{p}$ is the vector of checks, considered as fixed, with $b = \{1,2,\dots,9\}$; $\boldsymbol{q}$ is the vector of cut effect, considered as fixed, with $c = \{1,2,\dots,7\}$ for annual, $c = \{1,4,5,6\}$ for wet season, and $c = \{2,3,7\}$ for dry season; $\boldsymbol{s}$ is the vector of block effect, considered as fixed, with $d = \{1,2,\dots,10\}$; $\boldsymbol{u}$ is the vector of check x cut interaction effect, considered as fixed; and $\boldsymbol{\varepsilon}$ is the residual effect with $\varepsilon \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$ where σ_ε^2 is the residual variance component.

The block effects from model (2) were deducted from observed data for each hybrid as a function of the hybrid position on the incomplete block design. Afterwards, the corrected data for each trait was used in the following equation:

$$y_{gc} = \mu + p_g + q_c + u_{g_{\times}c} + \varepsilon_{gc} \quad (3)$$

where $\boldsymbol{y}$ is the vector for corrected phenotypic data; $\boldsymbol{\mu}$ is the vector for the overall mean; $\boldsymbol{p}$ is the vector of hybrids, considered as fixed, with $g = \{1,2,\dots,272\}$; $\boldsymbol{q}$ is the vector of cut effect, considered as fixed, with $c = \{1,2,\dots,7\}$ for annual, $c = \{1,4,5,6\}$ for wet season, and $c = \{2,3,7\}$ for dry season; $\boldsymbol{u}$ is the vector of hybrid x cut interaction effect, considered as fixed; and $\boldsymbol{\varepsilon}$ is the residual effect with $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$. The Wald test implemented in ASReml was used to test the significance of the fixed effects. The broad heritability was calculated using model (3) but considered

the $\boldsymbol{p}$ effect of hybrids as random, with $\boldsymbol{p} \sim N(0, \boldsymbol{I}\sigma_g^2)$ with σ_g^2 is the variance component of hybrids, thus the $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\varepsilon^2}{c}}$, the random effects were tested by the likelihood ratio test (LRT).

Genotyping

Fresh leaf was used for DNA extraction with the Qiagen® kit, and samples were genotyped by sequencing (GBS) (Elshire et al., 2011) using the ApeKI enzyme and Illumina Hi-Seq 2500 platform. The Genome Analysis Toolkit (*GATK*) pipeline (McKenna et al., 2010; Depristo et al., 2011) was implemented to discover and call SNP markers using ploidy = 4 for genotype calling at the tetraploid level and then calls were "diploidized" based on the genotype likelihoods. Furthermore, sequencing reads were aligned with five different genomes: *Setaria virides* (Sv) (DOE-JGI, 2018a), *Setaria italic* (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). Default alignment parameters were used in 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/).

The SNP markers were filtered by median depth (Median DP) ≥ 8, minimum allele depth (MAD) ≥ 2, minor allele frequency (MAF) ≥ 0.01, and missing data ≤ 50%. Also, samples with genotype quality (QD) ≥ ten were eliminated and samples with DepthPerSample (DP) < 8 were set as missing. These filter criteria were applied in the diploid configuration, and the genotype of the selected markers was extended to the tetraploid level. 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. 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 evaluate the population structure according to the marker-based additive relationship ($\boldsymbol{G}_{Dip}$ and $\boldsymbol{G}_{Tetra}$) described by Endelman and colleagues (Endelman et al., 2018) and Vitezica and colleagues (Vitezica et al., 2013) following the equations below where $\boldsymbol{p}_i$ is the reference allele frequency, and $\boldsymbol{X}$ is the allele dosage matrix with genotypes denoted 0-2 for diploids and 0-4 for tetraploids. A graphic representation of the kinship matrix was built using the R package superheat (Barter and Yu, 2018).

$$\boldsymbol{W}_{Dip} = \left(\boldsymbol{X}_{Dip} - 2p_i\right)$$

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

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

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

Genomic association analysis (GWAS)

The adjusted means of hybrids from the model (3) were used to perform the GWAS analysis of the traits on the annual data, wet and dry season. This analysis was inspired by the work developed by Pajerowska-Mukhtar and colleagues (Pajerowska-Mukhtar et al., 2009) with tetraploid potato and association studies with field resistance to diseases. The hybrids 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 and colleagues (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., 2016b).

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{ZSt} + \boldsymbol{ZQv} + \boldsymbol{Zu} + \boldsymbol{\varepsilon} \quad (4)$$

where $\boldsymbol{y}$ is the vector of the adjusted phenotypes; $\boldsymbol{\beta}$ is the vector of fixed effects; $\boldsymbol{t}$ in the vector of SNP effects; $\boldsymbol{v}$ is the vector of effects for the subpopulations; $\boldsymbol{u}$ is the vector of polygenic effects with $\boldsymbol{u} \sim N(0, \boldsymbol{K}\sigma_g^2)$ where $\boldsymbol{K}$ is the relationship matrix built according to VanRaden (VanRaden, 2008) and σ_g^2 is the variance component of genotypes; and $\boldsymbol{\varepsilon}$ is the residual effect vector with $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$ where $\boldsymbol{I}$ is the identity matrix σ_ε^2 is the variance component of error. The incidence matrix $\boldsymbol{X}$ is the indicates fixed effects, $\boldsymbol{Z}$ is the incidence matrix of SNPs, $\boldsymbol{Q}$ is the incidence matrix of population, and $\boldsymbol{S}$ is the incidence matrix to structure the genetic models. Genetic models fitted are general, additive, and dominance. The general model uses each genotype class independently; only the difference between the levels matter for the F-test. For the additive model, the effect was evaluated by dosage of the reference allele. Dominance models are for simplex and duplex dominance. Basically the first dominance model evaluates the hypothesis that the presence of one or

more copies of a reference allele has a similar performance between them and different performance from the homozygote genotype of the alternative allele (Rosyara et al., 2016b). The second dominance model is specific for tetraploid configuration and evaluates if the duplex genotype had similar performance of nuliplex and simplex genotype or similar performance to triplex and tetraplex genotypes (Rosyara et al., 2016b). False discovery rate was controlled by setting the significance threshold with Bonferroni's multiple testing correction methods. Also, the influence of the first two principal components (PCs) from the SNP set was evaluated on the GWAS models for both ploidies. 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

The sequence of fifty nucleotide positions from both sides of significant SNPs was selected from the respective reference genome and compared to the other species using the BLAST tool (Altschul et al., 1990) to determined candidate genes. Gene function and homology were annotated using the NCBI platform (NCBI Resource Coordinators, 2017).

4.3. RESULTS

Agronomical and nutritional phenotypic variation

There was a significant effect of genotype and season (Supplementary Table S1), and genotype effect nested in 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 heritability considering the annual data ranged from 0.46 to 0.81 for lignin and field green weight (FGW), respectively. This trait showed a different trend when compared to the other traits:regrowth capacity (REG), crude protein (CP), in vitro organic matter digestibility (IVD), neutral detergent fiber (NDF), and lignin (LIG). FGW 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 in the wet season (27%).

As one would expect, the 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 $kg.ha^{-1}$, while its performance in the dry and wet season was 1062.26 $kg.ha^{-1}$ and 1773.83 $kg.ha^{-1}$, respectively (Table 1). Thus, the genotypes had a better performance for FGW, REG and

IVD in the wet season than in the dry one. In contrast, for CP, NDF, and LIG there were no significant differences between the two seasons.

Considering the correlations among the traits, a positive one was found between FGW and REG, and NDF and LIG (Figure 2). Also, a strong positive correlation was observed between IVD and CP. On the other hand, IVD showed a negative correlation with LIG and NDF, and between CP with NDF, and FGW, as well. It is important to highlight that the correlation trend among traits retained its directionality across seasons, yet the magnitudes of the correlation tended to be lower in the wet compared to the dry season.

Population structure and diversity analysis

For the diploid and tetraploid configurations, population structure was evaluated in terms of genotype frequencies, visual representation of 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 (Figure 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 possible heterozygote tetraploid genotypes 1 (12.41%), 2 (6.74%) and 3 (2.7%), that summed 21.85%. The heatmap of the kinship additive matrix showed differences between the diploid and tetraploid levels (Figure 3B). There, the shape and size of the clusters are different for the ploidies. In the diploid configuration, the number of clusters is easier to identify compared to the tetraploid configuration. However, the use of tetraploid data organized the population in three greater groups. The biplot from the first two principal components (PCA) explained 13.7% at the diploid level and 12.8% using tetraploid marker data (Figure 3C). The cloud of points for both ploidies had a triangular shape, but with a different orientation. The genomic structure from the diallel origin can be identified by the PCA analysis using diploid or tetraploid data configuration.

GWAS analyses for the different seasons

A complex interaction of genotype by environment was observed when different heritability and performance values were found for all traits across seasons as mentioned above. Overall, additive and dominant models of GWAS were evaluated using markers at the diploid and tetraploid

level, and only the significant results are described below. We verified that each trait showed a different response regarding the combination between season, ploidy, and GWAS model. For instance, no significant marker effect was found for IVD at the diploid level. However, at least one marker was identified in each season using the tetraploid level (Figure 4).

All 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, mapped on the chromosome eight of *Setaria virides* at the position 7,908,449. 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, 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 data and the marker for the dry season were aligned with *Urochloa mock* (*Um*) reference at the positions 8,160,655 and 128,132, respectively (Supplementary Fig. S1). These markers came from centroids in the *Um* genome reference (data not shown) originated by 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* chromosome 3B, but no function was annotated.

Significant markers were found for IVD in all environmental conditions (Table 2). For these traits, we used the first two principal components in GWAS models aiming to correct for the effect of population structure. The markers *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 former was aligned with *Urochloa mock* reference at 3,259,930 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 effect was significant in the dry season for FGW. It was located on *Sorghum bicolor* chromosome 02 at position 1,588,978. The alternative allele of this SNP is dominant and associated with a 172.5 kg ha^{-1} reduction in the trait when present. Furthermore, the gene LOC8084285 was annotated in this region, which corresponds to an uncharacterized protein.

Differences between diploid and tetraploid levels results

Concerning the different models, no pattern was observed for ploidy level within seasons (Table 2). For the diploid level, only the single copy dominance (*1-dom-alt/ref*) and general models allowed the identification of significant SNPs associated with the traits. Besides these two models, significant SNPs were also found using the tetraploid configuration in combination with two copy dominance (*2-dom-ref*) model.

As already mentioned, the same marker *Sv_Chr08_7908449* that was significant for REG in the annual data and dry season, was significant for both diploidized and tetraploid data sets (Table 2 and Figure 5A). The model had excellent fit as highlighted by the quantile-quantile plot (Figure 5B). This marker had dominant behavior for the reference allele. In turn, at the diploidized level, at least one copy of this allele ($A-$) was necessary to improve the REG from -1 for approximately -0.5 (Figure 5C). 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 same marker (*Um_8160655*) was identified by the diploid and tetraploid models (Table 2, Figure 6A and Figure 6B). However, for REG, the results were different with the ploidy models. We verified that the alternative allele exerts a dominant effect on the trait. In this case, the homozygous genotypes for the reference allele and all the heterozygous displayed reduced average values for this trait in comparison to genotypes with two copies (diploid adjustment) or four copies (tetraploid adjustment) of the alternative allele. Thus, only genotypes homozygous for the alternative allele showed the undesirable high levels of LIG.

Furthermore, for some traits, significant SNPs were found only using specific marker configurations (Table 2). For example, a significant association for FGW (dry season) was only identified with the diploid model. In contrast, for IVD, all the significant SNPs were identified by the tetraploid model (dry, wet season, and annual data).

The alignment of the significant SNPs on reference genomes revealed that these sequences were positioned on the chromosome 2 of *Sorghum bicolor*, chromosome 5 of *Setaria italica*, and chromosome 9 of *Setaria virides* (Supplementary Fig. S2). Moreover, the *Setaria italica* allowed more alignments in general, which were relatively well distributed across the genome.

4.4. 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 the rainy season and dormancy in the dry one. This results in an irregular supply of fodder across the year. According to Jones (Jones, 1979), this seasonal difference in forage growth is the main obstacle to animal production in tropical and subtropical regions. The results here reported showed that *Urochloa spp*. interspecific population has genetic variability in seasonal production to be explored. Genotypes with combined maximum production for both dry and wet seasons can be identified. 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).

Seasonal variation modifies the environment and promotes physiological and morphological reactions on the plant. For example, about 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 the 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.

The correlation between the agronomical and nutritional traits follow what has been observed previously for the genus *Urochloa*, as described for *U. humidicola* (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. These nutritional traits will concentrate in tillers following the plant senescence. 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, 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

New sequence generation genotyping approaches have generated a massive volume of genomic information for different species of animals and plants. This tool also is not restrictive, and species without a reference genome can be evaluated. In these cases, the available genome of related species can be used to discover variant nucleotides in the target population (He et al., 2014). This new genomic information can be used for many biological studies and applications in breeding such as genomic selection and GWAS analysis. Here, the genome of five grasses were used to discover SNPs in an interspecific *Urochloa spp*. population. These markers were evaluated in a genomic association study to find markers that are in linkage disequilibrium with quantitative loci for forage traits, following the descriptions of Collard & Mackill (Collard and Mackill, 2008). Furthermore, to our knowledge, this is the first study reporting the application of GBS in this type of population 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 other nutritional (IVD) traits. 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 (Singh et al., 1973). In addition, Asp is a reserve of organic nitrogen, so that its decrease during water stress suggests that rates of nitrogen 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. *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 forages studies, directly related with forage production. Therefore, further studies to characterize this region could help breeders understand the genetic base of forage development.

Additionally, 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 for abiotic and biotic stress (Pabon et al., 2007; Mendonça et al., 2013; Matias et al., 2016). Furthermore, it may indicate that there is 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 (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. Thus, IVD performance could indirectly evaluate CP.

Specifically, the significant SNPs related to Neutral Detergent Fiber (NDF) for the annual and dry season, aligned with *Urochloa mock* (*Um*) reference genome showed a similar final sequence. Probably, these markers are in linkage disequilibrium with different copies of the same gene scattered in the polyploid genome. Another possibility is that these sequences can be different haplotypes of the same region, and the ploidy level is confounding the alignment during the mock reference building. Moreover, repetitive DNA sequences 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. It highlights that these species

share genes and genomic regions with *Urochloa* spp. Among them, *Setaria italica* genome allowed more alignments and coverage. 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*. These result corroborates others of 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; International and Initiative, 2010).

Effect of ploidy

Most of the genetic studies in polyploids simplify the data to use diploid models. However, that can under- or over-estimate the real genetic control of important traits (Dufresne et al., 2014). This study corroborates that the marker configuration is essential in GWAS discoveries for forage traits of *Urochloa spp*. interspecific populations. It demonstrated that the same SNP was significant for both diploid and tetraploid levels in some cases. In these cases, a more detailed information about the performance of intermediate genotypic classes allowed for a better resolution of the substitution allele effects and the number of copies needed for a given phenotype. Although the proportions between homozygotes and heterozygotes were similar, the inclusion of dosage information promoted different kinship matrix configurations, and also different biplot of principal components. Consequently, the efficacy of the MAS approach relied essentially on the information about the performance of all the genotypic classes.

Furthermore, the allele dosage information has several advantages, including the development of highly saturated maps that facilitate the detection of regions associated with important crop traits through QTL mapping (Bourke et al., 2018; Ferreira et al., 2018). Thus, these results provide evidence that the use of allele dosage for GWAS in polyploid species is indeed significant especially to better understand the inter-allelic interactions among the different alleles, however difficult it is to utilize a larger population size in order to represent well each of the all the possible genotypic classes.

4.5. CONCLUSION

This study presents the first GWAS analysis in interspecific tetraploid hybrids of *Urochloa* spp, 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 different, conserved and pleiotropic. The season (dry or wet) may influence the genomic regions that are controlling the trait variability, as observed for digestibility. Unfortunately, *Urochloa* spp does not have reference genome yet, but the region around these markers can be further investigated and improve the knowledge about the genomic control of tropical forage traits. This information could also be applied towards selection assisted by markers. These results showed the importance of the tetraploid configuration of markers for the identification of SNPs and to understand the behavior of gene action relative to allele dose. They also showed that diploid configuration can be used to do GWAS in polyploid species but without the possibility to estimate allele dosage. It was possible to identify the same regions using diploid or tetraploid data but the genetic effect of the alleles present may be masked in regions with dominance or epistatic control. The tetraploid configuration normally has five classes of genotypes (0, 1, 2, 3 and 4), consequently it complicates computational analyses compared to diploidized configuration with only three classes (0, 1 and 2). To account for trustworthy information the tetraploid configuration will need a higher amount of phenotypic data. As described by Morgante and colleagues (Morgante et al., 2018), the simple models are better for datasets with a low number of observations. These authors conclude by simulation approach that the number of observations should increase following the complexity of the parameters. Normally in practical forage breeding programs, the size of the genotyped population is not generally enough to get high confidence using tetraploid configuration on GWAS. In other words, the majority of markers does not have enough observation of each possible genotype (0, 1, 2, 3 and 4). Thus, the use of diploidized configuration could be more realistic, due to a greater number of genotypes in each possible class (0, 1 and 2). Finally, this study contributed to the advancement in the tropical forage genomic understanding and after the validation, the significant SNPs could be useful to the breeding program of *Urochloa* spp. aiming to accelerate the selection of future cultivars by reducing the cost and the time to release them.

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.

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.

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.

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.

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. 1–30.

Bennetzen, J. L., J. Schmutz, H. Wang, R. Percifield, J. Hawkins et al., 2012 Reference genome sequence of the model plant *Setaria*. Nat. Biotechnol. 30: 555.

Bourke, P. M., V. W. Gitonga, R. E. Voorrips, R. G. F. Visser, F. A. Krens et al., 2018 Multi-environment QTL analysis of plant and flower morphological traits in tetraploid rose. Theor. Appl. Genet. 1–15.

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 149.

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: 386–93.

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 Biol. Sci. 363: 557–572.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire et al., 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491–501.

DOE-JGI, 2018a *Setaria viridis* v1.1. http://phytozome.jgi.doe.gov/.

DOE-JGI, 2018b *Sorghum bicolor* v3.1. http://phytozome.jgi.doe.gov/.

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.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto et al., 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:.

Endelman, J. B., C. A. Schmitz Carley, P. C. Bethke, J. J. Coombs, M. E. Clough et al., 2018 Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics genetics.300685.2018.

Endres, L., J. L. de Souza, I. Teodoro, P. M. G. Marroquim, C. M. dos Santos et al., 2010 Gas exchange alteration caused by water deficit during the bean reproductive stage. Rev. Bras. Eng. Agrícola e Ambient. 14: 11–16.

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.

Fendrych, M., L. Synek, T. Pečenková, H. Toupalová, R. Cole et al., 2010 The Arabidopsis Exocyst Complex Is Involved in Cytokinesis and Cell Plate Maturation. Plant Cell 22: 3053–3065.

Ferrão, L. F. V., J. Benevenuto, I. de B. Oliveira, C. Cellon, J. Olmstead 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.

Ferreira, R. C. U., L. A. de C. Lara, L. Chari, S. C. L. Barrios, C. B. do Valle et al., 2018 Genetic mapping with allele dosage information in tetraploid *Urochloa decumbens* (Stapf) R.D. Webster reveals insights into spittlebug (*Notozulia entreriana* Berg) resistance. bioRxiv 360594.

Figueiredo, U. J. de, 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.

Gale, M. D., and K. M. Devos, 1998 Comparative genetics in the grasses. Proc. Natl. Acad. Sci. 95: 1971–1974.

Grandin, T., 2015 Improving Animal Welfare: A Practical Approach (T. Grandin, Ed.). CAB International, Boston, MA, USA.

Grandke, F., P. Singh, H. C. M. Heuven, J. R. de Haan, and D. Metzler, 2016 Advantages of continuous genotype values over genotype classes for GWAS in higher polyploids: A comparative study in hexaploid chrysanthemum. BMC Genomics 17:.

Hayes, B. J., N. O. I. Cogan, L. W. Pembleton, M. E. Goddard, J. Wang et al., 2013 Prospects for genomic selection in forage plant species. Plant Breed. 132: 133–143.

He, J., X. Zhao, A. Laroche, Z.-X. Lu, H. Liu et al., 2014 Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci. 5:.

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.

International, T., and B. Initiative, 2010 Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463: 763–8.

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.

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.

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.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan et al., 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Liaw, a, and M. Wiener, 2002 Classification and Regression by randomForest. R news 2: 18–22.

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

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.

Marten, G. C., J. S. Shenk, and F. E. Barton, 1989 Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality. Agric. Handb. 95.

Matias, F., S. Barrios, C. Do Bearari Lucas; Valle, R. Mateus, A. P. Do et al., 2018 Contribution of additive and dominance effects on agronomical and nutritional traits, and multivariate selection on *Urochloa* spp. hybrids. Crop Sci.

Matias, F. I., S. C. L. Barrios, C. B. do Valle, R. G. Mateus, L. B. Martins et al., 2016 Estimate of genetic parameters in *Brachiaria decumbens* hybrids. Crop Breed. Appl. Biotechnol. 16: 115–122.

Mauri, J., V. H. Techio, L. C. Davide, D. L. Pereira, F. S. Sobrinho et al., 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 et al., 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

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.

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. Grasslands 1: 103–105.

Mourad, A. M. I., A. Sallam, V. Belamkar, S. Wegulo, R. Bowden 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.

NCBI Resource Coordinators, 2017 Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45: D12–D17.

Ouyang, S., W. Zhu, J. Hamilton, H. Lin, M. Campbell et al., 2006 The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res. 35: D883--D887.

Pabon, 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.

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

Rosyara, U. R., W. S. De Jong, D. S. Douches, and J. B. Endelman, 2016a Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. Plant Genome 1–10.

Rosyara, U. R., W. S. De Jong, D. S. Douches, and J. B. Endelman, 2016b Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. Plant Genome 9: 0.

Santos, P. M., P. G. da Cruz, L. C. de Araujo, J. R. M. Pezzopane, C. B. do Valle et al., 2013 Response mechanisms of *Brachiaria brizantha* cultivars to water deficitstress. Rev. Bras. Zootec. 42: 767–773.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei et al., 2009 The B73 maize genome: complexity, diversity, and dynamics. Science (80-.). 326: 1112–1115.

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.

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.

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. 45–56.

Van Soest, P. J., 1995 Nutritional ecology of the ruminant.

Sun, C., B. Wang, X. Wang, K. Hu, K. Li et al., 2016 Genome-Wide Association Study Dissecting the Genetic Architecture Underlying the Branch Angle Trait in Rapeseed (Brassica napus L.). Sci. Rep. 6:.

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.

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

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

Torres González, A. M., and C. M. Morton, 2005 Molecular and morphological phylogenetic analysis of *Brachiaria* and *Urochloa* (Poaceae). Mol. Phylogenet. Evol. 37: 36–44.

VanRaden, P. M., 2008 Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91: 4414–4423.

Varshney, R. K., L. Pazhamala, J. Kashiwagi, P. M. Gaur, L. Krishnamurthy et al., 2011 Genomics and physiological approaches for root trait breeding to improve drought tolerance in chickpea (*Cicer arietinum* L.), pp. 233–250 in Root Genomics,.

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

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.

Voorrips, R. E., 2002 MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. J. Hered. 93: 77–78.

World Health Organization and United Nations University, 2007 Protein and amino acid requirements in human nutrition. World Health Organ. Tech. Rep. Ser. 1–265.

Worthington, M. L., and J. W. Miles, 2015 Reciprocal Full-sib Recurrent Selection and Tools for Accelerating Genetic Gain in Apomictic *Brachiaria* pp. 19–30 in Molecular Breeding of Forage and Turf, Springer International Publishing, Cham.

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.

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

Zargar, S. M., B. Raatz, H. Sonah, Muslimanazir, J. A. Bhat et al., 2015 Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. J. Crop Sci. Biotechnol. 18: 293–308.

Zheng, H., R. C. Babu, Md., M. S. Pathan, L. Ali, N. Huang et al., 2000 Quantitative trait loci for root-penetration ability and root thickness in rice: Comparison of genetic backgrounds. Genome 43: 53–61.

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. 201800182.

FIGURES

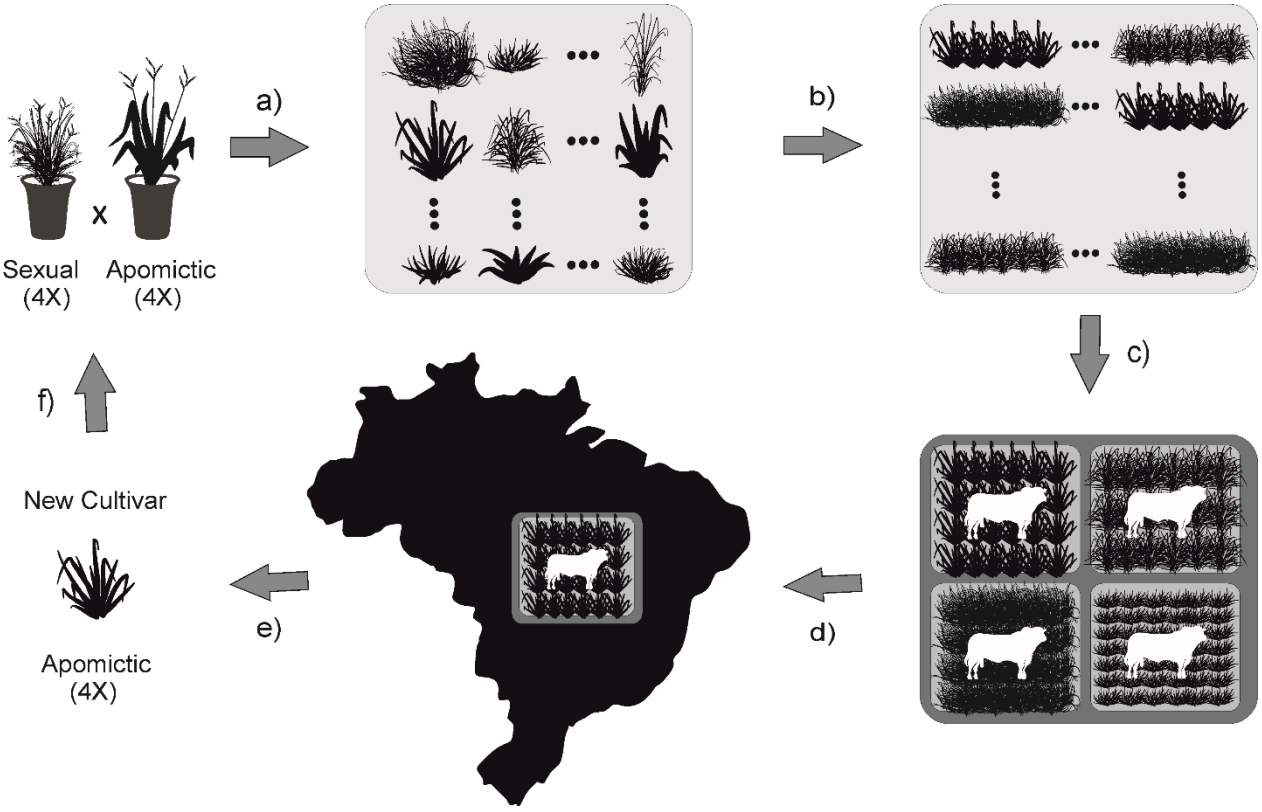

Fig 1. *Urochloa spp*. breeding program scheme to develop apomictic cultivars. a) Hybridization - single-cross between commercial apomictic cultivars and synthetic sexual parents (two years); b) Stage 1 - progeny evaluation based on one plant per plot (two years); c) Stage 2 - individuals selected in stage 1 are evaluated trials with more replicates (two years); d) Stage 3 - the hybrids selected in stage 2 are evaluated for animal performance (two years); e) Stage 4 - Regional multi-trial experiments considering the selected genotypes from stage 3 (two years); f) Seed production and release of the newest apomictic cultivar, which also enters the breeding program as a male parent (two-five years).

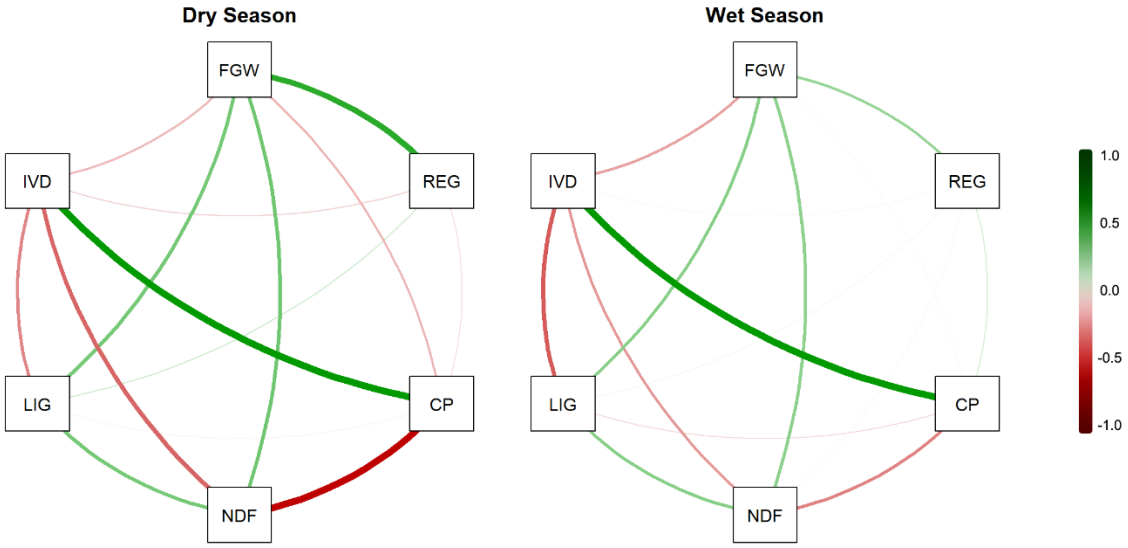

Fig 2. Correlation network between field green weight (FGW), regrowth capacity (REG), crude protein (CP), in vitro organic matter digestibility (IVD), neutral detergent fiber (NDF) and lignin in sulfuric acid (LIG).

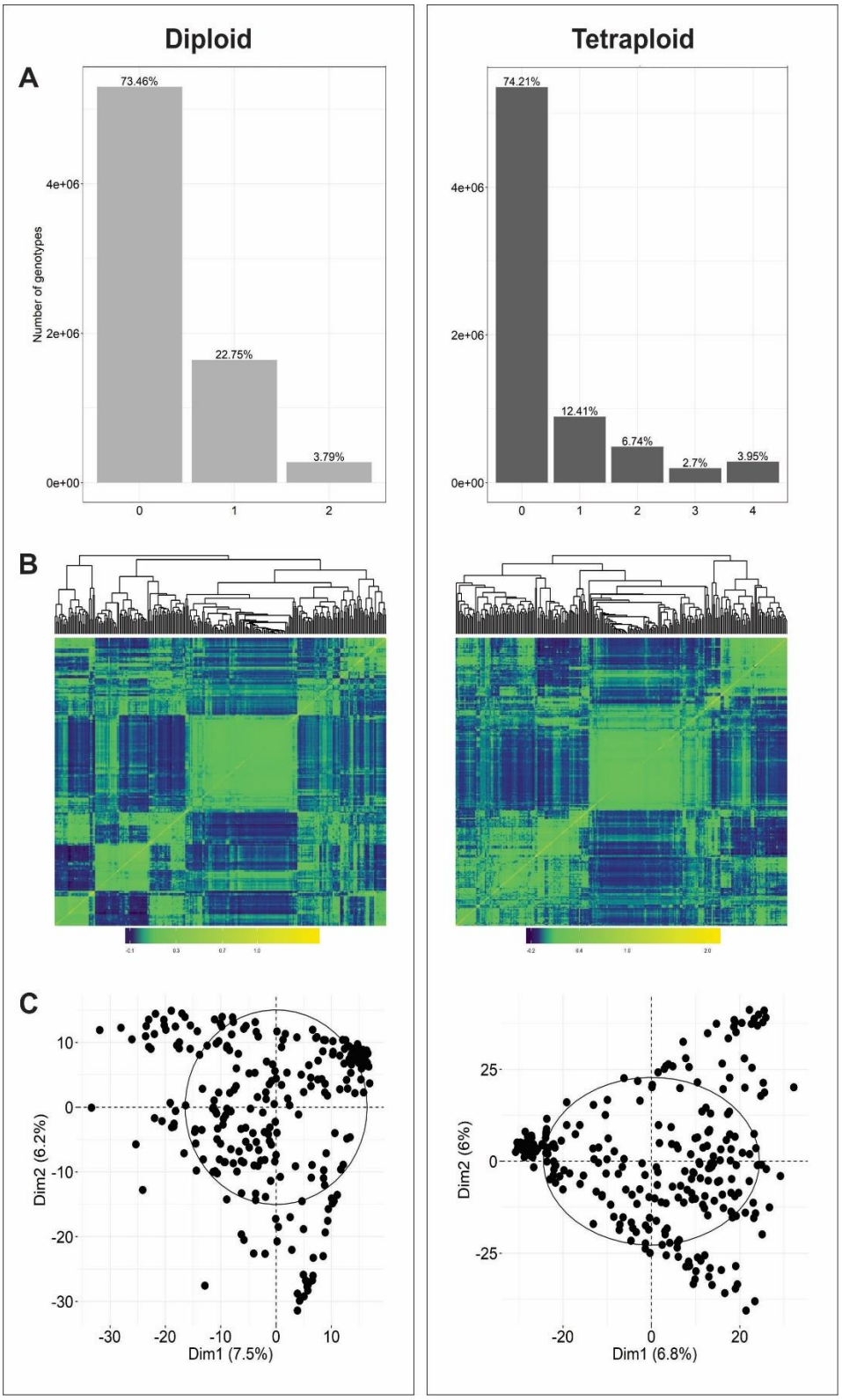

Fig 3. Population structure and diversity analysis using diploid and tetraploid marker configuration. A) The proportion of each class of genotype; B) Kinship matrix heatmap; and C) Biplot from the first two principal components.

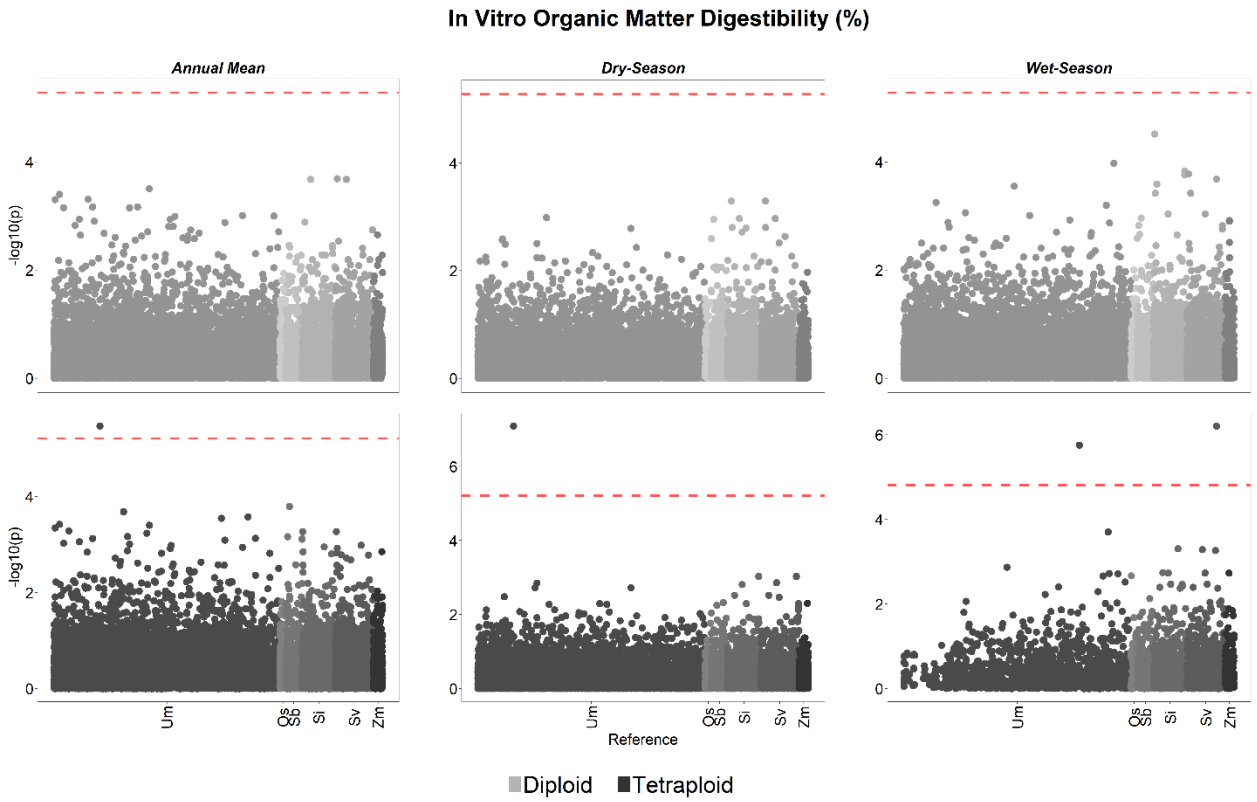

Fig 4. Manhattan plot of genome-wide association using diploid and tetraploid markers of 272 *Urochloa spp*. hybrids for in vitro organic matter digestibility (%). The values of -log10(p) in each Manhattan plot were sorted by position and identified by the following reference genomes: Si=*Setaria italica*, Sb=*Sorghum bicolor*, Sv=*Setaria virides*, Os=*Oryza sativa*, Zm=*Zea mays*, At=*Arabidopsis thaliana* and Um=*Urochloa mock*. The additive gene-action GWAS model was used for all diploid level scenarios. For the tetraploid level, general gene-action GWAS model is shown for annual and dry season whereas a dominance gene-action GWAS model is shown for wet season.

Um_8160655: AAGATCTTTT**C**ATCCGGAGGAAGTTTTTGGAGGCCTCGAAGCATCTCCATAACGGCT**GAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA**

Um_128132: CACCAAACCAACAAGCAAACCATGCACGTAGTGCCC**G**TGGACACACCAACATGTGCA**GAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA**

Fig 5. Comparison of the significant SNPs found for Neutral Detergent Fiber (NDF) for annual and dry season aligned with *Urochloa mock* (*Um*) reference genome at the positions 8,160,655 and 128,132, respectively. The SNP markers are highlighted in red and the coincident sequences in blue.

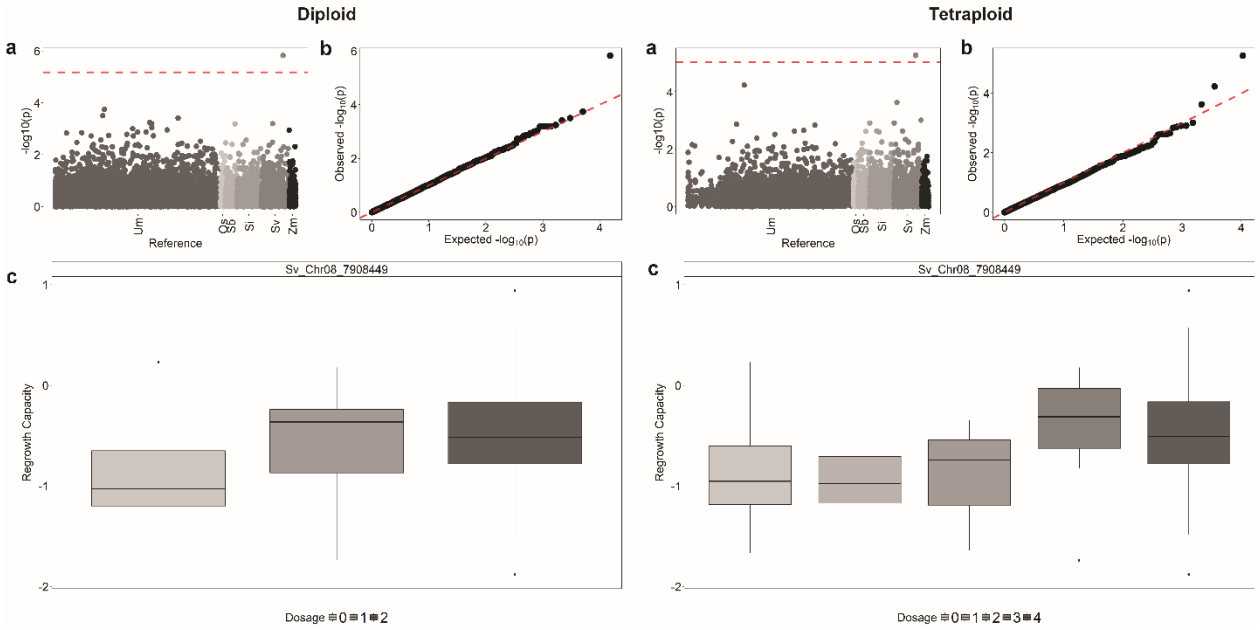

Fig 6. GWAS results of *Urochloa* spp. for regrow capacity (REG) on annual period using diploid and tetraploid configuration markers. A) Manhattan plot; B) QQplot; C) Box plot showing the trait average by genotype for significant SNPs. Reference genomes: Si=*Setaria italica*, Sb=Sorghum bicolor, Sv=Setaria virides, Os=*Oryza sativa*, Zm=*Zea mays*, At=*Arabidopsis thaliana* and Um=*Urochloa mock*.

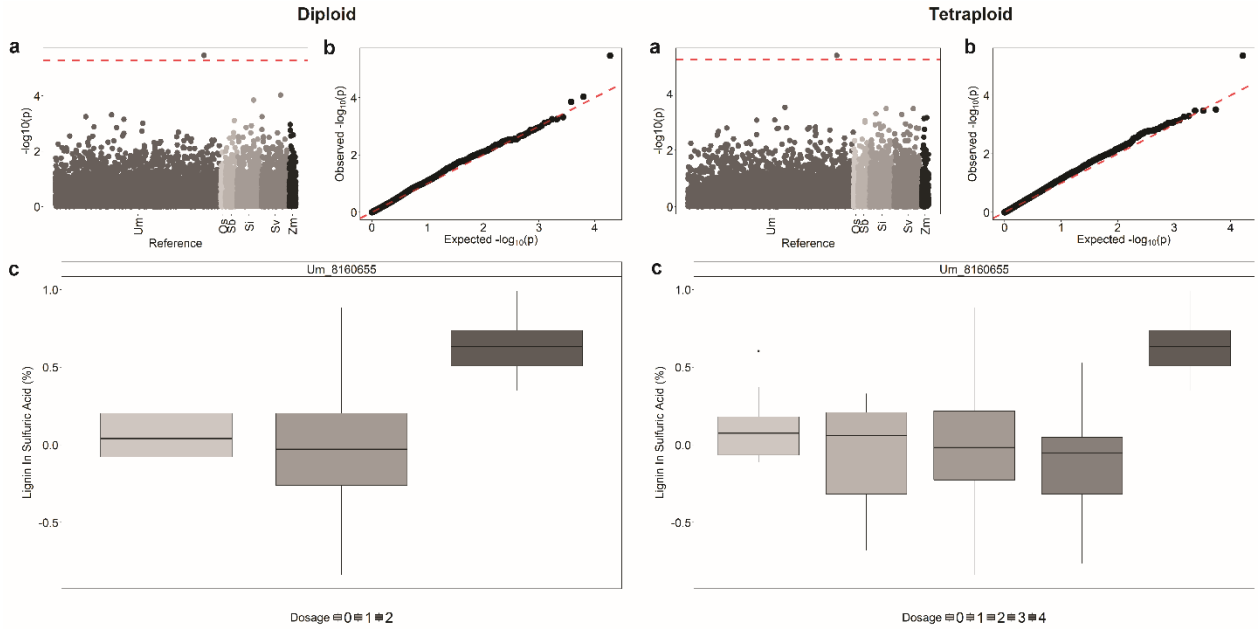

Fig 7. GWAS results of *Urochloa* spp. for lignin in sulfuric acid (LIG) on wet season using diploid and tetraploid configuration markers. A) Manhattan plot; B) QQplot; C) Box plot showing the trait average by genotype for significant SNPs. Reference genomes: Si=*Setaria italica*, Sb=*Sorghum bicolor*, Sv=*Setaria virides*, Os=*Oryza sativa*, Zm=*Zea mays*, At=*Arabidopsis thaliana* and Um=*Urochloa mock*.

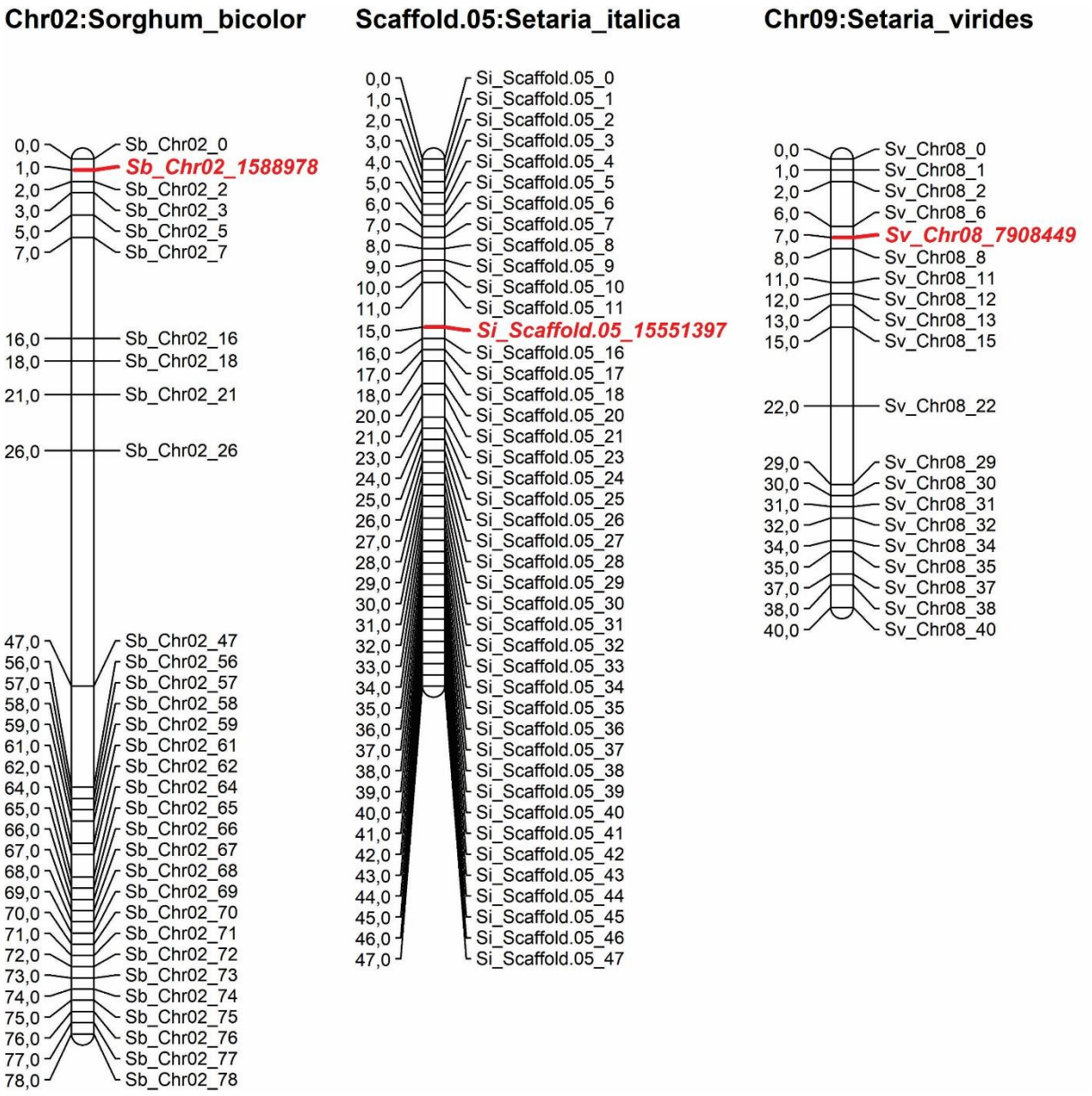

Fig 8. Relative positions regions (kb) with the presence of at least one Single Nucleotide Polymorphism (SNP) identified on reference genomes: *Sorghum bicolor*, *Setaria italica* and *Setaria virides*. The GWAS significant SNPs are highlighted in red.

TABLES

Table 1. Wald test to 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)

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 Season			
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 Season			
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 0.05 or (**) 0.01 level

Table 2. Annotated genes from significant SNP markers associated with *Urochloa* spp. forage traits identified by GWAS analysis.

Season	Trait	Ploidy	model	PC	Threshold	Reference	Chrom	Position	Score	MAF	Effect	Gene	Protein	R^2
Annual	REG	Dip	1-dom-alt	0	5.18	Sv	Chr08	7908449	5.83	0.25	0.45	trnD-GUC	tRNA-Asp	0.14
		Tetra	2-dom-alt		5.02				5.27	0.25	0.42			0.15
	NDF	Dip	1-dom-alt	0	5.18	Si	Scaffold_5	15551397	5.21	0.07	-2.54	LOC101778276	Exocyst Complex Component SEC15B	0.07
		Tetra	2-dom-ref		4.80	Um		8160655	4.94	0.46	1.98		Aligned with Triticum aestivum chromosome 3B	0.07
	IVD	Tetra	general	2	5.44	Um		128132	5.47	0.17	NA		Aligned with Triticum aestivum chromosome 3B	0.09
Dry	FGW	Dip	1-dom-alt	2	5.34	Sb	Chr02	1588978	5.59	0.18	-172.5	LOC8084285	Uncharacterized protein	0.05
	REG	Dip	1-dom-alt	0	5.18	Sv	Chr08	7908449	5.92	0.25	0.5	trnD-GUC	tRNA-Asp	0.11
		Tetra	2-dom-alt		5.02				5.9	0.25	0.5			0.13
	NDF	Tetra	general	0	5.21	Um		128132	5.37	0.17	NA		Alligned with Triticum aestivum chromosome 3B	0.11
	IVD	Tetra	general	2	5.44	Um		91613	7.09	0.02	NA		No significant similarity found	0.14
Wet	LIG	Dip	general	0	5.28	Um		8160655	5.46	0.49	NA		Alligned with Triticum aestivum chromosome 3B	0.11
		Tetra			5.21				5.35	0.46	NA			0.13
	IVD	Tetra	2-dom-ref	2	4.81	Um		3259930	5.76	0.41	-2.31	LOC101780209	Uncharacterized protein At3g52155	0.08
						Sv	Chr_08	7908449	6.21	0.25	-3.38	trnD-GUC	tRNA-Asp	0.06

FGW = Field Green Weight, TDM = Total Dry Matter, NDF = Neutral Detergent Fiber, Os = *Oryza sativaI*, Sb = *Sorghum bicolor*, Si = *Setaria italica*, Sv= *Setaria virides*, Um= *Urochloa mock*.

SUPPLEMENTARY TABLE

Table S1. Wald test to fixed effects of Intercept, Genotype, Season, Season/Cut and Season/Cut/Block 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).

Parameters	FGW		REG		CP		NDF		LIG		IVD	
Intercept	41785.70	**	124367.03	**	345430.50	**	1644665.20	**	85046.86	**	849886.60	**
Genotype	7316.89	**	4696.61	**	3582.20	**	2333.22	**	1877.51	**	2196.84	**
Season	3257.40	**	1309.74	**	9.27	**	907.89	**	26.63	**	374.09	**
Season/Cut	873.98	**	5150.87	**	0		0		0		0	
Block	188.72	**	73.27	**	21.39	**	25.42	**	11.76		5.21	

(*) significant at 0.05 or (**) 0.01 level. FGW = Field Green Weight, REG= Regrowth Capacity, CP= Crude Protein, NDF = Neutral Detergent Fiber, LIG= Lignin in Sulfuric Acid, IVD= Vitro Organic Matter Digestibility.

5. ON THE ACCURACY OF GENOMIC PREDICTION MODELS CONSIDERING MULTI-TRAIT AND ALLELE DOSAGE IN *Urochloa* SPP INTERSPECIFIC TETRAPLOID HYBRIDS

ABSTRACT

Currently, there is a lack on the 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 may reduce the accuracy to estimate the 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 multi-trait approach, accounting the correlation between the traits to adjust the 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 considering both single and multi-trait models. Conversely, combining GBLUP-AD and tetraploid information improved 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 can be used in forages breeding programs in order to reduce time.

Keywords: Polyploid; Genotyping-by-sequencing; Tropical forage; Dominance

5.1. 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). Therefore, GP 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 chromosomes 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 (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, once that 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 (Uitdewilligen et al. 2013). Usually, for most of the studies applying the GP in polyploid species, the genotypic information is parametrized at diploid level (Annicchiarico et al. 2015; Biazzi et al. 2017), and only a few studies consider a polyploid parametrization (Sverrisdóttir et al. 2017; Nyine et al. 2018).

The influence of allele dosage in polyploid species was recently evaluated in genomic association studies (Ferrão et al. 2018; Sharma et al. 2018) which concluded that different genomic regions are assessed when using molecular data in diploid versus tetraploid configuration. Thus, the missed information of heterozygous is one of the most critical problems caused when the polyploid genome is simplified as diploidized data (Voorrips et al. 2011; Hackett et al. 2013). For example, in tetraploid species, the genotypic classes simplex ($Aaaa$), duplex ($AAaa$), and triplex ($AAAa$) are summarized in a single class (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 of 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 new to tropical forage species compared to other crops (Hayes et al., 2013). A common group of polyploid forage species in tropical climates is *Brachiaria* and *Panicum*. This genus has a vast importance in promoting "beef farm pastures" in the livestock business (Montagner et al. 2012; Euclides et al. 2016). Additionally, it is wide adaptation to the soil and climate of Brazilian savannals and contributed

to make Brazilian beef vary competitive (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 subtropical countries are apomictic and tetraploids, such as *U. brizantha* cv Marandu(Jank et al. 2014), where an ideal genotype should have excellent agronomical and nutritional performance, to support and to feed cattle (Jank et al. 2011; Euclides et al. 2016).

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 hard-work and an expensive process due to the evaluation of animal performance apart from plant performance (Jank et al, 2014). For instance, one selection cycle in these species demands an average of two 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. A simulation study 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 the use of 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 ones of low heritability (Bauer and Léon 2008; Dos Santos et al. 2016; Fernandes et al. 2018). It has been successfully implemented using single by single trait (Jia and Jannink 2012; Guo et al. 2014) or indices (Schulthess et al. 2016; Lyra et al. 2017).

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 may provide higher simultaneous selection gains for the inversely correlated traits (Bauer and Léon 2008; Guo et al. 2014). Thus, our 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.

5.2. 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, 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 discovery single nucleotides polymorphisms using ploidy = 4 for genotype calling in 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 estimates the average insert size of the single-end reads, respectively. *Urochloa* spp does not have a 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 $\geq 50\%$ were eliminated. Also, samples with DepthPerSample (DP) < 8 were set as missing, and samples with genotype quality (QD) $\leq$ 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.

Phenotypes

The population was evaluated in the field (20°27'S; 54°57'W) for two 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 sexual interespecific 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 scale to determine 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. 1989). The calibration of the NIRS was performed previously by comparing the results obtained in the wet chemical analyzes to the spectrum read from these same samples in the NIRS for several nutritional characteristics (unpublished data).

We employed a two-step approach to estimate the phenotypic record of each hybrid. First, the block effect was estimated by a complete block design considering only the checks as described in the equation [1]. Afterwards, these effects were deducted from the observed data of each hybrid as a function at the field position. Then, the new corrected trait data were evaluated according to the equation [2].

$$y_{bcd} = \mu + p_b + q_c + s_d + u_{b_{\times}c} + \varepsilon_{bcd} \quad [1]$$

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

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 $b = \{1,2,\dots,9\}$; $\boldsymbol{p}^*$ is the vector of hybrids effect, considered as fixed, with $g = \{1,2,\dots,272\}$; $\boldsymbol{q}$ is the vector of cut effect, considered as fixed, with $c = \{1,2,\dots,7\}$ for agronomical traits and $c = \{3,4\}$ for nutritional traits; $\boldsymbol{s}$ is the vector of block effect, considered as fixed, with $d = \{1,2,\dots,10\}$; $\boldsymbol{u}$ is the vector of the check by cut interaction, considered as random, with $\boldsymbol{u} \sim N(0, \boldsymbol{I}\sigma^2_{b_{\times}c})$ 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 effect, considered as random, with $\boldsymbol{u}^* \sim N(0, \boldsymbol{I}\sigma^2_{g_{\times}c})$ where $\sigma^2_{g_{\times}c}$ is the variance component of described interaction; and $\boldsymbol{\varepsilon}$ is the residual vector with $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma^2_{\varepsilon})$ where σ^2_{ε} is the variance component of error.

All models used to obtain the genetic value of each hybrid and its 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:

$$\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{Za} + \boldsymbol{Td} + \boldsymbol{\varepsilon} \text{ [3]}$$

where $\boldsymbol{y}$ is the vector of genetic hybrids values from the equation [2], $\boldsymbol{\mu}$ is the intercept, $\boldsymbol{a}$ is the vector of additive effect with $\boldsymbol{a} \sim N(0, \boldsymbol{G}\sigma_a^2)$, $\boldsymbol{d}$ is the vector of dominance effect with $\boldsymbol{d} \sim N(0, \boldsymbol{D}\sigma_d^2)$, $\boldsymbol{\varepsilon}$ is the residual vector with $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$. σ_ε^2, σ_a^2, and σ_d^2 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:

$$\boldsymbol{W}_{Dip} = \left(\boldsymbol{X}_{Dip} - 2p_i\right)$$

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

Diploid dominance:

$$\boldsymbol{S}_{Dip} = 2p_i\boldsymbol{X}_{Dip} - 2p_i^2 - \boldsymbol{X}_{Dip}(\boldsymbol{X}_{Dip} - 1)$$

$$\boldsymbol{D}_{Dip} = \frac{\boldsymbol{S}_{Dip}\boldsymbol{S}_{Dip}'}{\sum 4p_i^2(1-p_i)^2}$$

Tetraploid additive:

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

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

Tetraploid dominance:

$$\boldsymbol{S}_{Tetra} = 6{p_i}^2 - 3p_i\boldsymbol{X}_{Tetra} + \frac{\boldsymbol{X}_{Tetra}(\boldsymbol{X}_{Tetra} - 1)}{2}$$

$$\boldsymbol{D}_{Tetra} = \frac{\boldsymbol{S}_{Tetra}\boldsymbol{S}_{Tetra}'}{\sum 6p_i^2(1-p_i)^2}$$

where p_i is the reference allele frequency, and $\boldsymbol{X}$ is the allele dosage matrix with genotypes.

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 divided 75% of individuals in a training population (TP) and 25% in validation population (VP). 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 2014) assuming 30,000 Gibbs samples, a *burn-in* of 5,000, 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:

$$y_{ni} = \mu + \beta_i + u_A + u_D + \varepsilon_{ni} \quad [4]$$

where $y_{ni} = (y_{1i}, \dots, y_{ni})'$ is the phenotypic data with i equal the number of traits $i = \{1,2,3,4\}$ and n the number of hybrids $n = \{1,2, \dots ,272\}$; $\boldsymbol{\mu}$ 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 $\boldsymbol{u_A} \sim MVN(0, \boldsymbol{G}\sigma_a^2)$ and $\boldsymbol{u_D} \sim MVN(0, \boldsymbol{D}\sigma_d^2)$; $\boldsymbol{\varepsilon}$ is the residual vector with $\boldsymbol{\varepsilon} \sim MVN(y_{ni}|\eta_{ni}, \boldsymbol{R})$. Where $MVN(.|.,.)$ denotes a multivariate-normal density with mean η_{ni} and covariance matrix $\boldsymbol{R}$; here, ηi is an r-dimensional vector whose entries are the expected phenotypic values of the n^{th} individual for each of the traits.

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 traits. This scheme mimics the situations in which the breeder desires to predict the performance of newly developed materials, without any phenotypic record (Fig.1A). 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 (Fig.1B). For each trait we tested three scenarios assuming a training set (TP) of 75, 50, and 25% of the total population size in a multi-trait genomic prediction approaches. 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 5,000 and *thinn*ed of 5.

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

Genetic gains and validation approach comparations

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 time to conclude the first stage of traditional tropical forage breeding program (four years for PS), and the time to get seedlings to extract DNA of GP (six months). The standardized selection intensity was fixed in 10% ($i = 1.76$). The genomic prediction approach INDG and MTM were compared with the phenotypic selection by the ratio among genetic gains with GP by genetic gains with PS, following the equation: $\Delta G_{genomic:PS} = {\Delta G_{genomic}}/{\Delta G_{PS}}$. In light of this, the values of σ_a and i were considered the same and the genetic gains were estimated only by $\Delta G = \frac{r}{L}$.

5.3. RESULTS

Phenotypic selection

Significant genetic effects were found for studied all traits (Table 1), highlighting the possibility of genetic gains when applying the phenotypic selection in this *Urochloa* spp population. The heritability 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 genetic gain was observed when employing the univariate phenotypic selection for CP (SG%=9.17) and the lowest for NDF (NDF=-0.52) (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. Regarding the univariate phenotypic selection, the population performance for a given trait in the

next generation is estimated by the response to selection $\boldsymbol{SG = \Delta \bar{X} * H^2}$ plus the original population mean $(\boldsymbol{\bar{X}_{pop}})$. Concerning the relation among traits, moderate to low correlations values were observed between them: 0.38^{**} (FGWxREG), -0.04^{ns} (FGWxCP), -0.24^{**} (FGWxNDF), -0.04^{ns} (REGxCP), -0.08^{ns} (REGxNDF) and 0.49^{**} (CPxNDF), where ** means correlation estimate significantly different from zero by the *t*-test and ^{ns} non-significant.

Single-trait genomic prediction

The prediction ability (r) by single-trait genomic models were 0.20, 0.15, 0.13, and 0.31 for crude protein, green weight, fiber, and regrowth capacity, respectively (Fig.2A). Slightly differences were observed among the r estimates obtained by the GBLUP-A and GBLUP-AD using 75/25 cross-validation (INDG). Similarly, for CP and FGW the GBLUP-A prediction ability was equal to or larger than those obtained when fitting the GBLUP-AD. 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 latter led to better results to predict REG and NDFr (Fig.2A).

Breeders should choose the genomic selection model not by only considering the prediction ability but also the selection coincidence. It is the proportion of coincident hybrids selected by the genomic prediction model and phenotypic selection, generally in the tails of a distribution. Thus, despite the small differences observed by r between GBLUP-A and GBLUP-AD, the latter performed better than GBLUP-A by comparing the selection coincidence using 10% of intensity of selection for all scenarios (Fig.2B). GBLUP-AD showed 0-5% higher selection coincidence than GBLUP-A for crude protein, and 5% for regrowth capacity. Concerning the dosage information, the tetraploid level combined to the GBLUP-A, in general, does not affect the selection coincidence. However, when it was combined with GBLUP-AD, the selection coincidence was improved for all traits (Fig.2B).

Multi-trait genomic prediction

The prediction ability of multi-trait genomic prediction models through the validation scheme 1 (VS1) were modest and slightly different, even considering different testing population sizes (Fig. 3A). 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 with the best advantage of increasing the training population size for prediction using a multi-trait framework. For most scenarios, the additive model led to slightly higher prediction accuracies for both

diploid and tetraploid parametrization. However, even with the small differences on the prediction accuracies the additive plus dominance models provided the more substantial selection coincidence, reaching the highest coincidences when combining with the tetraploid matrix.

The validation scheme 2 (VS2) showed the highest prediction abilities for all scenarios (Fig.3A and Fig.4A). For example, accounting only crude protein, the average of multi-trait prediction accuracy by VS1 were 0.16, 0.20, and 0.21 while the values by VS2 were 0.23, 0.33, and 0.37 for training set sizes of 25, 50, and 75%, respectively. The same trend was observed for all traits, where the prediction ability highly increased according to the training population size and scheme.

Regrowth ability showed the highest prediction values using both validations schemes, also for all training/testing populations sizes with prediction accuracies always superior to 0.25 (Fig.3A and Fig.4A). However, the difference concerning the other traits was diminished using VS2. For instance, the difference of prediction ability between REG and NDF, considering VS1 were around 0.10-0.20 and reduced to 0.02-0.06 using VS2.

Overall, the GBLUP-AD performed better than the GBLUP-A regarding the prediction ability, but as observed for single-trait predictions, the differences between the additive and additive-dominance GBLUP were tiny, being the most considerable differences observed in small training populations sizes, such as 25% (Fig.4A). Regarding the ploidy information, the differences in prediction abilities between diploid and tetraploid parametrizations were small for all scenarios. However, there is an increase in the selection coincidence according to the training set size and the ploidy considered to build the relationship matrices. . Furthermore, the tetraploid matrix led to highest coincidence estimates than the diploid matrix, achieving a maximum of 0.63 for FGW (Fig.4B).

Based on the standard error proposed by Bouckaert and Frank (2004), that accounts the size of validation and training population, shows that only when TP=75% and VP=25% there will be a significative difference between GBLUP-A and GBLUP-AD (Fig.3B and Fig.4B).

Genetic Gains

The genetic gain *per* unit of time using genomic prediction approaches was at least 1.383 times more efficient than phenotypic selection (Table 2). Using single-trait models (INDG) to select it is possible to increase the response to selection in 1.543, 3.322, 2.431, and 2.038 times concerning the phenotypic selection for FGW, REG, CP, and NDF, respectively. The gains obtained by multi-trait prediction VS1 were similar to those observed by the INDG approach.

However, in general, the estimates observed by MTM using the VS2 were the highest for all traits (Table 2).

5.4. DISCUSSION

Genomic prediction in polyploids

Currently, there is a lack on the information regarding the employment of genomic prediction in tropical forages breeding when compared to crop and temperate forages species. Furthermore, the genomic prediction has been extensively applied for diploid in detriment of polyploid species (You et al. 2018). The use of allele dosage in genomic prediction is important once that in theory the genetic value (GV_{ijkl}) of one tetraploid genotype $A_iA_jA_kA_l$ from a random mating population at equilibrium may be partitioned according to the equation described by Kempthorne (1957): $GV_{ijkl} = \mu + \alpha i + \alpha j + \alpha k + \alpha l + \beta ij + \beta ik + \beta il + \beta jk + \beta jl + \beta kl + \gamma ijk + \gamma ijl + \gamma ikl + \gamma jkl + \delta ijk$, 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 provided only by $GV_{ij} = \mu + \alpha i + \alpha j + \beta ij$. Therefore, simplify a tetraploid genotype as a diploid might insert a bias on the genomic predictions due to high order interactions and allele substitution effects to be estimated. In our study, we compared the impact of the ploidy parametrization over 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 admitted that these traits are polygenic and 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 set shows the best predictions (Guo et al. 2012; Pérez-Rodríguez et al. 2012; Combs and Bernardo 2013). Thus, we admitted that genome coverage is more important than depth and use the high-quality filtering for diploid to select markers in the quality control process. Our results indicated that the influence of the ploidy in this population depends on the trait. For example, diploid and tetraploid levels showed little differences in prediction ability between them for FGW, CP, and REG (Fig.2A, 3A, and 4A). 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), 60–80x for potato (Uitdewilligen et al. 2013). Alternatively, Griffin et al. (2011) using simulations showed that a read depth of 15x or more is required to identify allele in tetraploids. These depths are necessary once genotype calling in tetraploid species is more challenging than diploid species. In this case, there are five different genotype categories to be distinguish: nulliplex (0, aaaa), simplex (1, Aaaa), duplex (2, AAaa), triplex (3, AAAa), and tetraplex (4, AAAA) (Uitdewilligen et al. 2013). In this situations, low depth of reads is a barrier to cross and identify correctly the simplex, duplex, and triplex, due to be 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 cases when the genomic data has lower reading depth than expected for tetraploid genotype calling, the approach of grouping all "heterozygous" genotypes in one class, *i.e.*, considering the diploid parametrization, seems to be a good strategy to circumvent this problem and apply genomic tools.

As noted above, the dosage information in genomic tools needs to be evaluated carefully. Our results suggest that the use of quality filtration adequate for diploid level and extended for tetraploid level, it means, a couple of markers with a lower depth of reads than recommended for tetraploid species in the literature, was not a problem to perform the genomic predictions in *Urochloa* spp hybrids. As one can observe, the more considerable selection coincidence provide by tetraploid information under the diploid filtration criterions is a strong argument to be more permissive during the genotyping calling step in tetraploid species (Fig.2, 3 and 4). It is important to highlight that this *Urochloa* spp. population is an allotetraploid segmental with part of the genome with autotetraploid configuration and part with allotetraploid configuration (Mendes-Bonato et al. 2002; Worthington et al. 2016), and for a complete autotetraploid species the results could be different.

Slight differences in prediction accuracies between tetra and diploid matrices were found in our population. However, this differs from the findings of Nyine et al. (2018), who observed a significant reduction in the prediction accuracy when considering allele dosage in triploid banana. According to the same authors, this can be due to the variation on the minor allele frequency across loci, with have a significant impact on the estimation of SNP effects. In this context, our study is the first to compare both diploid and tetraploid matrices for interspecific hybrids, but we were more permissive for the tetraploid markers calling. Therefore, further studies accounting for a "recommended" read depth in polyploids 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. In this case, all recovered information 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 high correlations among the considered traits (Guo et al. 2014; Lyra et al. 2017) and broadly increase the prediction accuracy of genomic models. This scenario can be represented by the MTM-Validation Scheme 2, and it is similar to the "trait-assisted GS approach" proposed by Fernandes et al. (2018). However, we considered three traits are assisting the prediction of one. Furthermore, we considered a variation on the training sets the size for the MTM models.

Single and multi-trait models did not show significative differences in their prediction accuracies for the same training population size (Fig.2A, and 3A, TP=75%). This result is in accordance to 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 trait. It is similar to the CV1 proposed by Burgueño et al. (2012) and does not retrieve information among traits, commonly observed for multi-environments prediction models in crops (Lopez-Cruz et al. 2015; Souza et al. 2017). Thus, these similarities among prediction accuracies were expected.

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 (Figures 2A, 3A, and 4A). This result is in accordance to 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 the MTM-VS2 provided higher prediction accuracies when we used the smallest training population size (TP=25%). Furthermore, it suggest that a small number of individuals should be phenotyped for the "assisted trait" to reach larger accuracies estimates than the single (INDG) and multi-trait (MTM-VS1).

As already described no high correlation was observed among the traits. The highest value was between crude protein and fiber (0.49), followed by the correlation between green weight and regrowth (0.38). Also, no significant correlation was observed between FGW and CP, REG and CP, and REG and NDF. Thus, the absence of all phenotypic information of individuals in MTM-VS1 was a challenge to modeling than MTM-VS2, due to the model was using low values in the variance and covariance matrix of traits.

The prediction ability of INDG and MTM were lower than 0.45 for all traits (Fig.2 and Fig.3). 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 FGW 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 the *Urochloa* genome make it challenging to cover the whole genome.

GBLUP-A and GBLUP-AD models presented similar prediction abilities (Fig.2A, 3A, and 4A). In general, GBLUP-A was slightly superior to GBLUP-AD using the INDG approach. Probably, this dataset was not enough to cover and capture the nonadditive effects accurately, like was observed in alfalfa (Biazzi et al. 2017). In our study, we cannot conclude precisely the influence of additive and non-addictive marker effects on the prediction of new materials. Despite of this, the inclusion of non-additive 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), 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 (Fig.2B, 3B, and 4B). We observed that more levels of alleles dosages provide subtle but different configuration of additive and non-additive kinship matrix. These differences were not enough to improve the prediction accuracy of the models but could explain better 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. Indeed, using a tetraploid level, there are many more parameters to be estimated. Thus, even biologically correct, depending on the number and depth of markers, it is not possible to estimate all the genetic effects with accuracy. 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. Consequently, as the lower limit of the tetraploid model is diploid, we strongly recommend the former in genomic predictions.

The genomic prediction fes in *Urochloa* spp. breeding programs

Forages breeding programs focus on improved yield and quality of herbage aiming at 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 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 applying in tropical forages. We compared genetic gains achieved for critical agronomic traits when GP is used in a tropical forage breeding program (Table.2). Instead the low values of prediction accuracy, the gains in a unit of time provided for all scenarios larger genetic gains for genomic selection (GS) when compared to the phenotypic selection (Table.2). It is in accordance to 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 underdeveloped 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 forages 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 others different strategies were recommended in literature as use multi-trait models (Guo et al. 2014). For our dataset, the MTM approach improves significantly the prediction values, manly

using MTM-VS2 that also provide the most significant genetic gains (Table.2). INDG and MTM-VS1 had similar prediction abilities and genetic gains.

The *Urochloa* breeding program from the EMBRAPA beef-cattle follow-up the breeding scheme described by Jank et al. (2014). Each stage takes at least two years and adding one year for seed multiplication between them. The whole process takes 10–15 years. In a traditional program, the focus in the initial stage is to obtain new genotypes, in the intermediate stage is selected, 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 two or four years would 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 one 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 around $180 per sample (personal communication). Therefore, by computing the costs on the genetic gains' equation, the advantages to use genomic prediction approaches would be even higher than 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.

REFERENCES

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, et al (2015) Accuracy of genomic selection for alfalfa biomass yield in different reference populations. BMC Genomics 16:1020. doi: 10.1186/s12864-015-2212-y

Bassil N V., Davis TM, Zhang H, et al (2015) Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria × ananassa. BMC Genomics. doi: 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. doi: 10.1007/s00122-007-0662-6

Bennetzen, JL, Schmutz J, Wang H, et al (2012) Reference genome sequence of the model plant Setaria. Nat. Biotechnol. 30:555

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

Bouckaert RR, Frank E (2004) Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. Adv Knowl Discov data Min 3–12. doi: 10.1007/978-3-540-24775-3

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. doi: 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. doi: 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, et al (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. doi: 10.1038/hdy.2016.23

de los Campos G, Hickey JM, Pong-Wong R, et al (2013) Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics 193:327–345. doi: 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. doi: 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.

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

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. doi: 10.1371/journal.pone.0019379

Endelman JB, Carley CAS, Bethke PC, et al (2018) Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics. doi: https://doi.org/10.1534/genetics.118.300685

Euclides VPB, Montagner DB, Barbosa RA, et al (2016) Animal performance and sward characteristics of two cultivars of *Brachiaria brizantha* (BRS Paiaguás and BRS Piatã). 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. doi: 10.1007/s00122-017-3033-y

Ferrão LF V., 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. doi: 10.3389/fevo.2018.00107

Figueiredo UJ de, Nunes JAR, Valle CB do (2012) Estimation of genetic parameters and selection of *Brachiaria humidicola* progenies using a selection index. Crop Breed Appl Biotechnol 12:237–244.

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

Griffin PC, Robin C, Hoffmann AA (2011) A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to Poa grasses. BMC Biol 9:19. doi: 10.1186/1741-7007-9-19

Guo G, Zhao F, Wang Y, et al (2014) Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet. doi: 10.1186/1471-2156-15-30

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

Hackett CA, Bradshaw JE, Bryan GJ (2014) QTL mapping in autotetraploids using SNP dosage information. Theor Appl Genet 127:1885–1904. doi: 10.1007/s00122-014-2347-2

Hackett CA, McLean K, Bryan GJ (2013) Linkage Analysis and QTL Mapping Using SNP Dosage Data in a Tetraploid Potato Mapping Population. PLoS One. doi: 10.1371/journal.pone.0063939

Hayes BJ, Cogan NOI, Pembleton LW, et al (2013) Prospects for genomic selection in forage plant species. Plant Breed. 132:133–143.

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

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

Jank L, Barrios SC, do Valle CB, et al (2014) The value of improved pastures to Brazilian beef production. Crop Pasture Sci 65:1132–1137.

Jank L, Valle C, Resende R (2011) Breeding tropical forages. Crop Breed Appl Biotechnol S1:27–34. doi: 10.1590/S1984-70332011000500005

Jia Y, Jannink J-L (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–22. doi: 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.

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, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Li X, Wei Y, Acharya A, et al (2015) Genomic Prediction of Biomass Yield in Two Selection Cycles of a Tetraploid Alfalfa Breeding Population. Plant Genome 8:0. doi: 10.3835/plantgenome2014.12.0090

Li H, and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.

Liaw a, Wiener M (2002) Classification and Regression by randomForest. R news 2:18–22. doi: 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: 10.1534/g3.114.016097

Lyra DH, de Freitas Mendonça L, Galli G, et al (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. Mol Breed. doi: 10.1007/s11032-017-0681-1

Marten GC, Shenk JS, Barton FE (1989) Near-infrared reflectance spectroscopy (NIRS): Analysis of forage quality. Agric Handb 95.

Martin M (2011) Cutadapt removes adapter sequence from high-throughput sequencing reads. EMBnet.journal 17:10–12.

Matias F, Barrios S, Bearari Lucas; Valle C Do, 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 FI, Barrios SCL, Valle CB do, et al (2016) Estimate of genetic parameters in *Brachiaria decumbens* hybrids. Crop Breed Appl Biotechnol 16:115–122.

McKenna A, Hanna M, Banks E, 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

Mendes-Bonato AB, Pagliarini MS, Forli F, et al (2002) Chromosome numbers and microsporogenesis in *Brachiaria brizantha* (Gramineae). Euphytica 125:419–425. doi: 10.1023/A:1016026027724

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

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

Montagner DB, Nascimento Júnior D do, 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, et al (2018) Genomic Prediction in a Multiploid Crop: Genotype by Environment Interaction and Allele Dosage Effects on Predictive Ability in Banana. Plant Genome 11:0. doi: 10.3835/plantgenome2017.10.0090

Ouyang S, Zhu W, Hamilton J, et al (2006) The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res. 35: D883--D887.

Peng Z, Fan W, Wang L, et al (2017) Target enrichment sequencing in cultivated peanut (*Arachis hypogaea* L.) using probes designed from transcript sequences. Mol Genet Genomics 292:955–965. doi: 10.1007/s00438-017-1327-z

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

Pérez-Rodríguez P, Gianola D, González-Camacho JM, et al (2012) Comparison Between Linear and Non-parametric Regression Models for Genome-Enabled Prediction in Wheat. G3 Genes|Genomes|Genetics 2:1595–1605. doi: 10.1534/g3.112.003665

Resende RMS, Casler MD, de Resende MDV, et al (2014) Genomic selection in forage breeding: accuracy and methods. Crop Sci 54:143–156. doi: 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 1–10. doi: 10.3835/plantgenome2015.08.0073

Schmitz Carley CA, Coombs JJ, Douches DS, et al (2017) Automated tetraploid genotype calling by hierarchical clustering. Theor Appl Genet 130:717–726. doi: 10.1007/s00122-016-2845-5

Schnable PS, Ware D, Fulton RS, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schulthess AW, Wang Y, Miedaner T, et al (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. doi: 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. doi: 10.1371/journal.pone.0030906

Sharma SK, MacKenzie K, McLean K, et al (2018) Linkage Disequilibrium and Evaluation of Genome-Wide Association Mapping Models in Tetraploid Potato. G3 (Bethesda) g3.200377.2018. doi: 10.1534/g3.118.200377

Song J, Yang X, Resende MFR, et al (2016) Natural Allelic Variations in Highly Polyploidy *Saccharum* Complex. Front Plant Sci. doi: 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. Genes|Genomes|Genetics g3.117.042341. doi: 10.1534/g3.117.042341

Sverrisdóttir E, Byrne S, Sundmark EHR, et al (2017) Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. Theor Appl Genet 130:2091–2108. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1186/1471-2105-12-172

Worthington M, Heffelfinger C, Bernal D, et al (2016) A Parthenogenesis Gene Candidate and Evidence for Segmental Allopolyploidy in Apomictic Brachiaria decumbens. Genetics 116.

You Q, Yang X, Peng Z, et al (2018) Development and Applications of a High Throughput Genotyping Tool for Polyploid Crops: Single Nucleotide Polymorphism (SNP) Array. Front Plant Sci. doi: 10.3389/fpls.2018.00104

FIGURES

A

$$\begin{matrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{ni} \end{matrix} = \begin{bmatrix} Trait\_1 & Trait\_2 & Trait\_3 & Trait\_4 \\ \boldsymbol{NA} & \boldsymbol{NA} & \boldsymbol{NA} & \boldsymbol{NA} \\ \boldsymbol{NA} & \boldsymbol{NA} & \boldsymbol{NA} & \boldsymbol{NA} \\ \vdots & \vdots & \vdots & \vdots \\ value & value & value & value \end{bmatrix}$$

B

$$\begin{matrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{ni} \end{matrix} = \begin{bmatrix} Trait\_1 & Trait\_2 & Trait\_3 & Trait\_4 \\ \boldsymbol{NA} & value & value & value \\ \boldsymbol{NA} & value & value & value \\ \vdots & \vdots & \vdots & \vdots \\ value & value & value & value \end{bmatrix}$$

$$\vdots$$

$$\begin{matrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{ni} \end{matrix} = \begin{bmatrix} Trait\_1 & Trait\_2 & Trait\_3 & Trait\_4 \\ value & value & value & \boldsymbol{NA} \\ value & value & value & \boldsymbol{NA} \\ \vdots & \vdots & \vdots & \vdots \\ value & value & value & value \end{bmatrix}$$

Fig.1 - Multi-trait model validation schemes, A) an individual is not evaluated for any traits (VS1); B) one trait is not evaluated for part of individuals (VS2).

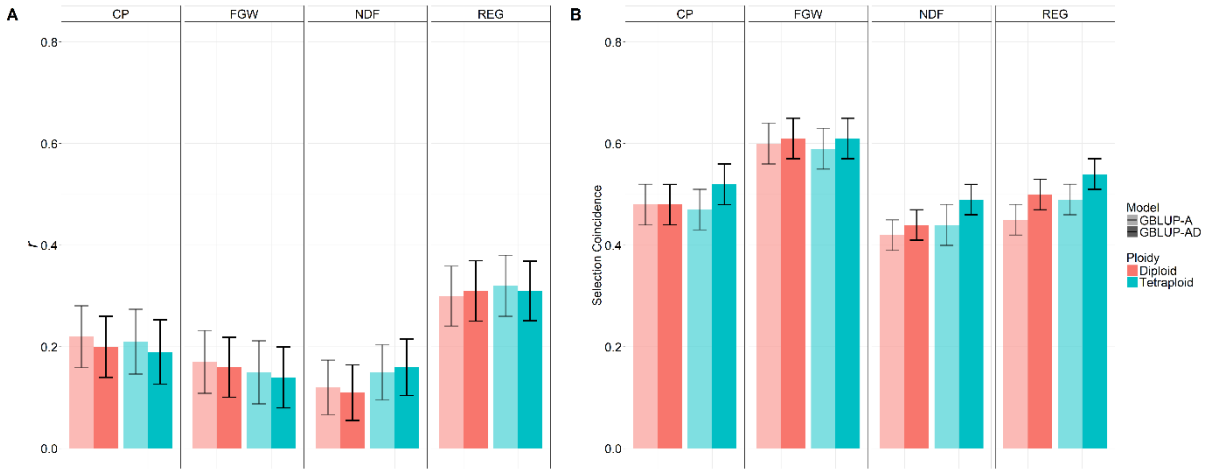

Fig.2 – INDG univariate (75/25) genomic prediction approach: (A) Predictive ability of INDG genomic prediction and (B) selection coincidence between the10% of the best hybrids selected by phenotypic selection and by genomic prediction carried out using two prediction models (GBLUP-A and GBLUP-AD) and 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.

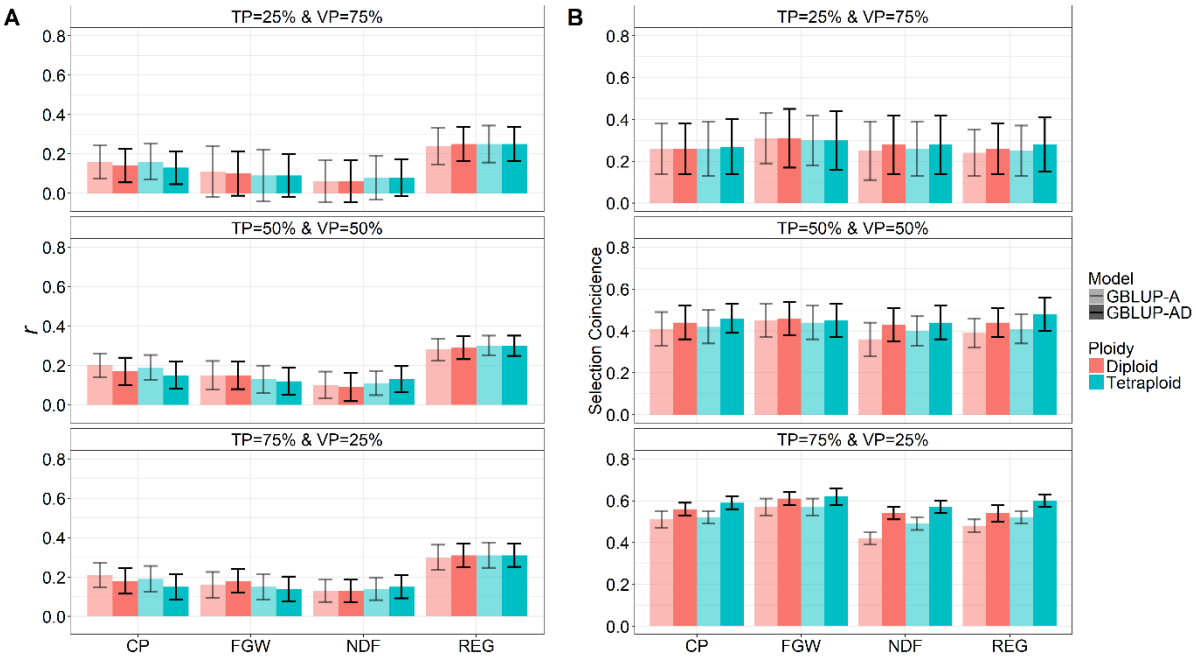

Fig.3 – Validation Scheme 1 (VS1) for Multi-Trait genomic prediction approach: (A) Predictive ability of INDG genomic prediction and (B) selection coincidence between the ten percent of the best hybrids selected by phenotypic selection and by genomic prediction carried out using two prediction models (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% & VP=25%}, 2 - {TP=50% & VP=50%} and 3 - {TP=25% & VP=75%}.

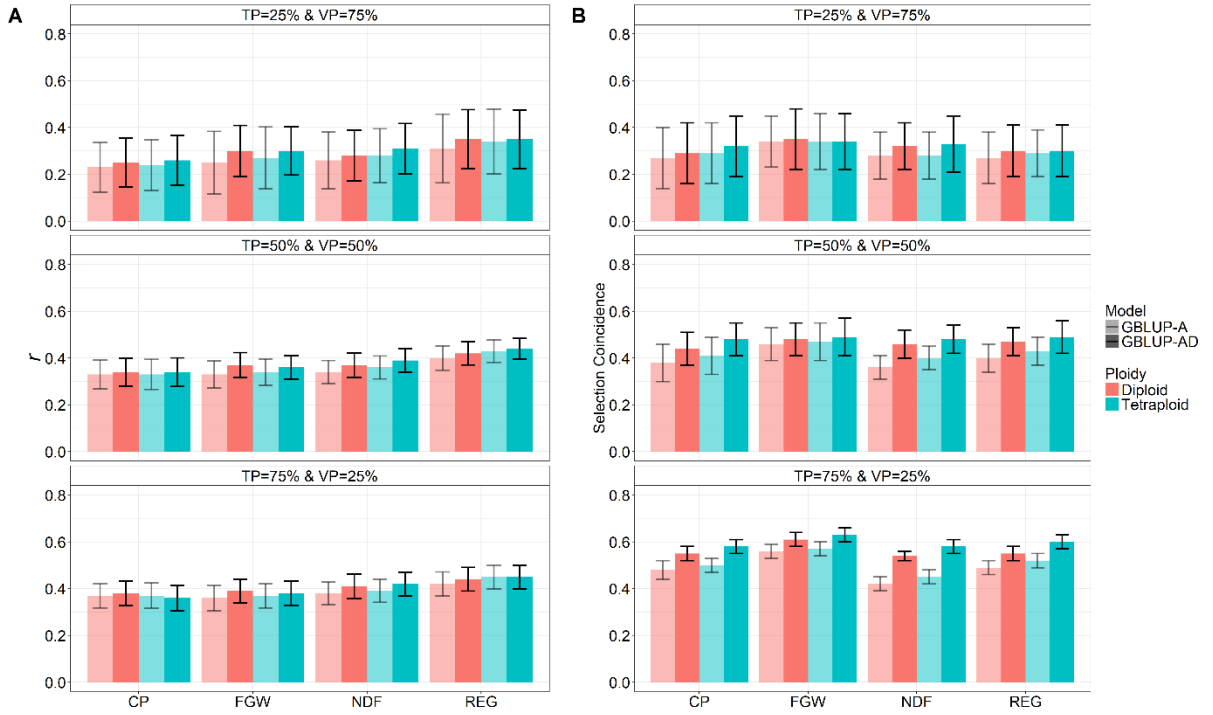

Fig.4 – Validation Scheme 2 (VS2) for Multi-Trait genomic prediction approach: (A) Predictive ability of INDG genomic prediction and (B) selection coincidence between the ten percent of the best hybrids selected by phenotypic selection and by genomic prediction carried out using two prediction models (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% & VP=25%}, 2 - {TP=50% & VP=50%} and 3 - {TP=25% & VP=75%}.

TABLES

Table 1. Wald test for fixed effects of Genotype, broad heritability ($\boldsymbol{H^2}$), average of population ($\boldsymbol{\bar{X}_{pop}}$), average of 10% best hybrids ($\boldsymbol{\bar{X}_{10\%\,Best}}$), average of 10% worst hybrids ($\boldsymbol{\bar{X}_{10\%\,Worst}}$), selection differential ($\boldsymbol{\Delta\,\bar{X} = \bar{X}_{10\%\,Best} - \bar{X}_{pop}}$) and response to selection ($\boldsymbol{SG\% = \Delta\,\bar{X} * H^2/\bar{X}_{pop}}$) 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 **
$\boldsymbol{H^2}$	0.81	0.75	0.68	0.52
$\boldsymbol{\bar{X}_{pop}}$	1468.26	3.23	15.80	65.91
$\boldsymbol{\bar{X}_{10\%\,Best}}$	1614.57	3.47	17.93	59.32
$\boldsymbol{\bar{X}_{10\%\,Worst}}$	447.88	1.83	13.11	70.81
$\boldsymbol{\Delta\,\bar{X}}$	146.31	0.24	2.13	-6.59
$\boldsymbol{SG\%}$	8.07	5.57	9.17	-0.52

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
	$\Delta G_{genomic}$	0.347	0.719	0.501	0.367
	$\Delta G_{genomic:PS}$	**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
	$\Delta G_{genomic}$	0.311	0.693	0.509	0.416
	$\Delta G_{genomic:PS}$	**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
	$\Delta G_{genomic}$	0.822	0.878	0.970	1.220
	$\Delta G_{genomic:PS}$	**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 ($\Delta G_{genomic}$) and ratio gain ($\Delta G_{genomic:PS}$).