JOÃO ROMERO DO AMARAL SANTOS DE CARVALHO ROCHA

ELEPHANTGRASS BREEDING FOCUSED ON PERSISTENCE AND DISCOVER CANDIDATE GENES

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de Doctor Scientiae.

VIÇOSA
MINAS GERAIS – BRASIL
2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
R672e
2019

Rocha, João Romero do Amaral Santos de Carvalho, 1990-
Elephantgrass breeding focused on persistence and discover candidate genes / João Romero do Amaral Santos de Carvalho Rocha. – Viçosa, MG, 2019.
xi, 63 f. : il. (algumas color.) ; 29 cm.

Texto em inglês.
Inclui apêndices.
Orientador: Pedro Crescêncio Souza Carneiro.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. *Cenchrus purpureus*. 2. Colheita. 3. Genoma. 4. Biocombustíveis. 5. Alimentação dos animais. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 584.9

JOÃO ROMERO DO AMARAL SANTOS DE CARVALHO ROCHA

ELEPHANTGRASS BREEDING FOCUSED ON PERSISTENCE AND DISCOVER CANDIDATE GENES

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de Doctor Scientiae.

APROVADA: 07 de março de 2019.

Luiz Antônio dos Santos Dias

Paulo César Emiliano

Marcos Vinicius Gualberto Barbosa da Silva

Juarez Campolina Machado
(Coorientador)

Pedro Crescêncio Souza Carneiro
(Orientador)

Aos meus pais, Roméro e Magda,
pelo empenho e apoio incondicional
para que eu chegasse até aqui.

DEDICO.

AGRADECIMENTOS

Primeiramente a Deus, pela vida e saúde.

Aos meus pais Roméro de Carvalho Rocha e Magda do Amaral Santos, à minha irmã Ana Carolina do Amaral Santos de Carvalho Rocha, meus eternos agradecimentos.

A Isadora Santos e Oliveira, minha namorada, incentivadora e companheira de todos os momentos.

Ao Professor Pedro Crescêncio Souza Carneiro, pela orientação, ensinamentos, disposição e amizade.

Ao Pesquisador Juarez Campolina Machado, pela amizade, pela oportunidade, pelos ensinamentos e pelas valiosas sugestões durante a elaboração de todos os trabalhos científicos.

Aos Professores/Pesquisadores Luiz Antônio dos Santos Dias, Paulo César Emiliano e Marcos Vinicius Gualberto Barbosa e Silva, pela disposição, pelo suporte e valiosa ajuda na finalização desse trabalho.

Aos amigos do Laboratório de Biometria e do Programa Feijão, especialmente aos amigos Tiago de Souza Marçal e Rodrigo Silva Alves.

Ao Programa de Pós-Graduação em Genética e Melhoramento e à Universidade Federal de Viçosa (UFV) pela oportunidade concedida.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudos.

E a todos, que de forma direta ou indireta, contribuíram para a realização deste trabalho.

MUITO OBRIGADO!

BIOGRAFIA

JOÃO ROMERO DO AMARAL SANTOS DE CARVALHO ROCHA, filho de Roméro de Carvalho Rocha e Magda do Amaral Santos, nasceu em 18 de julho de 1990 na cidade de Barra do Piraí, Rio de Janeiro, Brasil.

Em 2008, completou o ensino médio pelo Colégio Meta de Ensino Fundamental e Médio em Juiz de Fora - MG.

No ano de 2014, foi diplomado Engenheiro Agrônomo pela Universidade Federal de Viçosa (UFV), Viçosa-MG.

Em março de 2014, ingressou no curso de Mestrado em Genética e Melhoramento da Universidade Federal de Viçosa (UFV), Viçosa-MG, vindo a defender o título de Magister Scientiae, em 2015.

Em agosto de 2015, ingressou no curso de Doutorado em Genética e Melhoramento da Universidade Federal de Viçosa (UFV), Viçosa-MG, vindo a defender o título de Doctor Scientiae, em 2019.

SUMÁRIO

ABSTRACT ..... **viii**

RESUMO ..... **x**

GENERAL INTRODUCTION **1**

REFERENCES **3**

CHAPTER 1 ..... **5**

GENETIC INSIGHTS INTO ELEPHANTGRASS PERSISTENCE FOR BIOENERGY PURPOSE 5

ABSTRACT **6**

RESUMO **7**

1. Introduction **8**

2. Materials and methods **9**

2.1. Location and experimental conditions 9

2.2. Genetic material and experimental design 10

2.3. Measurement - Biomass yield 10

2.4. Statistical Analyses 10

2.4.1. Random regression model 10

2.4.2. Choice of the best fitted model 11

2.4.3. Extracting the genetics information 12

2.4.3.1. Variance components 12

2.4.3.2. Estimated genetic values 12

2.4.3.3. Accuracy 12

2.4.3.4. Eigenfunctions 13

2.4.3.5. Clones' persistence 13

2.5. Software 13

3. Results **14**

3.1. The best fitted model and the general genetic behavior 14

3.2. Heritability, genetic variance, phenotypic variance, and permanent variance trajectory for biomass yield over the multi-harvest 15

3.3. Elephantgrass clones' persistence 17

4. Discussions **18**

4.1. The best-fitted model 19

4.2. Genetic variability and general genetic behavior 19
4.3. Permanent environment, plasticity, and persistence 20
4.4. Insights about G x E interaction - driving the selection 21
4.5. Lowest persistence - supported by the genome dosage 22
4.6. Clones' persistence applied to bioenergy industries 23
5. Conclusions 25
6. References 25
7. Supporting Information 31
CHAPTER 2 34
DISCOVERING CANDIDATE GENES UNDERLYING BIOMASS DIGESTIBILITY IN ELEPHANTGRASS 34
ABSTRACT 35
RESUMO 36
1. Introduction 37
2. Materials and methods 38
2.1. Genetic material and experimental information 38
2.2. Phenotypic traits 39
2.3. Genotyping, quality control, and imputation 40
2.4. Single-step genome-based best linear unbiased prediction (ssGBLUP) 40
2.5. Simple repeatability plus G x C interaction model 42
2.6. Genome association study 43
2.7. Nucleotide sequence alignment, candidate genes and gene annotation 44
3. Results 44
3.1. Genetic variation, genotype by cutting interaction and accuracy under ssGBLUP model 44
3.2. Overall means and accuracy considering one cutting at a time 46
3.4. Genome association study 47
3.5. Annotation of M28_161 and M35_202 markers, in silico pathway analysis and allelic contribution 49
4. Discussion 51
4.1. Genetic variation, genotype by cutting interaction and prediction accuracy 51
4.2. Trait-marker association analysis 52
4.3. Markers' annotation and in silico pathway 53
5. Conclusions 56

6. References **56**
7. Supporting Information **62**
GENERAL CONCLUSIONS **63**

ABSTRACT

ROCHA, João Romero do Amaral Santos de Carvalho, D.Sc., Universidade Federal de Viçosa, March, 2019. **Elephantgrass breeding focused on persistence and discover candidate genes.** Adviser: Pedro Crescêncio Souza Carneiro. Co-advisers: Marcos Deon Vilela de Resende and Juarez Campolina Machado.

Due to environmental issues the world has been concerned with developing hazard mitigation plans. In the context of alternative energies, the biomass is studied from several candidate species (dedicated energy crops). Elephantgrass has been a notable option as a multi-purpose crop, e.g. bio-based products, co-products and biofuels, besides being used for animal feeding, this is mainly due to high photosynthetic efficiency, high biomass production, longevity, rapid growth, broad adaptation, desirable chemical properties and persistence. Regardless breeding efforts are imperative to contribute to Brazilian energetic matrix and animal production. To the best of our knowledge, in the bioenergy context, there are private companies that already installed in Brazil to use elephantgrass biomass to generate electricity (by the combustion of biomass) with installed capacity sufficient to supply a city with 200 thousand inhabitants of energy demand. While in the animal feeding context, it has been reported that the biomass digestibility can impact the animal performance since a small increase in dry matter digestibility (1%) can increase the daily weight gains in 3.2% for beef cattle. The first chapter (Genetic insights into elephantgrass persistence for bioenergy purpose) focused on assessing the biomass yield persistence for bioenergy purpose of 100 elephantgrass clones measured in six growth seasons in Brazil. To assess the clones' persistence, an index based on random regression models and genotype-ideotype distance was proposed. Results suggested the existence of wide genetic variability between elephantgrass clones, and that the yield trajectories along the harvests generate genetic insights into elephantgrass clones' persistence and G x E interaction. A gene pool that acts over the biomass yield (regardless of the harvest) was detected, as well as other gene pools, which show differences on genes expression (these genes are the major responsible for clones' persistence).It is noteworthy that the methodology used (random regression models) is adequate to achieve the aim of the work, as well as it can be applied with advantages in the study of the adaptability and

stability of any crop when compared with the methodologies until now used. The second chapter (Discovering candidate genes underlying biomass digestibility in elephantgrass) brings information about the first trait-marker association study reported for the elephantgrass. It was compared the single-step genome-based best linear unbiased prediction (ssGBLUP - including the genomic relationship) and the simple repeatability plus genotypes by cuttings interaction models. It was verified that genomic information allowed increases the accuracy for biomass quality traits on elephantgrass, even with a small number of markers. We found two SSR markers associated to biomass digestibility and several candidate genes that have functions involved in biosynthesis of cell wall molecules. These markers are relevant and their use can be crucial to accelerate the elephantgrass genetic breeding. In this sense, identify candidate genes that can be used through marker-assisted selection can help to develop selection procedures that optimize elephant grass breeding for different uses.

RESUMO

ROCHA, João Romero do Amaral Santos de Carvalho, D.Sc., Universidade Federal de Viçosa, março de 2019. **Melhoramento do capim-elefante focado na persistência e na descoberta de genes candidatos.** Orientador: Pedro Crescêncio Souza Carneiro. Coorientadores: Marcos Deon Vilela de Resende e Juarez Campolina Machado.

Devido a questões ambientais, o mundo tem se preocupado em desenvolver planos de mitigação de riscos. No contexto de energias alternativas, a biomassa é estudada a partir de várias espécies candidatas (cultivos energéticos dedicados). O capim-elefante tem sido uma opção notável como cultura para múltiplos usos, isto é, produtos de base biológica, co-produtos e biocombustíveis, além de ser utilizado na alimentação animal. Isto se deve principalmente a alta eficiência fotossintética, alta produção de biomassa, longevidade, rápido crescimento, ampla adaptação, propriedades químicas desejáveis e persistência. Não obstante, esforços de melhoramento são necessários para contribuir com a diversificação da matriz energética brasileira e para a produção animal. Para um melhor entendimento, no contexto bioenergético, existem empresas privadas que se instalaram no Brasil para utilizar a biomassa do capim-elefante para gerar eletricidade (por combustão da biomassa) com capacidade instalada suficiente para suprir a demanda de energética de uma cidade com 200 mil habitantes. Enquanto no contexto da alimentação animal, tem sido relatado que a digestibilidade da biomassa pode impactar no desempenho animal, uma vez que um pequeno aumento na digestibilidade da matéria seca (1%) pode aumentar os ganhos de peso diários para bovinos de corte em 3,2%. O primeiro capítulo (Compreensões genéticas da persistência do capim-elefante para fins de bioenergéticos) teve como foco a avaliação da persistência de 100 clones de capim-elefante para fins bioenergéticos, medidos em seis safras, no Brasil. Para avaliar a persistência dos clones, foi proposto um índice baseado em modelos de regressão aleatória e na distância genótipo-ideótipo. Os resultados sugeriram a existência de ampla variabilidade genética entre os clones de capim-elefante, e que as trajetórias de produtividade ao longo das colheitas geram compreensões genéticas sobre a persistência dos clones de capim-elefante e sobre a interação GxE. Um pool genético que atua sobre a produção de biomassa (independentemente da colheita) foi detectado, assim como outro pool gênico, que mostrou expressão diferencial de genes (estes genes são os principais responsáveis pela persistência dos clones). Vale ressaltar que a metodologia

utilizada (modelos de regressão aleatória) é adequada para se alcançar o objetivo do trabalho, bem como pode ser aplicada com vantagens no estudo da adaptabilidade e estabilidade de qualquer cultura, quando comparados com as metodologias até agora utilizadas. O segundo capítulo (Descobrindo genes candidatos relacionados à digestibilidade da biomassa em capim-elefante) traz informações sobre o primeiro estudo de associação marcador-característica relatado para o capim-elefante. Foram comparados os modelos single-step genome-based best linear unbiased prediction (ssGBLUP - incluindo o ralacionamento genômico) e o modelo de repetibilidade simples adicionado da interação genótipos por cortes. Verificou-se que a informação genômica possibilitou acréscimos na acurácia das características de qualidade da biomassa no capim-elefante, mesmo com um pequeno número de marcadores. Foram encontrados dois marcadores SSR (single sequence repeats) associados à digestibilidade da biomassa e vários genes candidatos que possuem funções envolvidas na biossíntese de moléculas da parede celular. Esses marcadores são relevantes e seu uso pode ser crucial para acelerar o melhoramento genético do capim-elefante. Nesse sentido, identificar genes candidatos que podem ser utilizados por meio da seleção assistida por marcadores podem ajudar a desenvolver procedimentos de seleção que otimizem o melhoramento do capim-elefante para diferentes usos.

GENERAL INTRODUCTION

The elephantgrass [Pennisetum purpureum Schumach.; Syn. Cenchrus purpureus (Schumach.) Morrone], is a very important forage in Brazil, for livestock, having its main use as cut and carry for the cattle (Botrel et al. 2000; Meinerz et al. 2011). Among the tropical grasses, elephantgrass stands out for its high yield and forage quality. It can be used in a variety of forms (pasture, grazing, silage, etc.), being one of the forages that most contributes for milk production in Central Region of Brazil (Botrel et al. 2000; Meinerz et al. 2011).

Due to environmental issues the world has been concerned with developing hazard mitigation plans. The composition of the world energy matrix is still strongly influenced by non-renewable sources of energy, with petroleum as the main raw material (EPE, 2016). According to the Brazilian Ministry of Mines and Energ (2016), the total demand for oil products has been reduced by 6.3% from 2014 to 2015, and this is in part to stimulating the renewable energy production.

In the context of alternative energies, the biomass has been studied with several candidate species (dedicated energy crops). Elephantgrass has been a notable option as dedicated energy crop, since the dedicated species must present photosynthetic efficiency, high biomass production, longevity, rapid growth, broad adaptation, desirable chemical properties (Anderson et al. 2008), nitrogen fixation ability (Morais et al. 2009) and persistence (Rocha et al. 2018). Breeding efforts may contribute to Brazilian energetic matrix diversification.

Pre-breeding actions such as the characterization, evaluation and exploration of the variability in elephantgrass germplasm have already been performed by Shimoya et al. (2001), Techio et al. (2002), Azevedo et al. (2012) and Rocha et al. (2017) at the phenotypic, genotypic and cytogenetic levels. In all these studies genetic variability was verified in elephantgrass, being this the basic premise for the beginning of any genetic breeding program.

Breeding efforts should be prioritized and may contribute to the Brazilian energy matrix diversification. Thus, application of genomic tools in forage breeding programs can reduce the time required for the development a superior cultivar reaching higher genetic gain, mainly for quantitative traits (Resende et al. 2012).

Studies with molecular markers began to present even greater importance from the development of single nucleotide polymorphisms (SNP) markers. Since then, new

technologies of genotyping, sequencing and new bioinformatics tools have allowed greater accuracy of information and lower price per sequenced base or molecular genotype, and reduced the selection time, which has resulted in an exponential growth of genomic information in plants (Pereira et al. 2018).

In general, there is greater availability of genomic information in species that can be used as temperate forages. The use of molecular markers in tropical forages is a fairly recent activity, and therefore, there are few results. Most tropical forage uses RAPD and SSR markers for short-term applications, generating knowledge advances about genomic diversity, cross-rate determination, and fingerprint identification (Azevedo et al. 2012, Negawo et al. 2017, Neagwo et al. 2018). In the medium to long term, there are few studies with tropical forages such as: identification of genomic regions associated with important agronomic traits, genetic maps, marker assisted selection and genomic selection.

Genomic information has modified the plant breeding format, instead of being a dependent process almost exclusively of phenotyping, it has become increasingly dependent of genotyping data (Varshney et al. 2014).

It is important to emphasize that genomics can aid the selection processes, but will not replace traditional plant breeding programs. Thus, investments in genomics should not be made in detriment of conventional genetic breeding or agronomic research, and the breeding programs should remain as protagonists (FAO 2011).

In this context, the forages genetic breeding associated to genomic tools will be one of the main vectors for new cultivars development. In this way, positive impacts are expected both on national livestock production and on the Brazilian energetic matrix diversification (Negawo et al. 2017, Neagwo et al. 2018, Pereira et al. 2018).

It should be noted that there are private companies that have already settled in Brazil to use elephantgrass biomass to generate electricity (by biomass combustion) with installed capacity sufficient to supply a city with 200 thousand inhabitants of energy demand (Sykué Bioenergia, 2019). In this sense, genetic information about elephantgrass persistence and genomic tools application are essential to ensure the energy supply and the breeding programs success.

Thus, the present work was carried out with the objective of generating genetic information about persistence, genotypes by cuttings interaction and marker-trait association destined to accelerate the genetic breeding programs of elephantgrass specific to the uses as fodder or as bioenergy crop.

REFERENCES

Anderson W, Casler M, Baldwin B. Improvement of perennial forage species as feedstock for bioenergy. In: Vermerris, W. (Ed.), **Genetic Improvement of Bioenergy Crops.** Springer, 2008; 308-345.

Azevedo ALS, Costa PP, Machado JC, Machado MA, Pereira AV, Lédo FJDS. Cross-species amplification of Pennisetum glaucum microsatellite markers in Pennisetum purpureum and genetic diversity of napier grass accessions. **Crop Sci** 2012;52, 1776-1785.

Botrel MA; Pereira AV, Freitas VP, Xavier DF. Potencial forrageiro de novos clones de capim-elefante. **Rev Bras Zoo.** 2000;29: 334-340.

Energy Research Company (EPE). **2016 Statistical Yearbook of electricity 2015 baseline year.** 230p. Brasília, DF, 2016.

Food and Agriculture Organization of United Nations (FAO, 2011). **Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines.** No. 9. Rome, 2011.

Meinerz GR, Olivo CJ, Agnolin CA, Moraes RS, Dullius AP, Mombach G, et al. Produção e valor nutritivo da forragem do capim-elefante em dois sistemas de produção. **Rev Bras Zoo.** 2011;40: 2673-2680.

Ministério das Minas e Energia. **Resenha energética brasileira, exercício de 2015.** 29p. Brasília, DF, 2016.

Morais RF, Souza JB, Leite JM, Soares LHB, Alves BJR, Boddey RM, et al. Elephant grass genotypes for bioenergy production by direct biomass combustion. **Pesq Agropec Bras.** 2009;44: 133-140.

Negawo AT, Jorge A, Hanson J, Teshome A, MuktaR MS, Azevedo ALS, et al. Molecular markers as a tool for germplasm acquisition to enhance the genetic diversity of a Napier grass (Cenchrus purpureus syn. Pennisetum purpureum) collection. **Trop. Grassl-Forrajes Trop.** 2018;6, 58-69.

Negawo AT, Teshome A, Kumar A, Hanson J, Jones CS. Opportunities for Napier grass (Pennisetum purpureum) improvement using molecular genetics. **Agronomy** 2017;7, 28.

Pereira JF, Azevedo ALS, Pessoa-Filho M, Romanel EAC, Pereira AV, Vigna BBZ, et al. Research priorities for next-generation breeding of tropical forages in Brazil. **Crop Breed Appl Biotechnol.** 2018;18, 314-319.

Resende MDV, Silva FF, Lopes OS, Azevedo CF. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada e Estatística Espacial.** Viçosa: universidade Federal de Viçosa/ Departamento de Estatística. 291p. 2012.

Rocha JRASC, Machado JC, Carneiro PCS, Carneiro JC, Resende MDV, Ledo FJS, Carneiro JES. Bioenergetic potential and genetic diversity of elephantgrass viamorpho-agronomic and biomass quality traits. **Ind Crops Prod.** 2017;95 :485-492.

Rocha JRASC, Marçal TS, Salvador FV, Silva AC, Machado JC, Carneiro PCS. Genetic insights into elephantgrass persistence for bioenergy purpose. **PLoS ONE.** 2018;13: e0203818.

Shimoya A, Ferreira RP, Pereira AV, Cruz CD, Carneiro PCS. Comportamento morfo-agronômico de genótipos de capim-elefante. **Rev Ceres.** 2001;48: 1-19.

Sykué Bioenergia. **História.** 2019 Avaiable in: <https://sykue.com.br/historia.html> Acessed: 08 mar 2019.

Techio VH, LC Davide, AV Pereira, E. Bearzoti. Cytotaxonomy of some species and of interspecific hybrids of Pennisetum (Poaceae, Poales). **Genet. Mol. Biol.** 2002;2: 203-209.

Varshney RK, Chen W, LI Y, Bharti AK, Saxena RK, Schlueter JA, et al. Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. **Nature Biotechnol.** 2012;30, 83-89.

CHAPTER 1

GENETIC INSIGHTS INTO ELEPHANTGRASS PERSISTENCE FOR BIOENERGY PURPOSE

VIÇOSA
MINAS GERAIS - BRASIL
2019

ABSTRACT

ROCHA, João Romero do Amaral Santos de Carvalho, D.Sc., Universidade Federal de Viçosa, March, 2019. **Genetic insights into elephantgrass persistence for bioenergy purpose.** Adviser: Pedro Crescêncio Souza Carneiro. Co-advisers: Marcos Deon Vilela de Resende and Juarez Campolina Machado.

Persistence may be defined as high sustained yield over multi-harvest. Genetic insights about persistence are essential to ensure the success of breeding programs and any biomass-based project. This chapter focuses on assessing the biomass yield persistence for bioenergy purpose of 100 elephantgrass clones measured in six growth seasons in Brazil. To assess the clones' persistence, an index based on random regression models and genotype-ideotype distance was proposed. Results suggested the existence of wide genetic variability between elephantgrass clones, and that the yield trajectories along the harvests generate genetic insights into elephantgrass clones' persistence and G x E interaction. A gene pool that acts over the biomass yield (regardless of the harvest) was detected, as well as other gene pools, which showed differences on genes expression (these genes are the major responsible for clones' persistence). The lower and higher clones' persistence was discussed based on genome dosage effect and natural biological nitrogen fixation ability applied to bioenergy industry. The huge potential of energy crops necessarily is associated with genetic insights into persistence, so just this way, breeding programs may breed a new cultivar that fulfills the bioenergy industries.

Key words: Cenchrus purpureus, random regression models, multi-harvest, gene pool.

RESUMO

ROCHA, João Romero do Amaral Santos de Carvalho, D.Sc., Universidade Federal de Viçosa, março de 2019. **Compreensões genéticas sobre a persistência de capim-elefante para fins bioenergéticos.** Orientador: Pedro Crescêncio Souza Carneiro. Coorientadores: Marcos Deon Vilela de Resende e Juarez Campolina Machado.

A persistência pode ser definida como alta produtividade sustentada ao longo de múltiplas colheitas. As compreensões genéticas sobre a persistência são essenciais para garantir o sucesso dos programas de melhoramento e de qualquer projeto baseado na biomassa. Este capítulo enfoca na avaliação da persistência da produção de biomassa para fins bioenergéticos considerando 100 clones de capim-elefante avaliados em seis estações de crescimento no Brasil. Para avaliar a persistência dos clones, foi proposto um índice baseado em modelos de regressão aleatória e na distância genótipo-ideótipo. Os resultados sugeriram a existência de ampla variabilidade genética entre os clones de capim-elefante, e as trajetórias de produtividade ao longo das colheitas geraram informações genéticas sobre a persistência dos clones de capim-elefante e sobre interação GxE. Um pool gênico que atua sobre a produção de biomassa (independentemente da colheita) foi detectado, assim como outro pool gênico, que mostrou diferenças na expressão dos genes (esses genes são os principais responsáveis pela persistência dos clones). A alta e baixa persistência dos clones foi discutida com base no efeito da dosagem do genoma e na capacidade natural de fixação biológica de nitrogênio aplicada à indústria de bioenergia. O grande potencial das culturas energéticas está necessariamente associado às compreensões genéticas relacionadas à persistência, dessa forma, os programas de melhoramento podem desenvolver uma nova cultivar que atenda às indústrias de bioenergia.

Palavras chave: Pennisetum purpureum, modelo de regressão aleatória, múltiplas colheitas, pool gênico.

1. Introduction

Elephantgrass [Pennisetum purpureum Schumach.; Syn. Cenchrus purpureus (Schumach.) Morrone] has potential as multi-purpose crop, e.g., bio-based products, co-products and biofuels, besides the use as forage. The nutrient-rich juice can be used as substrate for fungal-protein production (Takara and Khanal 2011) and microbial oil production (Chen et al. 2016). The dry biomass can be used both to produce chemicals composites (Ridzuan et al. 2016a, b; Ituen et al. 2016; Fartini et al. 2016), to be burned in boilers for energy generation (Fontoura et al. 2015; Rocha et al. 2017a, b) or be used for cellulosic ethanol conversion (Yasuda et al. 2014; Scholl et al. 2015).

The use of a given genotype for energy purposes should be mainly based on the knowledge of its calorific value and its yield biomass (Liu et al. 2009). Although raw material quality significantly impacts on bioenergy conversion, the greatest economic driver of raw material production is biomass yield. As biomass yield per unit area increases, transport expenses and demand on arable land decreases, and overall economic returns will increase (Anderson et al. 2016).

Besides high biomass yield, the biomass energy industries desire a persistent (high sustained yield) dedicated energy crop cultivar along the growth seasons. This allows better scaling and scheduling of planting, consequently, better store of the raw material according to the biomass demand and reducing overall costs. Furthermore, due to the perenniality of biomass crops cultivars, they may not be regularly substituted in a plantation (Porter et al., 2007).

Perennials crops must be sufficiently persistent to maintain their yield performance along to the subsequent growth seasons (Wilkins and Humphreys 2003; Conaghan and Casler 2011; Bouton 2012). The persistence could be affected by a large number of factors such as: environment (e.g., disease, temperature, drought, etc.) and management (e.g., harvest and grazing) of the crop (Riday and Brummer 2006). In addition to these factors, the genetic contribution for persistence control may be highlighted.

High biomass yield potential and greater output:input ratio are the breeder's wishes regarding any dedicated energy crop. Therefore, to achieve the plant ideotype, many steps are involved in the plant breeding process. In general, at any breeding program of energy crops, the phenotyping step will occur regarding the target traits, that

is, the data are obtained along the yield trajectory in different growth seasons for the same traits, which are denoted by longitudinal data according to Meyer (2000).

Genetics models such as random regression (RR) models deal with longitudinal data very well (Schaeffer 2004) because they capture the change of a trait continuously over the trajectory with fewer parameters than the multi-trait models (Sun et al. 2017), that is, a parsimonious covariance structure within a continuous scale (infinite dimensional) which provides estimated genetic values at specific times (harvest) or as a trend over time. Besides that, the RR models partition the variance into genetic and permanent environmental effect without assuming constant during the whole evaluated period (Kranis et al. 2007).

Understanding the yield trajectory along the growth seasons/harvests may determine the success of any biomass-project; even generate genetic insights into elephantgrass clones' behavior, which is useful to the breeding programs. To achieve these highlights, we accessed the biomass yield persistence of elephantgrass clones for bioenergy purpose.

2. Materials and methods

2.1. Location and experimental conditions

The experiment was carried out at the experimental field of Embrapa Dairy Cattle Research Center, located in the municipality of Coronel Pacheco, MG, Brazil (lat 21º 33' 18'' S, long 43º 15' 51" W, at 417 m asl), in a red-yellow latosol with the following chemical properties: pH (5.4); H+Al (2.31 cmolc dm^{-3}); P and K (1.1 and 23 mg dm^{-3}, respectively); and the following exchangeable cations: Al^{3+}, Ca^{2+}, and Mg^{2+} (0.2, 1.4, and 0.7 cmolc dm^{-3}, respectively). The planting was carried out in December 2011, in 0.20 m deep furrows, and 80 kg ha^{-1} P_2O_5 fertilizer was applied at planting. After the establishment stage, at 30 days after planting, elephantgrass plots were cut to 0.30 m stubble height (uniformity harvest), beginning the first of the six harvests. Maintenance fertilization was performed with 300 kg ha^{-1} of the N-P_2O_5-K_2O formulation (20:05:20 blended granular fertilizer), after the uniformity harvest and after all harvests. Fertilizers were applied according to the soil analysis.

Six harvests were carried out for this study. Aiming at using them as bioenergetic feedstock, the first (September 28th, 2012) and the second (June 04th, 2013)

harvests were made at 250 growth days; the third harvest (April 15th, 2014), at 315 regrowth days. Nevertheless, at the fourth harvest (January 15th, 2015), at 275 regrowth days, only the propagation material for the network assay of elephantgrass was collected, i.e., no data field information was registered. At 315 regrowth days, the fifth harvest was performed (November 26th, 2015), and the last harvest was carried out at 210 regrowth days (June 22th, 2016). Weather data for the term of the present assay are shown in supplementary Fig S1 and Fig S2.

2.2. Genetic material and experimental design

One hundred genotypes of the Elephantgrass Active Germplasm Bank (BAGCE, supporting information S2 Table) were evaluated. Plots (1.5 m x 4 m) consisted of a single 4 m row. Plots were planted side by side, spaced 1.5 m apart and allocated in a 10 x 10 simple lattice design, with two replications.

2.3. Measurement - Biomass yield

The elephantgrass was harvest and weighted in a 3m section from the middle of the rows to obtain the gross fresh biomass weight per plot. Previously, randomly fresh sub-samples of three complete plants from each plot were harvest and weighted (fresh biomass weight) and oven dried at 56 °C for 72 hours until reaching constant weight (dry biomass weight). After that, the material was ground until passing through a 1 mm mesh. The dry biomass yield was estimated using the fresh and dry biomass weights of the sub-sample fractions and the fresh biomass weight of the gross sample.

2.4. Statistical Analyses

2.4.1. Random regression model

Initially, several random regression models were tested in order to identify the one which best fit the biomass yield trajectory. The following general model was used:

$$y_{ijk} = R_k + \sum_{m=0}^{M_b-1} \beta_m \Phi_{ijm} + \sum_{m=0}^{M-1} \alpha_{im} \Phi_{ijm} + \sum_{m=0}^{M-1} p_{im} \Phi_{ijm} + e_{ijk}$$

The random regression models were fitted on Legendre polynomials of age at measuring (harvest day) for random and fixed effects, considering various orders of fit. y_{ijk} is the i^{th} genotype measured (i = 1, 2, …, ng, where ng is the total number of genotypes) on the j^{th} harvest day (j = 1, 2, …, nh, where nh is the last harvest day) on the k^{th} replication. R_k is the fixed effect of replication (k = 1, 2). β_m is the fixed regression coefficient fitted through quartic degree (order 5 or M_b = 5) Legendre polynomials to the common average trajectory of all genotypes. The random effects: α_{im} and p_{im} are the random regression coefficient for the Legendre polynomial of order m for the genetic effect and permanent environmental effects for ik^{th} plot (ik = 1, 2, …, np, where np is the total number of plots), respectively. Φ_{ijm} is the m^{th} Legendre polynomial for j^{th} harvest day, standard from -1, to +1, from i^{th} clone; M is the fit order, ranged from 1 to 5, of the Legendre's polynomial for the genetic and permanent environmental effects, respectively; e_{ijk} is the residual random effect associated with y_{ijk}. In the matrix notation, the above model is described as follow:

$$y = X\beta + Z\alpha + Wp + e$$

y is the data vector; β is the vector of the effects of the replication (assumed as fixed) added to the overall mean; α a is the vector of genetic effects (assumed as random); p is the vector of the permanent environment (random); e is the vector of residue (random). X, Z, and W represent the incidence matrices for these effects. It was assumed that the fixed part of the model accounts for systematic harvest effect, so that $\alpha \sim N(0, K_g \otimes I_{ng})$, $p \sim N(0, K_p \otimes I_{np})$, α and p are uncorrelated and $e \sim N(0, R)$, where I_{ng} and I_{np} are identity matrices of appropriated size ng and np, respectively. $\otimes$ denotes the direct product, K_g and K_p are covariance coefficient matrix for genetic and permanent environment effect, respectively. R represents a matrix of residual variances. Several models with different residual variances structures (e.g., unstructured, diagonal, and homogeneous) were tested.

2.4.2. Choice of the best fitted model

The maximum degree of the orthogonal polynomials fitted was tested to determine the most appropriate combination. The random regression models were

compared using the following statistical criteria: likelihood ratio test (LRT) (Rao 1973) and the Schwarz's Bayesian information criterion (BIC), $BIC = -2LogL + pLog\left[n - r\left(x\right)\right]$ where: $LogL$ is the logarithm of the likelihood function; p is the number of estimated parameters; n is the number of observations and $r\left(x\right)$ the rank of incidence matrix of fixed effect (Wolfinger 1993). Associated to the BIC, a model will only be chosen if all constraints for variance parameters are positive according to ASReml warning code for parameters (Gilmour et al. 2015).

2.4.3. Extracting the genetics information

2.4.3.1. Variance components

It was used the following estimator according to Kirkpatrick et al. (1990), to obtain the variance components $\left(\hat{V}_x\right)$ on the original scale.

$$\hat{V}_x = \Phi_{ijm}\hat{K}_x\Phi'_{ijm}$$

The term Φ_{ijm} was defined in the section Random regression model; $\hat{K}_x$ is the estimated coefficient covariance matrix for the random effect (genetic or permanent environment).

2.4.3.2. Estimated genetic values

The genetic values $\left(EVG_{ij}\right)$ were estimated as following:

$$EVG_{ij} = \sum_{m=0}^{M} \hat{\alpha}_{im}\Phi_{ijm}$$

Where, $\hat{\alpha}_{im}$ is the random regression coefficient of order m for the genetic effects of i^{th} clone.

2.4.3.3. Accuracy

The accuracy $\left(\hat{r}_{ij}\right)$ was estimated as following:

$$\hat{r}_{ij} = \sqrt{1 - \frac{PEV_{ij}}{\hat{V}_g}}$$

Where, PEV_{ij} is the prediction error variance, obtained by the diagonal elements of the transformed coefficient pertaining to clone i and harvest j of the inverse of the left rand side of mixed models equation; $\hat{V}_g$ is the estimated genetic variance.

2.4.3.4. Eigenfunctions

Additionally, the eigenfunction (Ψ_i) of the genetic coefficient covariance matrix was calculated to provide genetic insights about the studied trait according to Kirkpatrick et al. (1990).

$$\Psi_i = \sum_{m=0}^{M} (c_{\Psi i})_m \Phi_m$$

Where, $(c_{\Psi i})_m$ is the m^{th} element of the i^{th} eigenvector of K_g and Φ_m is the normalized value of the m^{th} Legendre polynomial.

2.4.3.5. Clones' persistence

The clones' persistence ($Persistence_i$) was obtained by the distance from each clone in relation to the ideotype (genotype-ideotype distance), considering all estimated genetic values in the range of 250 to 1615 days. The ideotype - $\max(EVG_j)$ - was defined as the maximum estimated genetic value in each day in the experimental period. The following algorithm was used:

$$Persistence_i = \frac{\dfrac{1}{\sum_{j=250}^{1615} \left[EGV_{ij} - \max(EVG_j) \right]^2}}{\sum_{k=1}^{100} \left\{ \dfrac{1}{\sum_{j=250}^{1615} \left[EGV_{kj} - \max(EVG_j) \right]^2} \right\}} \times 100$$

2.5. Software

The statistical analyses have been performed with the ASReml 4.1 (Gilmour et al. 2015) and R (R Development Core Team 2015) softwares.

3. Results

3.1. The best fitted model and the general genetic behavior

The goodness of fit of the models is presented in Table 1. According to the Schwarz's Bayesian information criterion (BIC) and ASReml warning code for parameters, the best model is denoted by Leg4.1.D with diagonal residual variance and was adopted to describe the changes in the variance and covariance components for elephantgrass biomass yield over multi-harvest. When the models without the genetic or permanent environmental effects were tested by the likelihood ratio test, genetic variability (p-value < 0.01) and significant permanent environmental effect (p-value < 0.01) were detected for all models.

Table 1 Different models fitted with orthogonal Legendre polynomials, number of parameters (p), Schwarz Bayesian information criteria (BIC), and likelihood ratio test (LRT) for genetic and permanent environmental effect.

Model[a]	Fitted order for effect		p	LogL convergence	Constraints[b]	BIC	LRT (Genetic)	LRT (Perm. env.)
	Genetic	Perm. env.						
Leg3.1.H	3	1	8	Converged	P	4679.75	188.66**	20.68**
Leg3.2.H	3	2	10	Converged	P/B	4693.53	148.88**	20.70**
Leg4.1.H	4	1	12	Converged	P/?	4642.74	253.26**	34.26**
Leg4.2.H	4	2	14	Not converged	P/?	-	-	-
Leg3.1.D	3	1	12	Converged	P	4554.36	274.20**	29.88**
Leg3.2.D	3	2	14	Converged	B/P	4550.85	221.22**	47.18**
Leg4.1.D	4	1	16	Converged	P	4549.17	306.98**	32.56**
Leg4.2.D	4	2	18	Converged	B/P	4534.36	265.30**	61.16**
Leg3.1.US	3	1	22	Converged	P	4749.23	185.38**	14.96**
Leg3.2.US	3	2	24	Not converged		-	-	-
Leg4.1.US	4	1	26	Converged	P/?	4714.61	247.58**	24.54**
Leg4.2.US	4	2	28	Converged	P/?	4728.15	215.34**	24.80**

[a]The models tested are referred to as Legm$_a$.m$_p$..x, where m$_a$ and m$_p$ represent the Legendre's polynomials orders adjusted for genetic and permanent environmental random effects, respectively, and x may assume homogeneous (H), diagonal (D) or unstructured (US) residual variance structure.

[b]ASReml warning code for parameters, P - positive definite, B - fixed at a boundary, ? - liable to change from P to B.

[**]significant at 1% by the chi-squared test.

Chi-squared tabulated: 6.63 for 1% significance level.

Fig 1 shows the general shape of the biomass yield trajectory over the harvests and all random genetic curves. The graph indicates the wide variability that exists around the average curve. Thus, elephantgrass has different biomass yield curves.

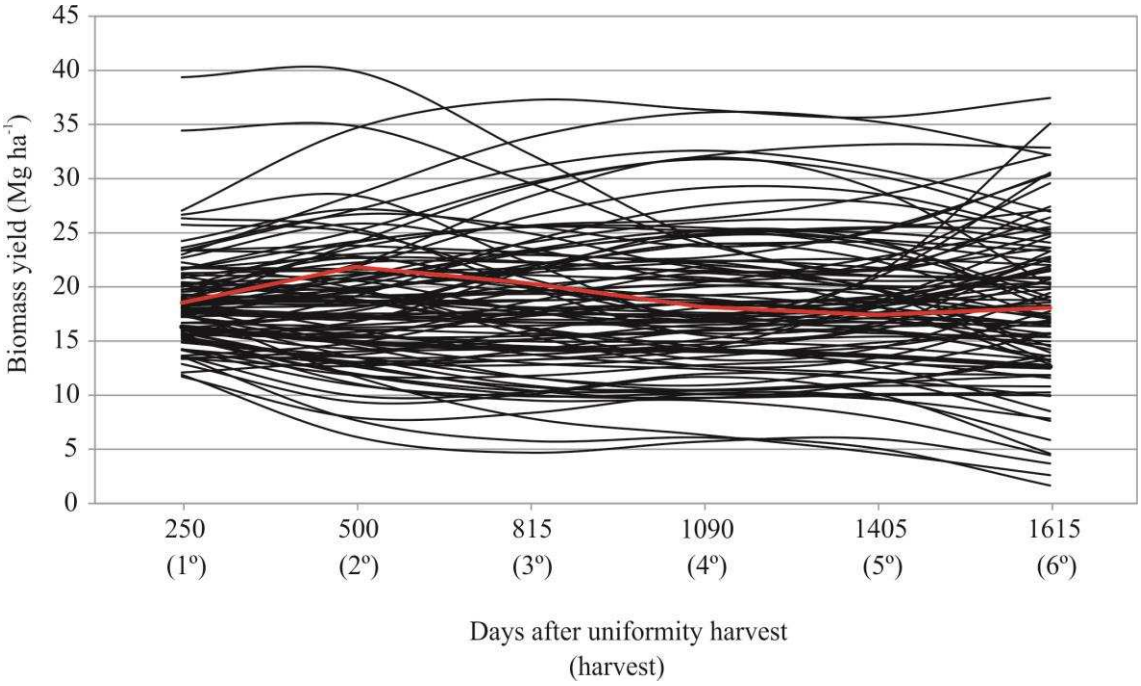

Fig 1. Estimated genetic values for biomass yield over multi-harvest for 100 elephantgrass clones. Each black line represents one clone, and the red line represents the average biomass yield curve.

3.2. Heritability, genetic variance, phenotypic variance, and permanent variance trajectory for biomass yield over the multi-harvest

Fig 2 shows that the phenotypic variance trajectory was not stable over the multi-harvest. The phenotypic variance reached the peak in the sixth harvest, i.e., the greatest phenotypic variability occurred in the sixth harvest followed by the second one. A constant trajectory was observed for permanent environment variance.

The genetic variance trajectory was increasing over the multi-harvest, with a slight decrease around the fifth harvest (Fig 2). The heritability estimates ranged from 0.45 to 0.75. In general, heritability values decrease from the first to the second harvest (indicating that the second harvest is strongly influenced by the environment) and remain above 0.69 from the third harvest onwards (Fig 2). In addition, the genetic values for the fourth harvest (1090 days, without phenotypic data) were predicted with 91.65% average accuracy.

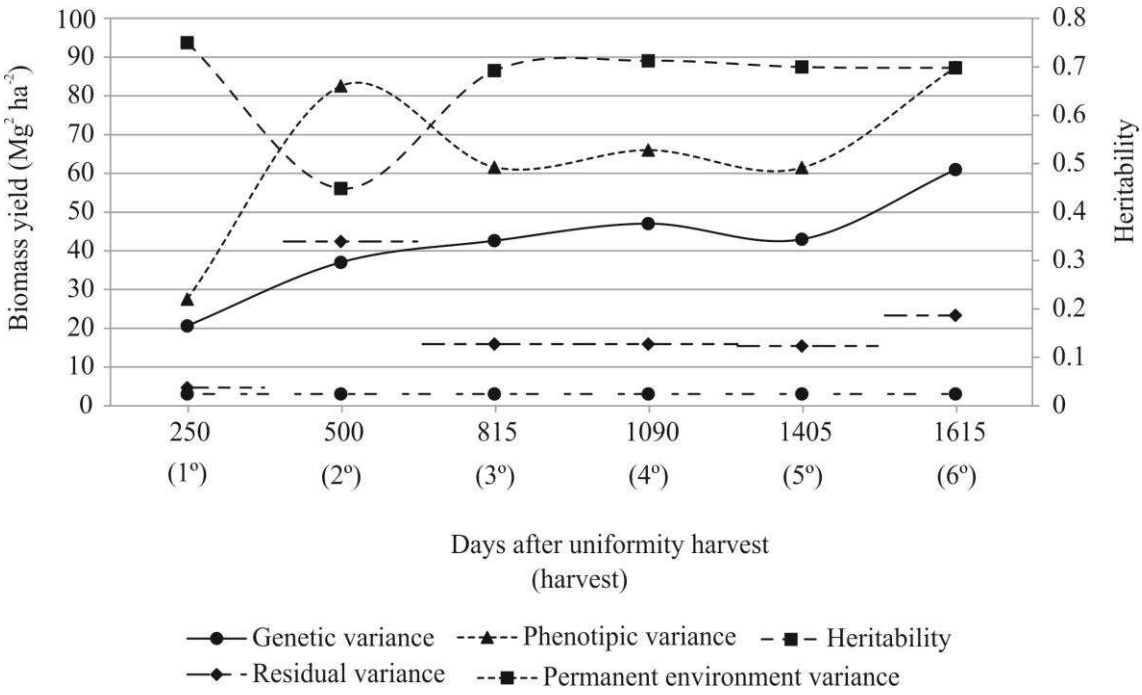

Fig 2. Heritability, genetic and permanent environmental variance, and phenotypic variance trajectory.

All trajectories were estimated from the random regression model (Leg4.1.D) fitted by Legendre polynomials for biomass yield over multi-harvest. See Supplementary information Fig 2 for accumulated rainfall and temperature data.

Fig 3 reveals that the first eigenfunction (associated with the largest eigenvalue) is nearly constant over the multi-harvest, indicating that 86% of the genetic variation is

explained by a gene pool that acts over the biomass yield, regardless of the growth season.

The second eigenfunction (around 11% of the genetic variation, Fig 3) represents other gene pool that shows differences in genes expression under different environment conditions, i.e., biomass yield reverses between the first two harvests (250 and 500 days) in relation to the third, fourth, and fifth harvests (815, 1090, and 1405 days) and reverses again in relation to the last one (1615 days). This genetic factor is the major responsible for the genotypes by environments (growth seasons) interaction. The third and fourth eigenfunction was not interpreted due to their small genetic variation proportion (3.31 and 0.26%, Fig 3).

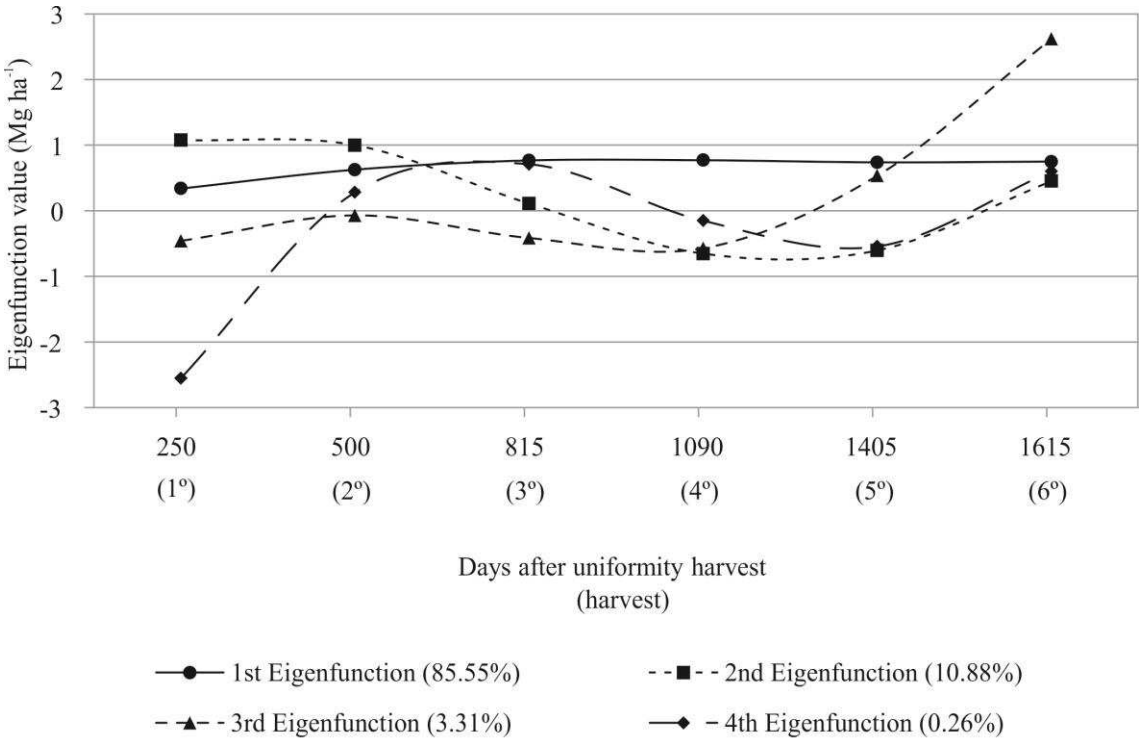

Fig 3. Estimates of the four eigenfunctions. Their proportional eigenvalues for the genetic covariance function are in parentheses**.**

3.3. Elephantgrass clones' persistence

The experimental biomass yield means were 12.51, 29.60, 19.60, 15.28, and 19.07. Mg ha^{-1} for 250, 500, 815, 1405 and 1615 days after uniformity harvest, respectively. Fig 4 shows the ten most and the five least persistent elephantgrass clones.

Clone 2 was not the most yielded at all harvests; however, it sustained the biomass yield with the highest performance from the third harvest (815 days) onwards.

Among the ten most persistent clones, clones 95 and 55 showed the highest biomass yield in the first two harvests, i.e., these clones achieved the biomass yield peak quickly, but were not able to persist over the multi-harvest (Fig 4). Clones 49, 100 and 99 showed the lowest persistence (Fig 4). Clones 99 and 100 were the only ones that completely died in the plots of the first replication in the fifth harvest (1405 days) and in all plots in the sixth harvest (1615 days).

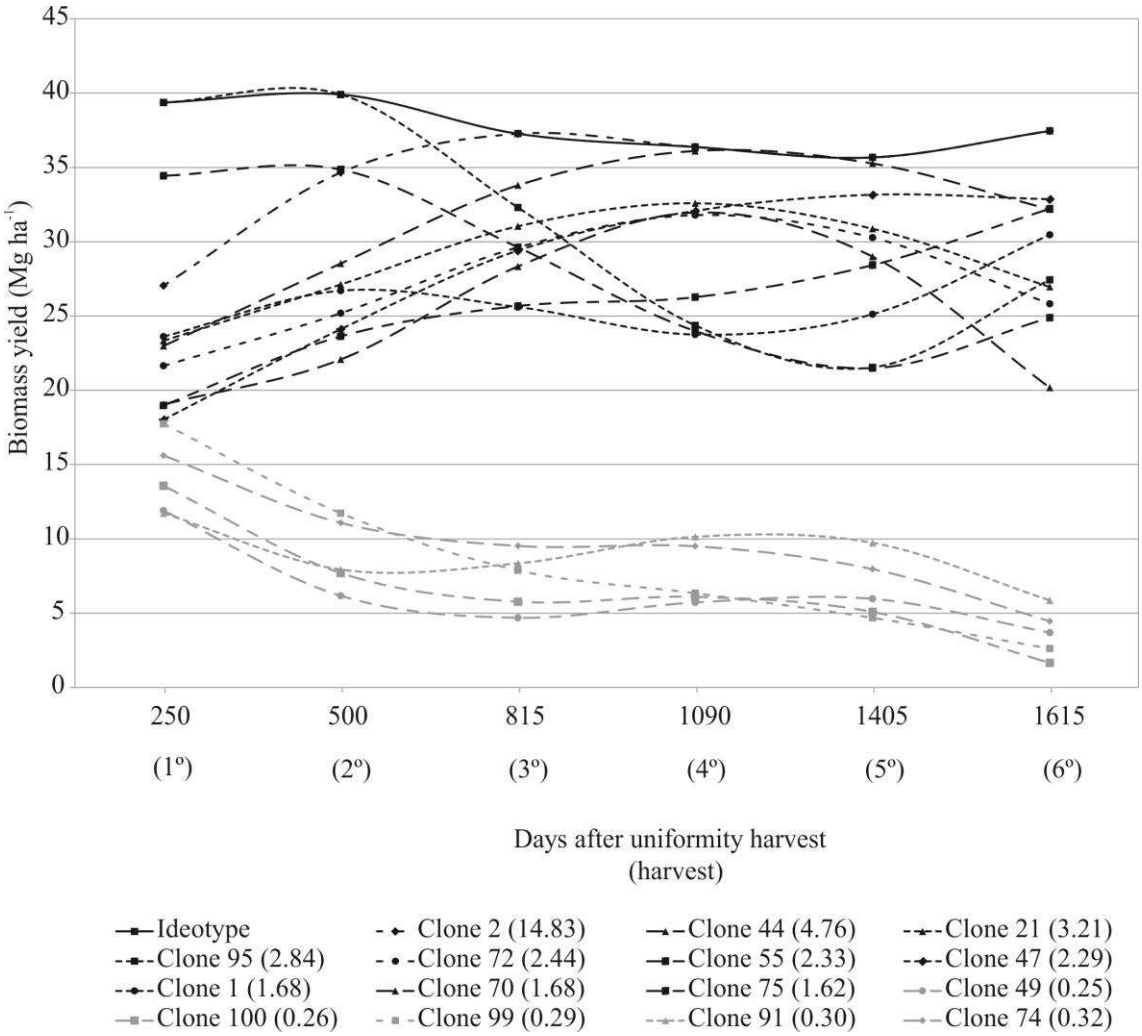

Fig 4. Elephantgrass biomass yield trajectory over the multi-harvest. The ten most persistent and the five least persistent clones. Persistence values are in parentheses.

4. Discussions

4.1. The best-fitted model

Statistical methods for analyzing data of perennials need to appropriately model the genetic effects over time (Faveri et al. 2015). A suitable method for modeling the genetic trajectory over time is the random regression model, as commonly used in the animal sciences (Kranis et al. 2007; Sun et al. 2017). It is noteworthy that the experimental data used in this work shows unequally spaced sample intervals over the harvests (250, 500, 815, 1090, 1405, 1615 days after the uniformity harvest) and that no phenotypic data was collected in the fourth harvest (at 1090 days after the uniformity harvest). Kirkpatrick et al. (1990) relate that, under this condition, the random regression models are sufficient, being the adequate methodology.

Different criteria have been used to find the polynomial order of the model with the best fit and parsimony (Corrales et al. 2015). In the present study, the best-fitted model was indicated by the BIC criterion and ASReml warning code for parameters (Legendre polynomials of the fourth order, for genetic effect, and first order, for permanent environmental effect with diagonal residual variance structure – Leg4.1.D, table 1). From this model, all interpretations were performed.

Simple repeatability and multi-trait models can be employed for longitudinal data analysis (Sun et al. 2017). However, the repeatability model considers the genetic correlation between the different harvests equal to one, i.e., exactly the same genes acting in the control of the trait over time (Mrode 2014). The repeatability model can be reproduced from the random regression model, considering the fitted order equal to one (intercept random regression model). Thus, the repeatability model is not suitable to represent the genetic behavior presented by the biomass yield over the multi-harvest. The multi-trait model assumes that the data are discontinuous, while in the present work the harvests are taken continuously. Therefore, the extrapolation of the genetic value to an unobserved harvest is not recommended. The multi-trait model would need to estimate 30 parameters, while the chosen random regression model (Leg4.1.D) required only 16 parameters, thus being parsimonious, in addition, the multi-trait model does not allow computing the random permanent environment effects (i.e., the permanent environment effect is confused or overlaid with the temporary environment effect).

4.2. Genetic variability and general genetic behavior

The high amplitude that the random genetic curves deviations showed in relation to general biomass yield trajectory suggests high genetic variability (Fig 1). Wide genetic variability in the studied clones was expected since they are clones belonging to a germoplasm bank (BAGCE) and have not yet been genetically improved for bioenergetics purposes (Rocha et al. 2017a). According to Azevedo et al. (2012) and Rocha et al. (2017b), these clones presented genetic variability using single sequence repeats markers and the biomass yield trait, respectively. Knowledge about genetic diversity is the key to further improvement, and evaluation of diversity in germplasm is essential for the effective use of genetic resources in breeding programs. Assessing the diversity information would facilitate the progress in plant breeding (Nielsen et al. 2014).

4.3. Permanent environment, plasticity, and persistence

Besides the genetic and the temporary environmental effect that composes a phenotype, a permanent environmental effect occurs in longitudinal data. Permanent implies stability and a constant or common presence to repeated measures (Kruuk and Hadfield 2007; Schaeffer 2011). Kruuk and Hadfield (2007) have shown that permanent environmental effect may overlap with several factors, e.g., dominance or epistatic genetic effects, maternal genetic effects, common environmental effects, especially the phenotypic plasticity. It is worth mentioning that the permanent environmental effect is estimated by the variance among repeated measures in the same individual (i.e., in animal science no replications of the individual is used in the experiment as in plant science). Replications of individuals or families are very common in plant breeding experiments, and the permanent environmental effect is estimated by the variance among repeated measures in the plots (different genotypes-replications combination). Under this condition, additional effects occur due to differential competition between the same individual in different plots (due to experimental randomization).

The second harvest showed the greatest phenotypic variance and the greatest temporary environment or residual variance (Fig 2); this behavior may be explained by the rainfall and temperature data (Fig S2). The second harvest showed the most favorable environmental condition (e.g., temperature and rainfall, see Fig S2) for elephantgrass growth, which is confirmed by the highest biomass mean, 29.60 Mg ha^{-1} (51% more productive than the second most productive harvest). Thus, resources

availability (e.g., light, water, nutrient, temperature, etc.) can stimulate phenotypic plastic response subject to generate a large phenotypic variance.

According to Nicotra et al. (2010), phenotypic plasticity is the range of phenotypes that a single genotype can express as a function of its environment. Phenotypic plasticity depends on the genome plasticity, defined as a change in the genome structure or organization associated with environmental signals (Nicotra et al. 2010). In this context, the phenotypic plasticity can be considered as favorable or unfavorable changes for genotype adaptedness (van Eeuwijk et al. 2016). Nicotra et al. (2010) define adaptive plasticity as the phenotypic plasticity that increases the global fitness of a genotype. In the plant breeding context, the adaptive plasticity is equivalent to the adaptability proposed by Finlay and Wilkinson (1963).

Bradshaw (2006) relates that plasticity is related to stability. Plasticity can be a simple sign of weakness (of lack of fitness), but it can also be a sign of strength, attributed to maintenance mechanisms of fitness. Breeders attempt to select genotypes with consistent performance between a range of target environments in order to reduce the G x E effects (Gage et al. 2017). For instance, breeders try to produce cultivars that reliably perform despite year-to-year fluctuations in weather patterns. In the case where limited phenotypic plastic response confers stability, the low G × E contribution may have a desirable effect by enabling germplasm to predictably perform across environments (Gage et al. 2017). Plasticity not only gives an edge over competitors but also is essential for genotype persistence in new or changing environments (Morris 2014). In the plant breeding scenario, persistence was defined as high sustained yield over environmental changes (Wilkins and Humphreys 2003; Conaghan and Casler 2011; Bouton 2012) and is a relevant trait under several aspects of the bioenergy industry.

4.4. Insights about G x E interaction - driving the selection

When trajectory curves are non-constant, genotypes show plasticity (in new and changing environments – growth seasons), and when the curves intersect, a G x E interaction occurs (see Fig 4). According to van Eeuwijk et al. (2016), this type of G x E has more severe consequences for breeders as it will change the rank order of clones in function of the environmental conditions.

The main reason for the seasonality of elephantgrass yield over multi-harvest is the differential genes expression, i.e., the environmental effect promotes different levels of genes expression (even the nonexpression of the genes) that affect the elephantgrass biomass yield. The differential genes expression is the theoretical base of G x E interaction. The first eigenfunction (Fig 3) captured a gene pool that was equally expressed in all growth seasons (e.g., general adaptability genes). The second eigenfunction (Fig 3) clustered genes that expressed themselves depending on the environmental differences (these genes determine the persistence - specific adaptability genes). The uninterpreted eigenfunctions showed small eigenvalues (Fig 3), and according to Kirkpatrick et al. (1990), eigenfunctions with very small (or null) eigenvalues represent deformations for which there is little (or none) genetic variation.

The second harvest is not recommended for selection ($h^2 = 0.45$, Fig 2). Results also indicated that the first, third, fourth, fifth, and sixth harvests represent a more favorable scenario for selection (i.e., accuracy higher than 90%). The genetic breeding must handle inheritable traits, i.e., those with high heritability. The heritability of a trait will have an impact on selection decisions. Genetic progress tends to be much slower in lowly heritable traits. Conversely, with higher heritability, a faster progress is achieved with selection due to greater accuracy in selection decisions (Bullock 2014).

Forage breeding can be a complex task due to the plant perenniality, among several other factors (Sokolovic´ et al. 2011). Persistence is a complex trait affected by a large number of biotic, abiotic, and genetic factors, e.g., diseases and pests, mechanical harvesting equipment, intensity of harvest management, temperature, drought, plant competition (Riday and Brummer 2006) and genome plasticity. Thus, pyramiding of genes that express themselves in different environments would increase genome plasticity and consequently increase the genotypes' persistence.

4.5. Lowest persistence - supported by the genome dosage

The death of clones 99 and 100 in the last two harvests is a factor that explains the lowest persistence. Death may have been caused by the low perenniality of these clones, i.e., the genome dosage may interfere with the perenniality of elephantgrass clones.

Elephantgrass is allotetraploid (2n = 4x = 28, A'A'BB) with ploidy level variations (Martel et al. 1997). Pearl millet [Pennisetum Glaucum (L.) R. Br.; Syn.

Cenchrus americanus (L.) Morrone, 2n = 2x = 14, AA] has an annual growth habit and can produce interspecific hybrids with elephantgrass (Pantulu and Krishna 1982;Martel et al. 1997). For instance, clone 99 is triploid (2n = 3x = 21, AA'B), and clone 100 is hexaploid (2n = 6x = 42, AAA'A'BB). Triploids have an additional copy of A pearl millet genome, whereas the hexaploids have two copies of A pearl millet genome. The A' genome chromosomes are larger than the B genome chromosomes. Moreover, the B genome contributes to elephantgrass perennial life cycle (Anderson et al. 2008). However, the additional genome dose of the pearl millet (A genome in Clones 99 and 100) may reduce the perenniality due to the annual growth habit genes present in the A genome.

Clone 46 (Kizozi) also showed low persistence. Kizozi was previously studied by Techio et al. (2002) and confirmed as a wild species of the genus Cenchrus due to the somatic chromosome number (2n=54). Wild species could have less adaptive genes when compared with breeding cultivars, which leads them to low persistence. Some of the least persistent clones are tetraploid (e.g., clones 30, 34, 35, 6, 77, 74, 91, and 49). The low persistence presented by clone 49 (Mott) is due to its reduced plant height (dwarfing genes). Mott is specifically adapted to be used as forage, in the pasture, owing to its high nutritive value, and it has previously been identified as of low potential for bioenergy production (Rocha et al. 2017b). The other clones showed reduced biomass yield, low height, thin stalks (Rocha et al. 2017a), and consequently low persistence. These clones were classified as Napier, Mercker, or intermediate (Napier/Mercker) group (Lira e al. 2010), and were studied by Rocha et al. (2017b) regarding the aptitude for bioenergy purposes. The authors showed that Napier, Mercker, and intermediate groups have a low potential for biomass energy production.

Furthermore, biomass crops should be perennial because a cultivar cannot be regularly substituted in a plantation. Perenniality, unlike annual habit, would be advantageous due to costs reduction with the establishment of energy crops (Conaghan and Casler 2011).

4.6. Clones' persistence applied to bioenergy industries

The aim of supplying biomass to bioenergy is to achieve high energy yields per unit area and the best possible fuel quality. The energy yield comprises the biomass yield and the energy content of the biomass. Fuel quality is determined by the physical

and chemical properties and influences the entire process of thermal utilization (Prochnow et al. 2009). However, the greatest economic driver of raw material production is biomass yield (Anderson et al. 2016). Besides high energy yields per area, biomass energy industries look for a cultivar with higher persistence (high sustained yield). This fact allows reducing costs due to the better scaling and scheduling of planting, harvesting, and storing of the raw material, based on the biomass demand.

Persistence is an economically important trait for perennial forages due to the costs involved in sward establishment. This trait is dependent on the vigor of a plant and its ability to survive and contribute to yield and ground cover (Conaghan and Casler 2011), and thus, the clumps expansion capacity and the number of basal and axillary tillers may directly impact on persistence. The persistence of elephantgrass clones was measured using clones-ideotype distance over multi-harvest. This approach takes into account the yield stability and the high genetic values.

Clone 2 showed the highest sustained biomass yield over multi-harvest (highest persistence), which may be supported by the natural biological nitrogen fixation (BNF) ability of this clone, as reported by Morais et al. (2012). The higher natural nitrogen input (e.g., BFN) is related to higher biomass production and competitive advantages, especially under unfavorable environmental conditions. Furthermore, to make the input:output rate more favorable to energy balance, industrial nitrogen inputs (N fertilizers) must be minimized. Low nitrogen requirement is desirable not only for being a valued constituent in terms of conversion to energy but also for N fertilizer is a costly input (Na et al. 2016).

Understanding the yield trajectories patterns of elephantgrass clones allowed detecting the G x E interaction and assess the persistence of these plants. The eigenfunctions indicate valuable insights about the G x E interaction, i.e., there is a gene pool (general adaptability genes) that is expressed in all growth seasons and another gene pool (specific adaptability genes or persistence genes) that is expressed under different environmental conditions. These findings suggest that increase the elephantgrass persistence can be successfully achieved with breeding efforts, as a consequence of wide genetic variability for biomass yield and high heritability values over harvests. Moreover, the random regression model allows optimizing elephantgrass management techniques, as well as developing strategies for crosses (i.e., explore the genetic variability).

Future persistence studies applied to elephantgrass should integrate molecular markers information, besides the phenotypic data, aiming at finding several stable quantitative trait loci (QTLs) across multi-harvest. However, the instable QTLs detected may contribute to the persistence increment by recombining (genes pyramiding) unstable QTLs. QTLs studies would allow identifying the genetic basis of G x E in the form of QTLs x E interaction (van Eeuwijk et al. 2016). In addition, further studies on natural biological fixation nitrogen vs. persistence specifically designed for bioenergy purposes must be developed, mainly to keep a favorable energy balance and reduce costs with N fertilizer input. In this way, breeders will be able to rationally deal with the factors that determine persistence by using these factors in their favor. Moreover, they will breed cultivars that will be adopted in bioenergy industries, mainly due to the increase in persistence.

5. Conclusions

There is wide genetic variability for biomass yield between elephantgrass clones.

The yield trajectories along the harvests generate genetic insights into elephantgrass clones' persistence and G x E interaction.

A gene pool that acts over the biomass yield, regardless of the harvest, was detected, as well as other gene pools, which show differences on genes expression, these genes are the major responsible for clones' persistence.

The elephantgrass persistence can be successfully increased the achieved by means of breeding efforts.

6. References

Anderson WF, Casler MD, Baldwin BS. Improvement of perennial forage species as feedstock for bioenergy. In: Vermerris W, editor. **Genetic Improvement of Bioenergy Crops.** New York: Springer Science + Business Media; 2008. pp. 347-376.

Anderson WF, Sarath G, Edme S, Casler MD, Mitchell RB, Tobias CM, Hale AL, Sattler SE, Knoll JE. Dedicated Herbaceous Biomass Feedstock Genetics and Development. **Bioenergy Res.** 2016;9: 399-411. doi:10.1007/s12155-015-9709-8

Azevedo ALS, Costa PP, Machado JC, Machado MA, Pereira AV, Lédo FJS. Cross-species amplification of Pennisetum glaucum microsatellite markers in Pennisetum purpureum and genetic diversity of Napier grass accessions. **Crop Sci.** 2012;52: 1776-1785. doi: 10.2135/cropsci2011.09.0480

Bouton J. Breeding lucerne for persistence. **Crop Pasture Sci.** 2012;63: 95-106. doi:10.1071/CP12009

Bradshaw AD. Unravelling phenotypic plasticity – why should we bother? **New Phytol.** 2006;170: 644-648. doi:10.1111/j.1469-8137.2006.01761.x

Bullock KD. **Genetic Principles.** In: Garrick D, Bullock D, Enns RM, Imumorin I, Mistzal I, Kappes S, editors. Beef Sire Selection Manual, 2nd ed. National Beef Cattle Evaluation Consortium; 2014. Pp. 14-16.

Chen XF, Huang C, Xiong L, Wang B, Qi GX, Lin XQ, Wang C, Chen XD. Use of elephant grass (Pennisetum purpureum) acid hydrolysate for microbial oil production by Trichosporon cutaneum. **Prep Biochem Biotechnol** 2016;46: 704-708. doi: 10.1080/10826068.2015.1135453

Conaghan P, Casler MD. A theoretical and practical analysis of the optimum breeding system for perennial ryegrass. **Ir J Agric Food Res.** 2011;50: 47-63.

Corrales, JD, Munilla S, Cantet RJC. Polynomial order selection in random regression models via penalizing adaptively the likelihood. **J Anim Breed Genet.** 2015;132: 281-288. doi:10.1111/jbg.12130

Fartini MS, Majid MSA, Ridzuan MJM, Amin NAM, Gibson AG. Compressive properties of Napier (Pennisetum Purpureum) filled polyester composites. **Plast. Rubber Compos.** 2016;45: 136-146. doi: 10.1080/14658011.2016.1149911

Faveri J, Verbyla AP, Pitchford WS, Venkatanagappa S, Cullis BR. Statistical methods for analysis of multi-harvest data from perennial pasture variety selection trials. **Crop and pasture science.** 2015;66: 947-962. doi: 10.1071/CP14312

Finlay KW, Wilkinson GN. The analysis of adaptation in a plant-breeding programme. **Aust. J. Agric. Res.** 1963;14: 742–754.

Fontoura CF, Brandão LE, Gomes LL. Elephant grass biorefineries: towards a cleaner Brazilian energy matrix? **J Clean Prod.** 2015;96: 85-93. doi: 10.1016/j.jclepro.2014.02.062

Gage JL, Jarquin D, Romay C, Lorenz A, Buckler ES, Kaeppler S, et al. The effect of artificial selection on phenotypic plasticity in maize. **Nat. Commun.** 2017;8: 1348. doi:10.1038/s41467-017-01450-2

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R. **ASReml User Guide Release 4.1 Structural Specification**, VSN International Ltd, Hemel Hempstead, UK; 2015.

Ituen E, James A, Akaranta O, Sun S. Eco-friendly corrosion inhibitor from Pennisetum purpureum biomass and synergistic intensifiers for mild steel. **Chin. J.Chem. Eng**. 2016;24: 1442-1447. doi: 10.1016/j.cjche.2016.04.02

Kirkpatrick M, Lofsvold D, Bulmer M. Analysis of the inheritance, selection and evolution of growth trajectories. **Genetics.** 1990;124: 979-993.

Kranis A, Su G, Sorensen D, Woolliams JA. The application of random regression models in the genetic analysis of monthly egg production in turkeys and a comparison with alternative longitudinal models. **Poultry Science.** 2007;86: 470-475. doi:10.1093/ps/86.3.470

Kruuk LEB, Hadfield JD. How to separate genetic and environmental causes of similarity between relatives. **J Evol Biol.** 2007;20: 1890-1903. doi: 10.1111/j.1420-9101.2007.01377.x

Lira MA, Cunha MV, Pereira AV. **Melhoramento genético do capim-elefante.** In: Lira MA, Dubeux Júnior JCB, Mello ACL, editors. Capim-elefante: Fundamentos e Perspectivas. Recife, IPA/UFRPE; 2010. pp. 31-48.

Liu X, Shen Y, Lou L, Ding C, Cai Q. Copper tolerance of the biomass crops elephant grass (Pennisetum purpureum Schumach), vetiver grass (Vetiveria zizanioides) and the upland reed (Phragmites australis) in soil culture. **Biotechnol Adv.** 2009;27: 633-640. doi: 10.1016/j.biotechadv.2009.04.017

Martel E, De Nay D, Siljak-Yakovlev S, Brown S, Sarr A. Genome size variation and basic chromosome number in pearl millet and fourteen related Pennisetum species. **J. Hered.** 1997;88: 139-143. doi: 10.1093/oxfordjournals.jhered.a023072

Meyer K. Random regressions to model phenotypic variation in monthly weights of Australian beef cows. **Livest Prod Sci.** 2000;65: 19-38. doi: 10.1016/S0301-6226(99)00183-9

Morais RF, Quesada DM, Reis VM, Urquiaga S, Alves BJR, Boddey RM. Contribution of biological nitrogen fixation to Elephant grass (Pennisetum purpureum Schum.). **Plant Soil.** 2012;356: 23-34. doi.org/10.1007/s11104-011-0944-2

Morris MRJ. Plasticity-mediated persistence in new and changing environments. **Int J Evol Biol**. 2014;2014: 1-18. doi:10.1155/2014/416497

Mrode RA. **Linear models for the prediction of animal breeding values**. MA :CABI, Boston; 2014.

Na C, Sollenberger LE, Fedenko JR, Erickson JE, Woodard KR. Seasonal changes in chemical composition and leaf proportion of elephantgrass and energycane biomass. **Ind Crops Prod.** 2016;94: 107-116. doi; 10.1016/j.indcrop.2016.07.009

Nicotra AB, Atkin OK, Bonser SP, Davidson AM, Finnegan EJ, Mathesius U, Poot P, Purugganan MD, Richards CL, Valladares F, van Kleunen M. Plant phenotypic plasticity in a changing climate. **Trends Plant Sci.** 2010;15: 684–692. doi:10.1016/j.tplants.2010.09.008

Nielsen NH, Backes G, Stougaard J, Andersen SU, Jahoor A. Genetic Diversity and Population Structure Analysis of European Hexaploid Bread Wheat (Triticum aestivum L.) Varieties. **PLoS ONE.** 2014;9: e94000. doi:10.1371/journal.pone.0094000

Pantulu JV, Krishna MR. Cytogenetics of pearl millet. **Theor Appl Genet**. 1982;61: 1-17. doi: 10.1007/BF00261503

Porter JR, Kirsch MMN, Streibig J, Felby C. Choosing crops as energy feedstocks. **Nat Biotechnol.** 2007;25: 716-717. doi:10.1038/nbt0707-716

Prochnow A, Heiermann M, Plöchl M, Amon T, Hobbs PJ. Bioenergy from permanent grassland – a review: 2. Combustion. **Bioresour. Technol.** 2009;100: 4945-4954. doi:10.1016/j.biortech.2009.05.069

R Development Core Team. R: **A language and environment for statistical computing. R Foundation for Statistical Computing.** Vienna, Austria; 2015. Available from: http://www.R-project.org/.

Rao CR. **Linear Statistical Inference and its Applications.** John Wiley & Sons; 1973.

Riday H, Brummer EC. Persistence and Yield Stability of Intersubspecific Alfalfa Hybrids. **Crop Sci.** 2006;46: 1058-1063. doi:10.2135/cropsci2005.0272

Ridzuan MJM, Majid MSA, Afendi M, Kanafiah SNA, Zahri JM, Gibson AG. Characterisation of natural cellulosic fibre from Pennisetum purpureum stem as potential reinforcement of polymer composites. **Mater. Des.** 2016a;89: 839-847. doi:10.1016/j.matdes.2015.10.052

Ridzuan MJM, Majid MSA, Afendi M, Mazlee MN, Gibson AG. Thermal behaviour and dynamic mechanical analysis of Pennisetum purpureum/glass-reinforced epoxy hybrid composites. **Compos. Struct.** 2016b;52: 850-859. doi: 10.1016/j.compstruct.2016.06.026

Rocha JRASC, Machado JC, Carneiro PCS, Carneiro JC, Resende MDV, Ledo FJS, Carneiro JES. Bioenergetic potential and genetic diversity of elephantgrass viamorpho-agronomic and biomass quality traits. **Ind Crops Prod.** 2017a;95: 485-492. doi: 10.1016/j.indcrop.2016.10.060

Rocha JRASC, Machado JC, Carneiro PCS, Carneiro JC, Resende MDV, Pereira AV, Carneiro JES. Elephant grass ecotypes for bioenergy production via direct combustion of biomass. **Ind Crops Prod.** 2017b;95: 27-32. doi: 10.1016/j.indcrop.2016.10.014

Schaeffer LR. Application of random regression models in animal breeding. **Livest Prod Sci.** 2004;86: 35-45. doi: 10.1016/S0301-6226(03)00151-9

Schaeffer, LR. Cumulative permanent environmental effects for repeated records animal models. **J Anim Breed Genet.** 2011;128: 95-99. doi:10.1111/j.1439-0388.2010.00894.x

Scholl AL, Menegol D, Pitarelo AP, Fontana RC, Zandoná Filho A, Ramos, LP, Dillona AJP, Camassola M. Ethanol production from sugars obtained during enzymatic hydrolysis of elephant grass (Pennisetum purpureum, Schum.) pretreated by steam explosion. **Bioresour. Technol.** 2015;192: 228-237. doi: 10.1016/j.biortech.2015.05.065

Sokolovic´ D, Radovic´ J, Tomic´ Z. Perennial forage grasses, from breeding to healthy ruminant feed. **Biotechnology Anim. Husbandry.** 2011;27: 599-614. doi:10.2298/BAH1103599S

Sun J, Rutkoski JE, Poland JA, Crossa J, Jannink JL, Sorrells ME. Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. **Plant genome** 2017;2: 1-12. doi: 10.3835/plantgenome2016.11.0111

Takara D, Khanal SK. Green processing of tropical banagrass into biofuel and biobased products: An innovative biorefinery approach. **Bioresour. Technol.** 2011;102: 1587-1592. doi:10.1016/j.biortech.2010.08.106

Techio VH, Davide LC, Pereira AV, Bearzoti E. Cytotaxonomy of some species and of interspecific hybrids of Pennisetum (Poaceae, Poales). **Genet. Mol. Biol.** 2002;25: 203-209. doi:10.1590/S1415-47572002000200014

van Eeuwijk FA, Bustos-Korts DV, Malosetti M. What should students in plant breeding know about the statistical aspects of genotype x environment interactions? **Crop Sci.** 2016;56: 2119-2140. doi: 10.2135/cropsci2015.06.0375

Wilkins PW, Humphreys MO. Progress in breeding perennial forage grasses for temperate agriculture. **J Agric Sci.** 2003;140: 129-150. doi:10.1017/S0021859603003058

Wolfinger, R.D. Covariance structure in general mixed models. **Comm Stat Simul Comput.** 1993;22B: 1079-1106. doi:10.1080/03610919308813143

Yasuda M, Ishii Y, Ohta K. Napier grass (Pennisetum purpureum Schumach) as raw material for bioethanol production: pretreatment, saccharification, and fermentation. **Biotechnol Bioprocess Eng.** 2014;19: 943-950. doi:10.1007/s12257-014-0465-y

7. Supporting Information

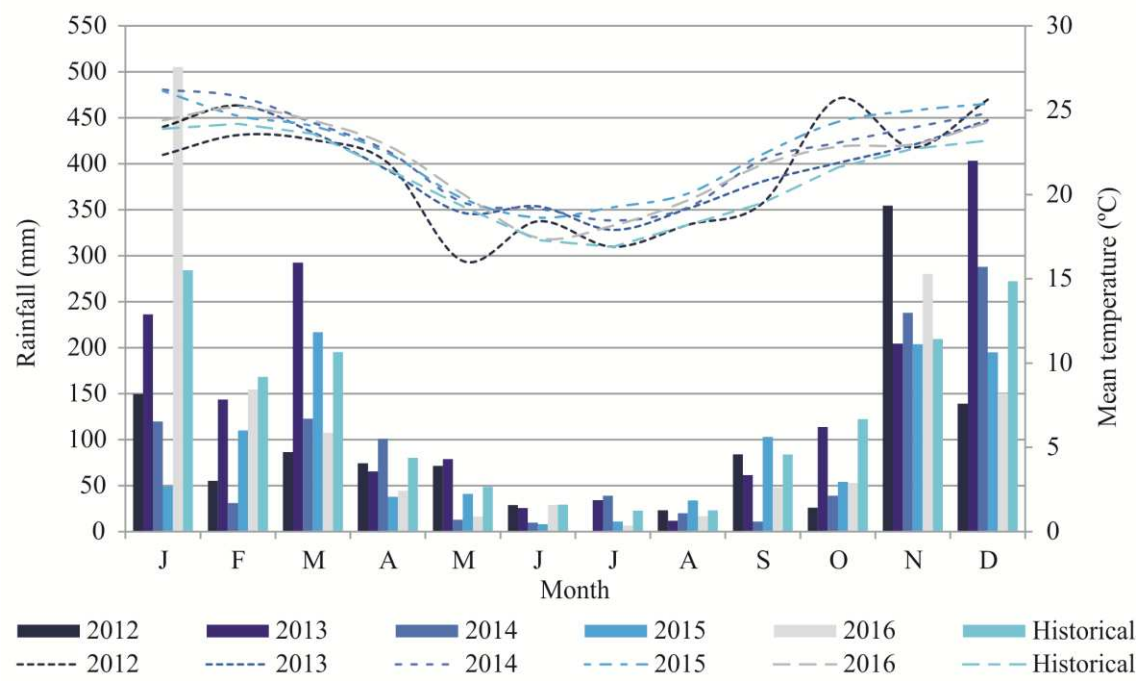

Fig S1. Rainfall (bars) and average temperature (lines) during the current assay and historical data for 30 years.

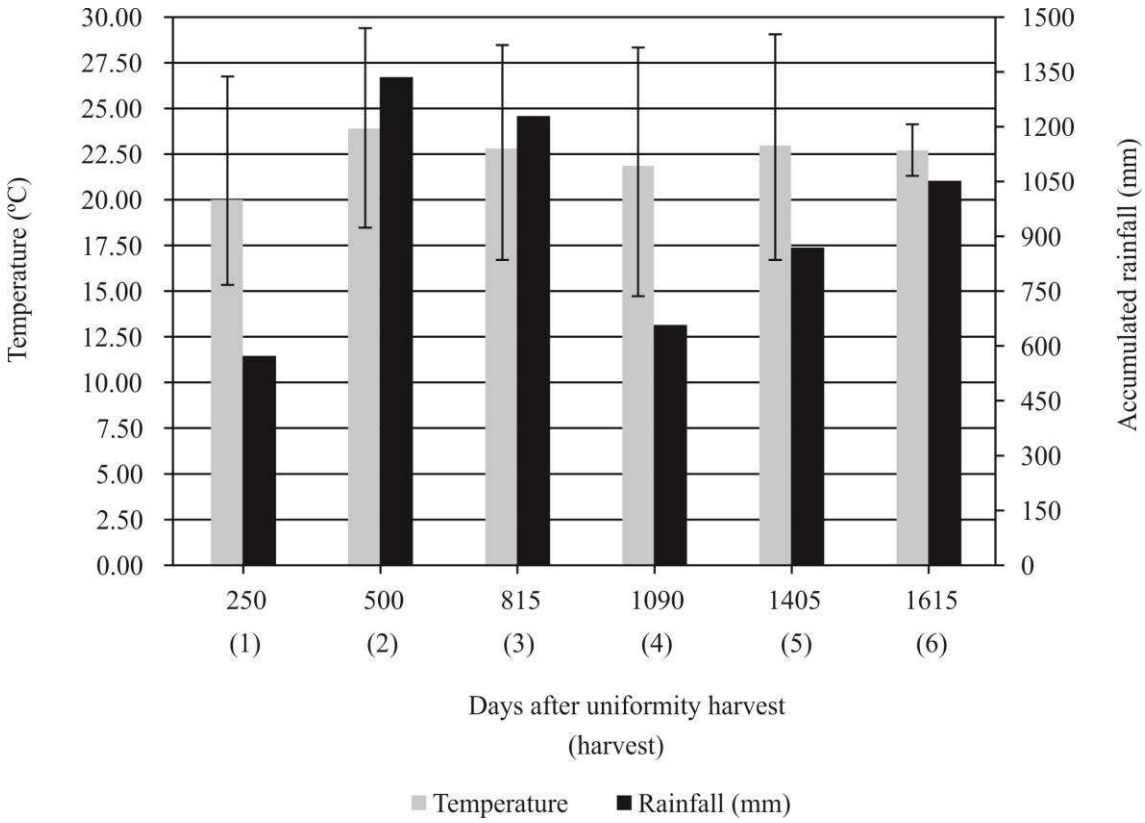

Fig S2. Accumulated rainfall (black bars) and average temperature (grey bars) with the maximum and minimum average temperature (deviation lines) during the multi-harvest.

Table S1. Registrations names of clones of the Active Elephantgrass Germplasm Bank (BAGCE) maintained by Embrapa Dairy Cattle Research Center and their respective code. †Code of accessions.

Code †	BAGCE registration	Code	BAGCE registration	Code	BAGCE registration	Code	BAGCE registration
1	Elefante da Colômbia	26	Mineiro	51	Guaco	76	12 AD IRI
2	BAGCE 2	27	Mole de Volta Grande	52	Cuba-115	77	07 AD IRI
3	Tres Rios	28	Porto Rico	53	Cuba-116	78	Pasto Panamá
4	Napier Volta Grande	29	Napier	54	Cuba-169	79	BAGCE 92
5	Mercker Santa Rita	30	Mercker Comum	55	King Grass	80	09 AD IRI
6	Pusa Napier Nº 2	31	Terezópolis	56	Roxo Botucatu	81	11 AD IRI
7	Gigante de Pinda	32	Taiwan A-26	57	Mineirão IPEACO	82	05 AD IRI
8	Napier Goiano	33	Duro de Volta Grande	58	Vruckwona Africano	83	06 AD IRI
9	Mercker S. E. A.	34	Mercker Comum Pinda	59	Cameroon	84	01 AD IRI
10	Taiwan A-148	35	Turrialba	60	BAGCE 69	85	04 AD IRI
11	Porto Rico 534-B	36	Taiwan A-146	61	Guaçu	86	13 AD IRI
12	Taiwan A-25	37	Cameroon - Piracicaba	62	Napierzinho	87	03 AD IRI
13	Albano	38	Taiwan A-121	63	IJ 7125	88	02 AD IRI
14	Híbrido Gigante da Colômbia	39	Vrukwona	64	IJ 7126	89	08 AD IRI
15	Pusa Gigante Napier	40	T241 Piracicaba	65	IJ 7127	90	Pioneiro
16	Elefante Híbrido 534-A	41	BAGCE 50	66	IJ 7136	91	Banhado
17	Costa Rica	42	BAGCE 51	67	IJ 7139	92	Roxo Farroupilha
18	Cubano de Pinda	43	Elefante Cachoeiro Itapemirim	68	IJ 7141	93	Roxo de Canguçu
19	Mercker Pinda	44	Sem Pelo	69	Goiano	94	Roxo do Itassú
20	Mercker 86 México	45	Capim Cana D'África	70	CAC 262	95	BRS Capiaçu
21	Taiwan A-144	46	Kizozi	71	Ibitinema	96	CNPGL 91-06-3
22	Napier S.E.A.	47	Gramafante	72	Australiano	97	CNPGL 96-25-3
23	Taiwan A-143	48	Roxo	73	BAGCE 82	98	BRS Canará
24	Pusa Napier Nº 1	49	Mott	74	13 AD	99	CNPGL 94-49-6
25	Elefante de Pinda	50	BAGCE 59	75	10 AD IRI	100	PCM 0701

CHAPTER 2

DISCOVERING CANDIDATE GENES UNDERLYING BIOMASS DIGESTIBILITY IN ELEPHANTGRASS

VIÇOSA
MINAS GERAIS - BRASIL
2019

ABSTRACT

ROCHA, João Romero do Amaral Santos de Carvalho, D. Sc., Universidade Federal de Viçosa, march de 2019. **Discovering candidate genes underlying biomass digestibility in elephantgrass.** Adviser: Pedro Crescêncio Souza Carneiro. Co-adviser: Marcos Deon Vilela de Resende and Juarez Campolina Machado.

Elephantgrass can be used for different purposes including bioenergy and animal feeding. As an orphan plant, genomic information in this species is rare and studies associating candidate genes to traits targeted by breeders are inexistent. In order to identify candidate genes that could be exploited for marker-assisted selection in elephantgrass this study aimed at verifying significant association of single sequence repeats markers (SSR) with eight traits linked to bioenergetics and animal feeding. The single-step genome-based best linear unbiased prediction (ssGBLUP) and genome association (GA) methods were applied to analyze the dataset. It was verified that the genomic information applied to the elephantgrass allowed increases on accuracy considering the biomass quality traits, even with a small number of markers. We found that one marker (M28_161) was significantly associated with large values of biomass digestibility. That marker was found with moderate linkage disequilibrium with another marker (M35_202) that, in general, is detected in genotypes showing small values of biomass digestibility. Bioinformatics analysis revealed both markers have orthologous regions in other C4 grasses as Setaria viridis, Panicum hallii and Panicum virgatum and these regions are located closely to candidate genes involved in biosynthesis of cell wall molecules (xyloglucan and lignin), which support their association with biomass digestibility. The markers and the candidate genes identified here are useful for breeders that need to change biomass digestibility in elephantgrass. This is the first trait-marker association study for this species.

Key words: Cenchrus purpureus, genome, ssGBLUP, GA, MAS.

RESUMO

ROCHA, João Romero do Amaral Santos de Carvalho, D. Sc., Universidade Federal de Viçosa, março de 2019. **Desvendando genes candidatos relacionados à digestibilidade da biomassa em capim-elefante.** Orientador: Pedro Crescêncio Souza Carneiro. Coorientadores: Marcos Deon Vilela de Resende e Juarez Campolina Machado.

O capim-elefante pode ser usado para diferentes finalidades, incluindo bioenergia e alimentação animal. A informação genômica nesta espécie é rara e estudos associando genes candidatos a características alvos dos melhoristas são inexistentes. A fim de identificar genes candidatos que podem ser explorados para a seleção assistida por marcadores em capim-elefante, este estudo teve como objetivo verificar associação significativa de marcadores single sequence repeats (SSR) ligados a sete características relacionadas a produção bioenergética e à alimentação animal. Foram aplicados para analisar o conjunto de dados os métodos single-step genome-based best linear unbiased prediction (ssGBLUP) e de associação genômica (GA). Verificou-se que a informação genômica aplicada ao capim-elefante permitiu aumento na acurácia considerando as características de qualidade da biomassa, mesmo com um pequeno número de marcadores. Descobrimos um marcador SSR (M28_161) com associação significativa com os maiores valores de digestibilidade da biomassa. Esse marcador foi encontrado em desequilíbrio de ligação moderado com outro marcador SSR (M35_202) que, em geral, é detectado em genótipos que apresentam pequenos valores de digestibilidade da biomassa. Análises de bioinformática revelaram que ambos os marcadores apresentam regiões ortólogas em outras gramíneas C4 como Setaria viridis, Panicum hallii e Panicum virgatum, e estas regiões estão localizadas próximas de genes candidatos que estão envolvidos na biossíntese de moléculas da parede celular (lignina e xiloglucano), o que suporta a associação com digestibilidade da biomassa. Os marcadores e os genes candidatos identificados aqui são úteis para os melhoristas que precisam alterar a digestibilidade da biomassa em capim-elefante. Este é o primeiro estudo de associação característica-marcador para esta espécie.

Palavras chave: Cenchrus purpureus, genoma, ssGBLUP, GA, MAS.

1. Introduction

Elephantgrass [Pennisetum purpureum Schumach. syn. Cenchrus purpureus (Schumach.) Morrone] is a perennial tropical grass with high photosynthetic efficiency (C4 photosynthetic pathway) that is naturally found in several African countries (Negawo et al. 2017). It has a range of adaptation to different levels of altitude, precipitation and soils and important agronomic traits (Pereira et al. 2010; Fontoura et al. 2015) being the most interesting one its high biomass production. Depending on the environmental characteristics, some cultivars can produce as much as 300 tons ha^{-1} $year^{-1}$ of green biomass reaching 170 kg dry matter $ha^{-1}day^{-1}$ (Gomide et al. 2015; Pereira et al. 2016). Based on that, elephantgrass has been aimed for multiple purposes including production of bio-based compounds (Chen et al. 2016; Takara and Khanal 2011), production of bioactive molecules with pharmaceutical-industry interest (Mambe et al. 2016) and co-products (Madakadze et al. 2010; Ituen et al. 2016; Ridzuan et al. 2016a; Ridzuan et al. 2016b; Fartini et al. 2016). It has also been a targeted of bioenergy programs with its dry matter production yearly being greater than sugarcane and eucalyptus, the most used biomass energy sources in Brazil (Fontoura et al. 2015; Rocha et al. 2017). Nevertheless, the most common use for elephantgrass is animal feeding, especially for dairy cattle.

Elephantgrass is a tetraploid species (2n = 4x = 28) with two genomes (A'A'BB) (Reis et al., 2014). Evidence have shown that its genome is homologous to the genome of pearl millet (Pennisetum glaucum (L.) R. Br. syn. Cenchrus americanus (L.) Morrone) and cytogenetics studies have been usually reported for these species in order to identify hybrids (Anderson et al. 2008; Reis et al., 2014). However, as a non-model ('orphan') plant, genomic information has been rarely reported for elephantgrass. Although the first RNAseq study (Zhou et al. 2018) and a de novo whole genome sequencing have been recently published (Wang et al. 2018), much are needed to lead the way for a next-generation breeding of elephantgrass (Pereira et al. 2018). For instance, trait-marker association studies are inexistent mainly due to the low number of molecular markers available. Among the few available ones there are SSR markers transferred to elephantgrass from pearl millet (Azevedo et al. 2012). These markers have been used for genetic diversity studies aiming to enhance the diversity in genebank collections (Negawo et al. 2018). Even though SSR markers appear in the genome in a frequency lower than SNPs, they can be used for genomic association studies. For

instance, SSR markers have been used to identify loci associated with yield components and fiber quality in a collection of upland cotton accessions (Qin et al. 2015), with resistance to Sclerotinia sclerotiorum in Brassica napus (Gyawali et al. 2016) and with morphological quantitative traits that contribute to yield and juice quality in sugarcane (Siraree et al. 2017).

The interest in performing genomic association is to dissect the genetic basis of important traits by identifying candidate genes and markers linked to them. When considering multipurpose traits (i.e., traits interesting for both animal feeding and bioenergy), the development of selection procedures based on molecular marker information can optimize the genetic breeding (Biazzi et al. 2017). Among the common traits for forage and bioenergetics use there are agronomic traits (height, green and dry biomass) and quality or nutritive value traits (acid and neutral detergent fiber, biomass digestibility, dry matter concentration and lignin). Since elephantgrass are commonly vegetatively propagated and crosses are not widely used, the use of germplasm collections for association mapping is interesting for being based on historical and naturally occurred recombination events. Thereby, the goal of this study was to verify the predictive accuracy considering the genomic relationship information and significant association between SSR markers and eight traits, evaluated in different cuts, in a collection of 100 elephantgrass genotypes. Our goal extended to the identification of candidate genes that can be used for marker-assisted selection.

2. Materials and methods

2.1. Genetic material and experimental information

In December 2011, one hundred elephantgrass genotypes (the registration genotypes' names are shown in Supplementary Table S1) were planted in 0.20 m deep furrows, with 80 kg ha^{-1} P_2O_5 fertilizer applied at planting. The red-yellow latosol soil of the area in the experimental station of Embrapa Gado de Leite in Coronel Pacheco, MG, Brazil (lat 21º33'18'' S, long 43º15'51" W, at 417 masl) had the following chemical properties: pH (5.4), H+Al (2.31 cmolc dm^{-3}), P (1.1 cmolc dm^{-3}), K (23 mg dm^{-3}) and exchangeable cations Al^{3+} (0.2 cmolc dm^{-3}), Ca^{2+} (1.4 cmolc dm^{-3}) and Mg^{2+} (0.7 cmolc dm^{-3}). The plots consisted of a single 4 m row where rows were planted side by side, spaced 1.5 m apart. Plots were allocated in a 10 x 10 simple lattice design, with

two replications. At 30 days after planting, elephantgrass plots were cut to 0.30 m stubble height (uniformity cutting). The first of the six growth seasons started at this time. Maintenance fertilization was made with 300 kg ha^{-1} of the $N\text{-}P_2O_5\text{-}K_2O$ formulation (20:05:20 blended granular fertilizer), after all cuttings. Fertilization was executed according to the soil analysis.

Six cuttings were carried as follows: first and second cuttings (September 28th, 2012 and June 04th, 2013, respectively) were performed at 250 regrowth days, third cut (April 15th, 2014) at 315 regrowth days, fourth cut (January 15th, 2015) at 275 regrowth days, fifth cut (November 26th, 2015) at 315 regrowth days and the sixth cut (June 22th, 2016) was performed at 210 regrowth days. The fourth cut was destined to collect the propagation material, i.e., no phenotypic data.

2.2. Phenotypic traits

The plants were phenotyped at first, second, third, fifth and sixth cuts for each of the following traits: (i) height (m) was obtained from the arithmetic mean of the height of three random plants samples, in each plot, measured from the ground level to the curve of the last completely expanded leaf; (ii) green biomass (Mg ha^{-1}) was obtained from a cut at 7.5 cm stubble height in a 3 m section from the middle of rows, using a gasoline-powered trimmer and then collected by hand. The 3 m section was immediately weighed in the field to provide estimates of green biomass; (iii) dry biomass (Mg ha^{-1}) was quantified by multiplying the green biomass by the dry matter concentration (%); (iv) acid detergent fiber (g Kg^{-1}), (v) neutral detergent fiber (g Kg^{-1}) and (vi) lignin content (g Kg^{-1}) were determined following the methodology proposed by Goering and Van Soest (1970) for the first and second cuts.; (vii) biomass digestibility (g Kg^{-1}) was determined by the method described by Tilley and Terry (1963) for the first and second cuts; and (viii) dry matter concentration (%), for the first and second cuts, that was obtained by sampling three complete plants from each plot, which were dried in a kiln after weighing (fresh weight) until weight stabilization. Samples were weighed (dry weight) again, and then the dry matter concentration was determined by the ratio between dry weight and fresh weight. This trait was used as a common denominator for the estimation of biomass digestibility. For acid detergent fiber, neutral detergent fiber and biomass digestibility, random samples of three complete plants from each plot were collected before cutting the experimental plots.

Then, these samples were dried in a forced air circulation oven at 56°C for 72 hours. After drying, samples were ground to small particles (1 mm) in a Wiley type grinder and analyzed as described above.

For the third, fifth and sixth cuts the traits acid and neutral detergent fiber, lignin biomass digestibility and dry matter concentration were obtained via Near Infrared Spectroscopy (NIRS). The data generated, on the first and second cuttings and from other experiments, by the traditional methodologies of biomass quality analysis, were used for NIRS calibration.

2.3. Genotyping, quality control, and imputation

Single sequence repeats (SSR) markers were used for genotyping as described by Azevedo et al. (2012). Due to multiallelic nature of SSR markers, each allele was considered as a marker (totalizing 111 markers). For each marker, individuals were coded as 0 (absence of allele) or 1 (presence of allele) according to Viana et al. (2016). SSRs with more than 15% missing values (i.e. call rate of at least 85%) and/or frequency of minor allele above 1% were removed. We considered the following imputation algorithms for the missing value data point on a matrix of markers (M_i):

$$M_i = \begin{cases} \text{if}: & p_i \le 0.5 \rightarrow M_i = 0 \\ \text{if}: & 0.5 < p_i \le 1 \rightarrow M_i = 1 \end{cases}$$

In this algorithm p_i is the allele frequency associated with the presence of the marker in the locus i. The algorithm was directly implemented as an R function. After the quality control, 87 markers were used**.**

The SSR markers used at this work are distributed on pearl millet chromosomes 1, 3, 4, 5, 6, and 7 according to Allouis et al. 2001, Budak et al. (2003), Mariac et al. (2006) and Rajaram et al. (2013).

2.4. Single-step genome-based best linear unbiased prediction (ssGBLUP)

The mixed model methodology was adopted for statistical analyses via ssGBLUP (Legarra et al. 2009; Aguilar et al. 2010). The statistical model was denoted by: $y = Xm + Za + Zg + Wb + Ti + Tr + Qp + \varepsilon$, where:

y = vector of response across the five cuttings (just with phenotypic data);

m = vector of the effects of the measurement-replication combination (assumed as fixed) added to the overall mean;

a = vector of genetic effects (assumed as random);

g = vector of residual genetic effects (assumed as random);

b = vector of block effects (assumed as random);

i = vector of the genotype by cutting effects;

r = vector of the residual genotype by cutting interaction effects;

p = vector of the permanent environment effects (random);

ε = vector of residues (random)

X, Z, W, T, and Q represent the incidence matrices for these effects.

Distributions of random effects are considered $a \sim N(G \otimes \sigma_a^2)$; $g \sim N(I \otimes \sigma_{rg}^2)$; $b \sim N(I \otimes \sigma_b^2)$; $i \sim N(G_i \otimes \sigma_i^2)$; $r \sim N(I \otimes \sigma_{ri}^2)$; $p \sim N(I \otimes \sigma_p^2)$; $\varepsilon \sim N(I \otimes \sigma_e^2)$ where G is a matrix of genomic additive relationship, G_i is a matrix of genomic interaction (genotype by cutting interaction) relationship, I is the identity matrix of appropriate dimension, σ_a^2, σ_{rg}^2, σ_b^2, σ_i^2, σ_{ri}^2, σ_p^2 and σ_e^2 are the additive, residual genetic, block, genotype by cutting interaction, residual genotype by cutting interaction, permanent environment and the residual variances components, respectively. The model above includes the residual genetic effects and residual genotype by cutting interactions, according to Oakey et al. (2016).

The additive relationship matrix structures (G) were used according to Resende et al. (2014), which is denoted by: $G = \frac{Z^* Z^{*'}}{\sum_i^n p_i(1-p_i)}$; where: $Z^* = Z - P$, where Z is a matrix containing marker genotypes, and P is a matrix with p_i elements in column i.

Due to the presence of 10 genotypes ungenotyped and 90 genotypes genotyped, it was adopted the inverse of genomic relationship matrix (H^{-1}), according to Legarra et al. (2009): $H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{*-1} - A_{22}^{-1} \end{bmatrix}$; where A^{-1} is the inverse of pedigree relationship for all elephantgrass genotypes and A_{22}^{-1} is the inverse of pedigree relationship only for the genotyped elephantgrass genotypes. A and A_{22} are an identity

matrix for this study because there is no information about the pedigree. To compute the exact inverse of G to compose the H^{-1} matrix, it was used the following algorithm, $G^{*} = 0.95G + 0.05A_{22}$ (Masuda et al. 2016).

From the matrix, H^{-1} the ssGBLUP procedure was run according to the specified model. For the random effects of the model, the significance for the likelihood ratio test was tested by using the chi-square test with one degree of freedom. The ssGBLUP was performed with the ASReml 4.1 software (Gilmour et al. 2015).

The accuracy $\left(r_{\hat{g}g}\right)$ for the clone effect (additive plus residual genetic effects) considering the ssGLUP model was estimated as following:

$$r_{\hat{g}g} = \sqrt{1 - \frac{PEV_a + PEV_{rg}}{\hat{\sigma}_a^2 + \hat{\sigma}_{rg}^2}}$$

Where, PEV_a and PEV_{rg} is the prediction error variance, obtained by the diagonal elements of inverse of the left rand side of mixed models equation for additive and residual genetic effects, respectively.

2.5. Simple repeatability plus G x C interaction model

Additionally, it was run the model without considered the relationship (Simple repeatability plus G x C interaction model) between the genotypes, as following:

$y = Xm + Zg + Wb + Ti + Qp + \varepsilon$, where:

y = vector of response across the five cuttings (just with phenotypic data);

m = vector of the effects of the measurement-replication combination (assumed as fixed) added to the overall mean;

g = vector of genetic effects (assumed as random);

b = vector of block effects (assumed as random);

i = vector of the genotype by cutting effects;

p = vector of the permanent environment effects (random);

ε = vector of residues (random)

X, Z, W, T, and Q represent the incidence matrices for these effects.

Distributions of random effects are considered $g \sim N(I \otimes \sigma_g^2)$; $b \sim N(I \otimes \sigma_b^2)$; $i \sim N(I \otimes \sigma_i^2)$; $p \sim N(I \otimes \sigma_p^2)$; $\varepsilon \sim N(I \otimes \sigma_e^2)$ where I is the identity matrix of

appropriate dimension, σ^2_g, σ^2_b, σ^2_i, σ^2_p and σ^2_e are the genetic, block, genotype by cutting interaction, permanent environment and the residual variances components, respectively.

The accuracy $\left(r_{\hat{g}g}\right)$ for the clone effect (genetic effects) considering the simple repeatability plus G x C interaction model was estimated as following:

$$r_{\hat{g}g} = \sqrt{1 - \frac{PEV_g}{\hat{\sigma}^2_g}}$$

Where, PEV_g is the prediction error variance, obtained by the diagonal elements of inverse of the left rand side of mixed models equation genetic effects, respectively.

2.6. Genome association study

The 90 elephantgrass genotyped and phenotyped were used in the genome association study, i.e., by means of R package sommer, GWAS2 function (Covarrubias-Pazaran 2016), it was verified the significant association between markers and phenotypic traits. In the presence of genotype by cutting interaction, the association study was performed for each cutting, using the model: $y = Xm + Mu + Za + Zg + Wb + \varepsilon$, where:

y = vector of response for each cutting;

m = vector of the effects of the replication (assumed as fixed) added to the overall mean;

u = vector of markers (assumed as fixed);

a = vector of genetic effects (assumed as random);

g = vector of residual genetic effects (assumed as random);

b = vector of block effects (assumed as random);

ε = vector of residue (random)

X, Z, M, and T represent the incidence matrices for these effects.

Distributions of random effects are considered $a \sim N(G \otimes \sigma^2_a)$; $g \sim N(I \otimes \sigma^2_{rg})$; $b \sim N(I \otimes \sigma^2_b)$ and $\varepsilon \sim N(I \otimes \sigma^2_e)$. Where G is a matrix of genomic additive relationship, I is the identity matrix of appropriate dimension, σ^2_a, σ^2_{rg}, σ^2_b and σ^2_e are the additive, residual genetic, block, and the residual variance, respectively.

Markers with -log (P-value) up to false discovery rate method (FDR) threshold was considered a candidate marker. To compute FDR was considered a 0.02 threshold level and considering P-value lower than 0.05.

It was used the R package LDcorSV by means of LD.Measures function to compute the linkage disequilibrium (r^2) as describe by Hill e Robertson (1968): $r^2 = \frac{[p(AB) - p(A)p(B)]^2}{p(A)p(B)[1-p(A)][1-p(B)]}$; where, $p(AB)$ is the frequency of the haplotype AB, $p(A)$ and $p(B)$ are the frequency of the alleles A and B, respectively. Thus, r^2 ranged from 0 (when the two markers are in perfect equilibrium) to 1 (when the two markers provide identical information). Manhattan plots and quantile-quantile (QQ) plots were obtained by the "sommer" package (Covarrubias-Pazaran 2016) in R project (R Development Core Team 2015).

2.7. Nucleotide sequence alignment, candidate genes and gene annotation

A list of candidate genes (marked significantly associated with the traits) was assembled by BLAST, with default parameters, using the plant comparative genomics portal Phytozome (Goodstein et al. 2012) based in related species (reference genomes, e.g., Setaria italica, Setaria viridis, Panicum virgatum and Panicum halli). BLAST against the genome of Cenchrus americanus was performed at the NCBI webpage (https://blast.ncbi.nlm.nih.gov/Blast.cgi). For these searches, the sequences of the primers published by Azevedo et al. (2012) were used. For gene annotation, orthologs genes in Arabidopsis thaliana were searched at the TAIR (The Arabidopsis Information Resource) database (Reiser et al. 2017).

3. Results

3.1. Genetic variation, genotype by cutting interaction and accuracy under ssGBLUP model

Initially, the ssGBLUP model was used to fit the full data set, enabling testing the genetic variability under genotype by cutting interaction as well as testing the residual genetic and the residual genotype by cutting interaction effects. Genetic variability ($p<0.05$) was detected for all traits (height, green biomass, dry biomass, dry

matter concentration, acid detergent fiber, neutral detergent fiber and lignin) except for biomass digestibility (Fig 1). Regarding the genotype by cutting interaction, significant effect ($p<0.05$) for all traits was observed (Fig 1). It is noteworthy that, in the individual analysis of each cutting, significant genetic variability was detected for all traits.

The log-likelihood ratio tests (LRT) for residual genetic effect suggest that this effect should be considered ($p<0.05$) for height, green biomass and dry biomass. However, for the biomass quality traits (i.e., dry matter concentration, acid detergent fiber, neutral detergent fiber, biomass digestibility and lignin) the residual genetic effect was not significant ($p>0.05$) (Fig 1.). For residual genotype by cutting interaction it was detected significant effects ($p<0.05$) just for green biomass, dry biomass, dry matter concentration and lignin (Fig 1).

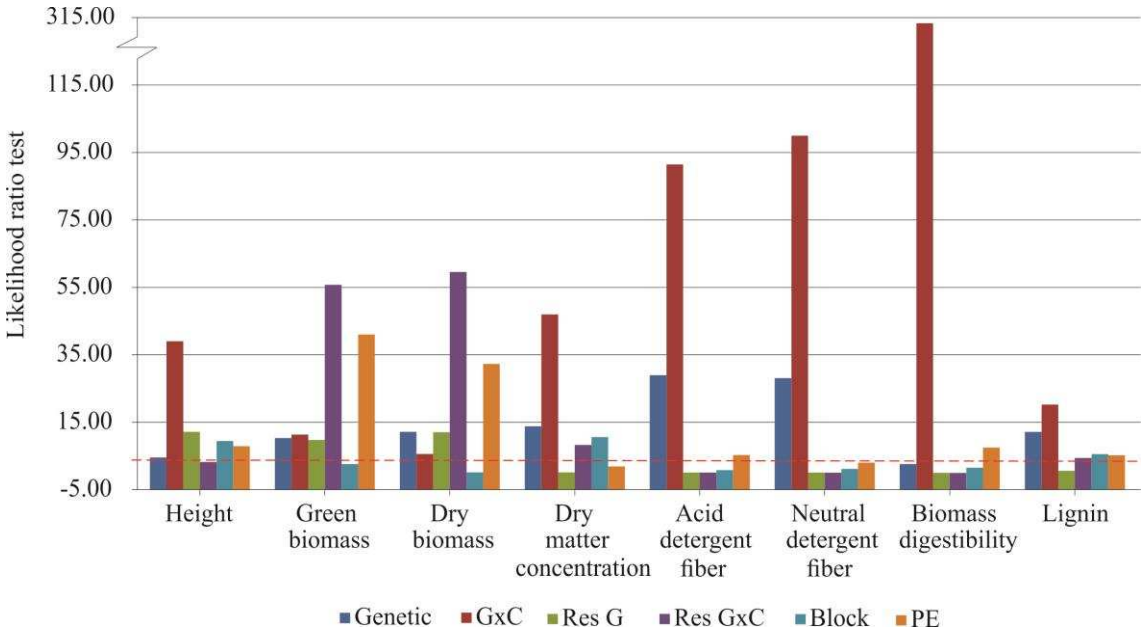

Fig 1. Likelihood ratio tests for genetic effects (Genetic), genotype by cutting effects (GxC), residual genetic effects (Res G), residual genotype by cutting interaction effects (Res GxC), block effects (Block) and permanent environment effects (PE) considering ssGBLUP model. All bars above the dashed red line are significant by chi-square test at 5% probability ($X^2_{5\%} = 3.84$).

The accuracy values for the ssGBLUP model, i.e., considering the H^{-1} matrix, ranged from 0.58 (biomass digestibility) to 0.84 (acid detergent fiber). When the model not consider the relationship between the genotypes (simple repeatability plus genotype by cutting interaction model), the accuracy ranged 0.59 (biomass digestibility) to 0.89

(dry biomass). We can also note that the inclusion of the relationship matrix via ssGBLUP model provided increases on accuracy for the biomass quality traits (Fig 2), exception for biomass digestibility, therefore, molecular markers (genomic relationship) may be a promising tool to raise the prediction accuracy for elephantgrass breeding.

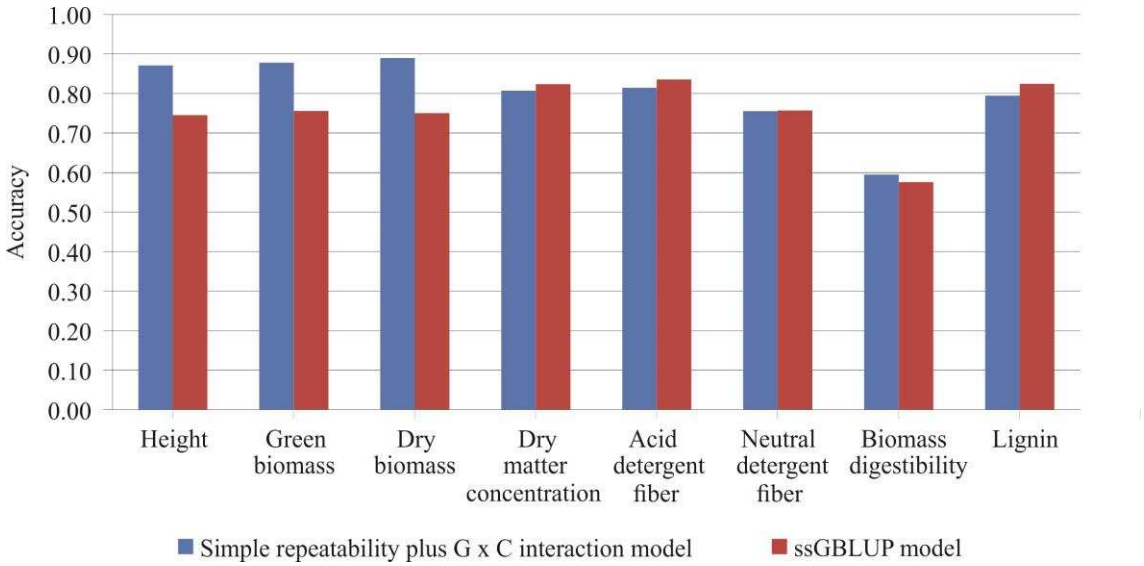

Fig 2. Accuracy of breeding value for the simple repeatability plus genotype by cutting interaction model and ssGBLUP model.

3.2. Overall means and accuracy considering one cutting at a time

Since the significant effect of genotype by cutting interaction was observed for all traits (ssGBLUP model, Fig 1), the trait-marker association was performed by cutting. The overall means and accuracy of each cut are illustrated in Fig 3. The accuracy values ranged from 0.47 (acid detergent fiber and biomass digestibility on the fifth cut and biomass digestibility on the first cut) to 0.88 (dry matter concentration on the second cut). In general, the overall means for all traits varied widely over the multi-cutting (corroborating with significant genotype by cutting interaction). The second cut, for instance, showed the greatest values for height, green biomass, dry biomass, acid detergent fiber, neutral detergent fiber and lignin, whereas, showed the lowest values for biomass digestibility.

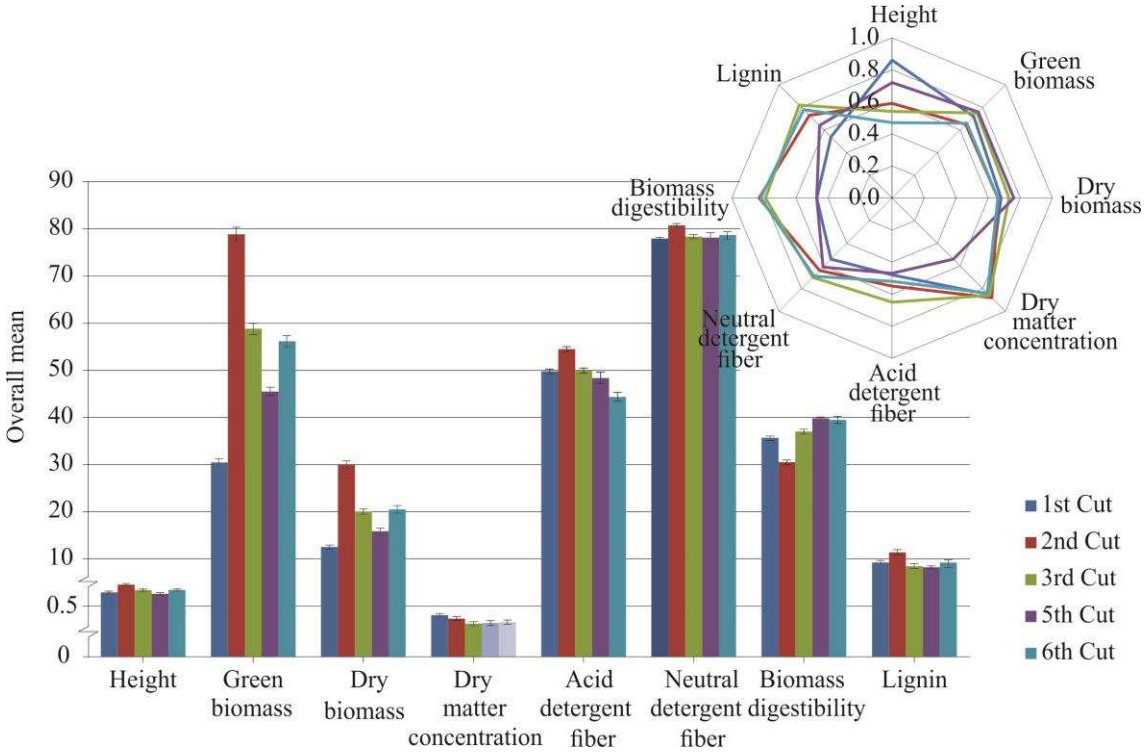

Fig 3. Overall means (bars plot) with the standard errors and accuracy of breeding values (radar plot) for five cuttings recorded in 90 genotypes of elephantgrass.

3.4. Genome association study

By analyzing 90 elephantgrass genotypes, we identified one allele of the M28 marker (M28_161) as significantly associated with biomass digestibility. This association was detected only for the first cut. The Manhattan plot (Fig 4, on the right) shows the -log(P-values) for all SSR affecting the biomass digestibility, while the quantile-quantile (QQ) plot (Fig 4, on left) displays significant deviations of observed -log(P-values) from those expected.

To identify the linkage of all markers evaluated in the present study, we performed linkage disequilibrium analysis (Fig 5). In this analysis, we were interested in identifying markers potentially linked to the marker M28_161 that is significantly linked to the large values of biomass digestibility. One allele of the SSR marker M35 (M35_202) showed moderate linkage disequilibrium (r^2 = 0.20) to M28_161 and, in general, its presence was detected in elephantgrass genotypes with small values of biomass digestibility. All other r^2-values were smaller than 0.11 varying from 0.00 to 0.11.

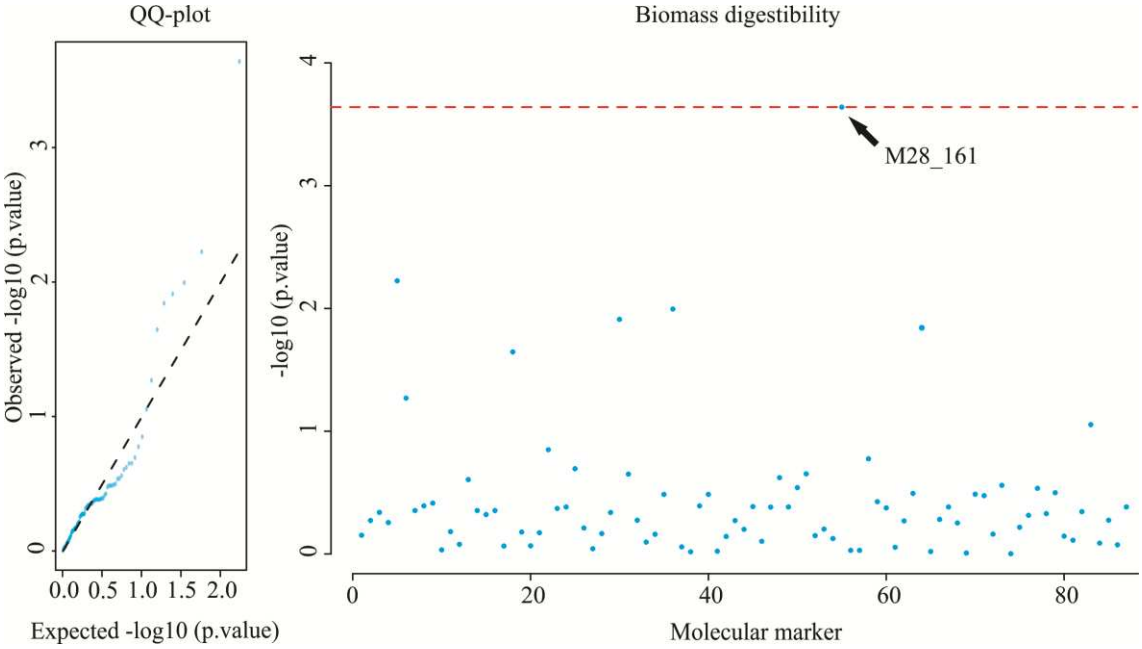

Fig 4. QQ-plot (on the left) and Manhattan plot (on the right) for genome association of biomass digestibility phenotype with SSR markers on the first cut of elephantgrass. Dashed red lines on Manhattan plot (false discovery rate) indicate the minimum threshold to select significant markers. The arrow highlights a major trait-marker association.

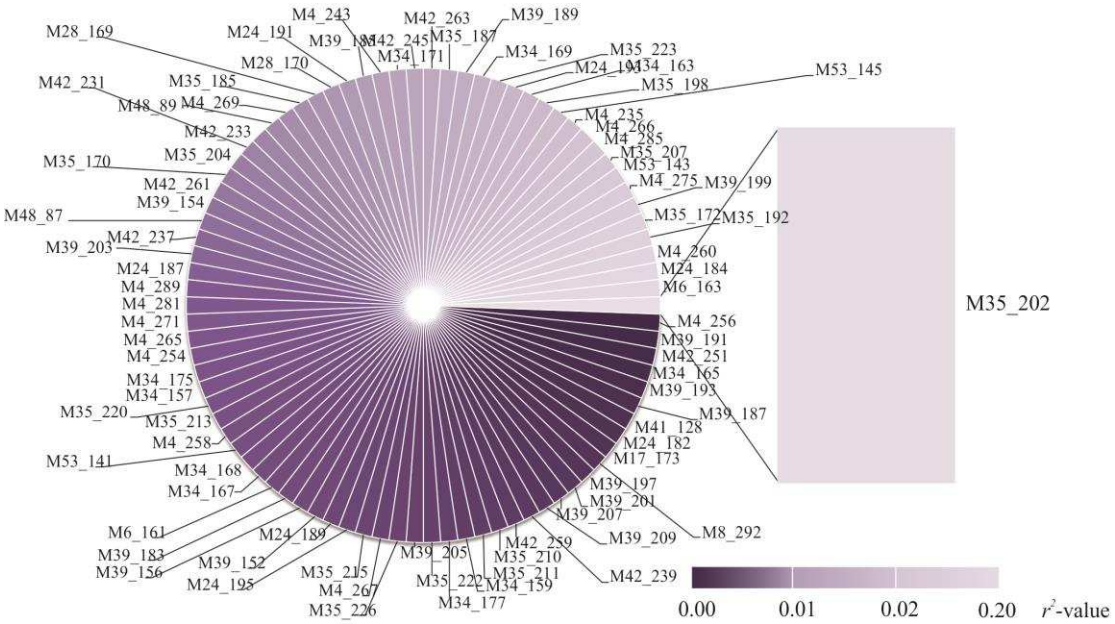

Fig 5. Linkage disequilibrium (r^2 -values) between all markers used in this study with the marker M28_161, which was found as significantly associated with biomass digestibility in elephantgrass.

3.5. Annotation of M28_161 and M35_202 markers, in silico pathway analysis and allelic contribution

BLAST search revealed that the two SSR markers (M28_161 and M35_202) are close to candidate genes that have annotated functions in other grasses (Table 1). In some cases (i.e. Setaria viridis chromosome 3 and Panicum halli chromosomes 5 and 7), the marker M28_161 is close to candidate genes linked to pathways influenced by plant hormones (salicylic acid or abscisic acid). However, in most cases, the markers (M28_161 and M35_202) are associated with candidate genes involved in synthesis of cell wall components. In Panicum halli, M28_161 is close to candidate genes involved in lignin biosynthetic process and cell wall organization (chromosome 7) and, in Panicum virgatum, that marker is close to a candidate gene involved in lignin catabolic process and oxidation-reduction process (located on chromosome 8). In Setaria viridis, both M28_161 and the M35_202 were found to have orthologous regions. The M28_161 marker is close to a candidate gene located on chromosome 3 that acts in the xyloglucan biosynthetic process, while the M35_202 marker is close to a candidate gene located on the chromosome 7 and acts in the lignin biosynthetic process. For Setaria italica, successful BLAST occurred for the sequence of M28_161 marker, but no candidate genes related to digestibility trait were found. On the other hand, for Cenchrus americanus no BLAST results were acquired when using the sequences of M28 and M35 markers.

Table 1. Candidate genes from other C4 species that were identified in genomic regions near to sequences orthologous to the SSR markers M28_161 and M35_202.

Marker	Candidate gene	Reference genome	Chromosome	Biological pathway	Ortholog locus on A. thaliana	Locus position*	Marker position
M28_161	Sevir. 3G340800	Setaria viridis	3	Xyloglucan biosynthetic process; salicylic acid mediated	AT2G20370	40,675,315	40,676,270
M28_161	Pahal. G00901	Panicum halli	7	Lignin biosynthetic process; cell wall organization	AT5G48930	29,711,375	29,776,838
M28_161	Pahal. G00889	Panicum halli	7	Salicylic acid mediated	AT5G05190	29,776,838	29,776,838
M28_161	Pahal. E03247	Panicum halli	5	Response to abscisic acid	AT3G05880	49,635,160	49,627,286
M28_161	Pavir. 8KG357500	Panicum virgatum	8	Lignin catabolic process, oxidation-reduction process	AT3G09220	72,822,520	72,841,944
M35_202	Sevir. 7G164200	Setaria viridis	7	Lignin biosynthetic process; cell wall organization	AT5G48930	23,027,630	23,030,255

*Gene that is closest to the SSR marker.

4. Discussion

4.1. Genetic variation, genotype by cutting interaction and prediction accuracy

Here we observed a significant effect of genotype by cutting for all the eight traits analyzed (Fig 1). Statistically, this indicates no coincident genotypes performance in relation to the different cuttings (van Eeuwijk et al. 2016), and genetically, it means that the cuttings promoted different levels of genes expression (Nicotra et al. 2010). Therefore, it can be inferred that selection should be carried out in a specific cutting. Faced with this fact, we opted to carry out the genomic association study considering a single cut at a time.

It is worth mention that the absence of genetic variability (i.e. for biomass digestibility) by ssGBLUP analysis does not mean that in fact there is no genetic variability, because, the interaction effect may reduce the genetic variability. Residual terms were added on the ssGBLUP model to capture nonadditive effects. For crops that we can propagate via clonal and are not lines (as elephantgrass) the nonadditive effect will represent epistatic and dominance effects (Oakey et al. 2016) and additive effects not explained by the genomic relationship matrix, therefore without inflating the residual estimates and without overlay to the other effects. The value of using the total genetic effect (additive plus nonadditive) for phenotypic evaluation would depend on the impact of the nonadditive proportion and of course of the target of the breeding program. If the breeding target is identifying the best clone and recommend it to farmers the total genetic effect reflect the potential of this clone. However, it is worth noting that the total genetic effect does not reflect the potential of a clone as a parent because nonadditive effects are not inherited (Oakey et al. 2016). The additive value may better reflect the potential for determining which clones to be crossed in a breeding program for elite genotypes development.

The not significant residual genetic effect for the biomass quality traits gives us indications that even with few markers (87 alleles) these were able to explain part of the additive fraction that acts in the genetic architecture of these traits. For the morpho-agronomic traits the residual genetic effect was significant, so the genomic relationship matrix constructed from 87 markers did not allow explanations about the additive genetic fraction for height, green biomass and dry biomass. This allows us to speculate about greater genetic complexity for these traits.

When we compared the predictive accuracy of the ssGBLUP and the simple repeatability plus genotype by cutting interaction models, we can observe that include relationship information on the model allowed increase the accuracy for biomass quality traits, even with a low number of molecular markers, exception for biomass digestibility.

In the biomass digestibility all cell components fractions could affect the forage digestibility. The main components are: cellulose, lignin, hemicellulose and cell wall proteins (Chen, 2014; Pauly and Keegstra 2016). In this way, the biomass digestibility is dependent of several other traits. Thus, its genetic complexity becomes greater which explain the fact that the predicted accuracy was not greater considering the genomic information.

With the RNAseq study made by Zhou et al. (2018), de novo whole genome sequencing recently published by Wang et al. (2018) and further efforts for the development of SNPs (single nucleotide polymorphisms) to elephantgrass, the gains in accuracy may be even greater. Studies with molecular markers began to present even greater importance from the development of single nucleotide polymorphisms (SNP) markers. Since then, new technologies of genotyping, sequencing and new bioinformatics tools have allowed greater accuracy of information and lower price per sequenced base or molecular genotype, and reduced the selection time (Pereira et al. 2018).

4.2. Trait-marker association analysis

In this study we performed a trait-marker association analysis based on SSR markers and eight important agronomic and quality traits of elephantgrass. These traits can dictate the end use of elephantgrass germplasms since they influence their utilization either as animal feeding or as bioenergy (Rocha et al. 2017). None of the agronomic traits (height, green and dry biomass) neither three of the quality traits (acid and neutral detergent fiber, dry matter concentration and lignin) were associated with markers. This is probably related to the low number of SSR markers used (18 SSR markers that originated 87 alleles) when compared to the genome size of elephantgrass, recently estimated in 2.1 Gb (Wang et al. 2018). However, one quality trait (biomass digestibility) was shown to be associated with the SSR marker M28_161 when the dataset from the first cut was analyzed (Fig 4). That marker is in linkage disequilibrium

with the marker M35_202 that, consequently, is considered to be associated with biomass digestibility. However, these markers are linked to different values of biomass digestibility since M28_161 is significantly associated with large values of biomass digestibility while M35_202 is more frequently detected in genotypes showing small values of biomass digestibility. This is an interesting result considering that digestibility is a valuable trait for many plant species and understanding its impact on plant quality and the genomic regions associated with it have been targeted by many research groups (Wang. et al. 2016; Grev et al. 2017; Leng et al. 2018).

4.3. Markers' annotation and in silico pathway

Based on the association of biomass digestibility with markers M28_161 and M35_202, we were interested in search for candidate genes underlining this phenotype. For this, we ran BLAST searches using the sequences of these markers and the genome of the related C4 grasses Setaria italica, Setaria viridis, Panicum virgatum, Panicum halli and Cenchrus americanus (Table 1). These species are closely related to elephantgrass as demonstrated by the phylogenetic studies (Bennetzen et al. 2012; Huang et al. 2016). In these BLAST analyses, no results were found for the C. americanus genome sequence. This was unexpected mainly because the SSR markers used here were developed from pearl millet (Qin et al. 2015) and because the A'A' genome of elephantgrass is homologous to A genome of pearl millet (Anderson et al. 2008;Reis et al 2014). It is unknown what could have limited the BLAST search. For the other species, we were able to identify six candidate genes that, in general, are associated with biosynthesis of cell wall molecules (Table 1).

One candidate gene annotated in the chromosome 3 of Setaria viridis (Sevir. 3G340800) is orthologous to the locus AT2G20370 of Arabidopsis thaliana and codes for a xyloglucan galactosyltransferase responsible for different functions including synthesis of cell wall materials (Table 1). The discovery of a candidate gene associated with xyloglucan is interesting and important to puzzle out the digestibility trait in elephantgrass. Xyloglucan is a component of the plant cell wall that can be solubilized only with strong chaotropic agents (see review in Pauly and Keegstra 2016). It is a type of hemicellulose that has the ability to bind to cellulose forming a cellulose-xyloglucan network linked through hydrogen bonds. Its importance relies on its ubiquitousness in the cell walls of all land plants. It is unclear

if xyloglucans hampers biomass digestibility but it can be related to lignin, a cell wall component that commonly negatively affects digestibility. This because xyloglucan binds to cellulose and cellulose forms an aggregate that, in some types of cells, can be embedded in a matrix that contains lignin (Pauly and Keegstra 2016).

Lignin, the second most abundant biopolymer on Earth (Boerjan et al. 2003), is a highly condensed phenylpropanoid matrix that is relatively difficult to be digested by ruminal microorganisms and intestinal enzymes (Liu and Yu 2011). Here we do not find marker-lignin association. This was probably due to lack of precision on lignin phenotyping step, since we found significant marker-digestibility association and the lignin is one of the main components of the forage biomass that affect the digestibility. For many plant species there is a strong negative correlation between lignin content and digestibility (Wu et al. 2013; Lam et al. 2017) that can impact animal performance since a small increase in dry matter digestibility (1%) can increase the daily weight gains for beef cattle by 3.2% (Casler and Vogel 1999). It does not mean that decreasing the lignin content is the only way to improve plant digestibility by it is commonly accepted that lignin is an interesting target for that purpose. In this context, the candidate genes Pahal. G00901, Pavir. 8KG357500 and Sevir. 7G164200, annotated in the genome of Panicum halli, Panicum virgatum and S. viridis respectively, are interesting assets. These candidate genes are orthologous of the loci AT5G48930, AT3G09220 and AT5G48930 in Arabidopsis (Table 1). AT5G48930 is a hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase involved in the phenylpropanoid pathway that has a role in the production of the hydroxycinnamyl alcohols (or monolignols) that serve as the building blocks of lignin (Fraser and Chapple 2011).

The lignin content is important not only to improve forages quality but it can also impact bioenergy production (Fu et al. 2011). In this case, the lignin content can be reduced or increased depending on how the biomass will be treated to generate energy. For instance, for the conversion of lignocellulosic biomass to ethanol, the polysaccharides from the cell wall need to be hydrolyzed to simple sugars and then be fermented to ethanol (Mielenz 2001). In this case, reducing the lignin content can increase ethanol production through conventional biomass fermentation (Fu et al. 2011). However, when considering biomass combustion, greater level of lignin is needed to rich high energy conversion (Dorez et al. 2014). Since the molecular basis for lignin content in elephantgrass is unknown, the candidate genes identified here can be

used as target for gene manipulation. Either by transgenesis or by gene editing, the manipulation of genes associated with lignin production can result in germplasm more ideally destined to animal feeding or bioenergy (Fu et al. 2011; Zhou et al. 2015; Lam et al. 2017).

Some candidate genes are associated with plant hormones (salicylic acid - SA or abscisic acid - ABA) and two of them show interesting features. The candidate gene Pahal. G00889 (annotated in the chromosome 7 of Panicum halli) is orthologus to the locus AT5G05190 in Arabidopsis where a hypothetical protein is involved in different functions including response to fungus (Wu et al. 2015). Meanwhile, the candidate gene Pahal. E03247 (also from P. halli) has as orthologous in Arabidopsis the locus AT3G05880 that codes for a small and highly hydrophobic protein involved in hyperosmotic salinity response and response to cold (Table 1). In this way, in the genomic region associated with biomass digestibility in elephantgrass, two genes associated with biotic and abiotic stresses could reside. Plant hormones, as SA and ABA, and their cross-talk play an important role in the molecular response of plants to biotic and abiotic stresses (Nguyen et al. 2016). Thus, validating these candidate genes in elephantgrass can increase our knowledge on how this species respond to abiotic and abiotic stresses like spittlebug attack and water stress, which are important targets for breeders in Brazil (Pereira et al. 2018).

In conclusion, elephantgrass biomass is an interesting feedstock for animal production and additionally it stands out in the biorefinery concept. Because genomic information for elephantgrass is rare, our findings revealed two SSR markers associated with biomass digestibility that may innovate the way elephantgrass breeding is conducted. For instance, marker-assisted selection (MAS) can be applied. Molecular markers associated with biomass digestibility are extremely relevant since they may drive the elephantgrass selection for different uses (e.g. bioenergy production, bio-based products, co-products, bioactive compounds and animal feed). Moreover, further investigation and validation of the candidate genes unrevealed here may contribute to a better understanding of biomass digestibility variation and its genomic basis. Additionally, the manipulation of these candidate genes may, therefore, be a useful biotechnological tool to increase the efficiency of conversion of elephantgrass biomass into bioenergy.

5. Conclusions

Genomic information allowed increases the predicted accuracy for biomass quality traits on elephantgrass.

The marker M28_161 was significantly associated with large values of biomass digestibility.

The marker M35_202 was detected in the elephantgrass genotypes that showed the small values of biomass digestibility.

Bioinformatics analysis revealed that the both markers are close to candidate genes involved in biosynthesis of cell wall molecules, which support their association with biomass digestibility.

The markers and the candidate genes identified here are useful for breeders that need to change biomass digestibility in elephantgrass.

6. References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **J Dairy Sci** 2010;93, 743-752.

Allouis S, Qi X, Lindup S, Gale M, Devos KM. Construction of a BAC library of pearl millet, Pennisetum glaucum. **Theor Appl Genet** 2001;102, 1200-1205.

Anderson WF, Casler MD, Baldwin .S. **Genetic Improvement of Bioenergy Crops** (eds. Vermerris W.) 347-376 (Springer Science + Business Media LLC , 2008).

Azevedo ALS, Costa PP, Machado JC, Machado MA, Pereira AV, Lédo FJDS. Cross-species amplification of Pennisetum glaucum microsatellite markers in Pennisetum purpureum and genetic diversity of napier grass accessions. **Crop Sci** 2012;52, 1776-1785.

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, et al. Reference genome sequence of the model plant Setaria. **Nat Biotechnol** 2012;30, 555-561.

Biazzi E, Nazzicari N, Pecetti L, Brummer EC, Palmonari A, Tava A, et al. Genome-wide association mapping and genomic selection for alfalfa (Medicago sativa) forage quality traits. **PLoS On**e 2017;12, e0169234.

Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. **Annu Rev Plant Biol.** 2003;54, 519-546.

Budak H, Pedraza F, Cregan PB, Baenzinger PS, Dweikat I. Development and utilization of SSRs to estimate the degree of genetic relationships in a collection of pearl millet germplasm. **Crop Sci** 2003;43, 2284-2290.

Casler MD, Vogel KP. Accomplishments and impact from breeding for increased forage nutritional value. **Crop Sci** 1999;39, 12-20.

Chen H. Chemical Composition and Structure of Natural Lignocellulose. In: **Biotechnology of Lignocellulose**. Springer, Dordrecht. 2014.

Chen XF, Huang C, Xiong L, Wang B, Qi GX, Lin XQ, et al. Use of elephant grass (Pennisetum purpureum) acid hydrolysate for microbial oil production by Trichosporon cutaneum. **Prep Biochem Biotechnol** 2016;46, 704-708.

Covarrubias-Pazaran G. Genome-assisted prediction of quantitative traits using the R package sommer. **PLoS One** 2016; 11, e0156744.

Dorez G, Ferry L, Sonnier R, Taguet, Lopez-Cuesta JM. Effect of cellulose, hemicellulose and lignin contents on pyrolysis and combustion of natural fibers. **J. Anal Appl Pyrol** 2014;107, 323-331.

Fartini MS, Abdul MSM, Ridzuan MJM, Amin NAM, Gibson AG. Compressive properties of Napier (Pennisetum Purpureum) filled polyester composites. **Plast Rubber Compos** 2016;45, 136-146.

Fontoura CF, Brandão LE, Gomes LL. Elephant grass biorefineries: towards a cleaner Brazilian energy matrix? **J Clean Prod** 2015;96, 85-93.

Fraser CM, Chapple C. The phenylpropanoid pathway in Arabidopsis. **Arabidopsis Book** 9, e0152. 2011.

Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M Jr et al. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. **Proc Nat Acad Sci** 2011; 108, 3803-3808.

Gilmour AR, Gogel BJ. Cullis BR, Welham SJ, Thompson R. **ASReml user guide release 4.1** (VSN International Ltd, UK www.vsni.co.uk, 2015).

Goering HK, Van Soest PJ. **Forage fiber analysis: Apparatus, reagents, procedures and some applications** (eds. Goering, H.K. & Van Soest) 20 p. (U.S. Agricultural Research Service, 1970).

Gomide C.A M, Paciullo DSC, Lédo FJS, Pereira AV, Morenz MJF, Brighenti AM. Informações sobre a cultivar de capim-elefante BRS Kurumi. **Comunicado Técnico, Embrapa Gado de Leite**, 1-4. 2015.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Res** 2012;40, D1178-D1186.

Grev AM, Wells MS, Samac DA, Martinson KL, Sheaffer CC. Forage accumulation and nutritive value of reduced lignin and reference alfalfa cultivars. **Agron J**. 2017;109, 2749-2761.

Gyawali S, Harrington M, Durkin J, Kyla Horner K, Parkin IAP, Hegedus DD, et al. Microsatellite markers used for genome-wide association mapping of partial resistance to Sclerotinia sclerotiorum in a world collection of Brassica napus. **Mol Breed** 2016;36, 72.

Hill W, Robertson A. Linkage disequilibrium in finite populations. **Theor Appl Genet** 1968;38, 226-231.

Huang P, Shyu C, Coelho CP, Cao Y, Brutnell TP. Setaria viridis as a model system to advance millet genetics and genomics. **Front Plant Sci** 2016;7, 1781.

Ituen E, James A, Akaranta O, Sun S. Eco-friendly corrosion inhibitor from Pennisetum purpureum biomass and synergistic intensifiers for mild steel. **Chin J Chem Eng** 2016;24, 1442-1447.

Lam PY, Tobimatsu Y, Takeda Y, Suzuki S, Yamamura M, Umezawa T, et al. Disrupting flavone synthase II alters lignin and improves biomass digestibility. **Plant Physiology** 2017;174, 972-985.

Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. **J Dairy Sci** 2009;92, 4656-4663.

Leng P, Ouzunova M, Landbeck M, Wenzel G, Eder J, Darnhofer B, et al. Quantitative trait loci mapping of forage stover quality traits in six mapping populations derived from European elite maize germplasm. **Plant Breed** 2018;137, 139-147.

Liu N, Yu P. Molecular clustering, interrelationships and carbohydrate conformation in hull and seeds among barley cultivars. **J. Cereal Sci** 2011;53, 379-383.

Madakadze IC., Masamvu TM, Radiotis T, Li J, Smith DL. Evaluation of pulp and paper making characteristics of elephant grass (Pennisetum purpureum Schum) and switchgrass (Panicum virgatum L.), **Afr J Environ Sci Technol** 2010;4, 465-470.

Mambe FT, Voukeng KI, Beng V, Kuete V. Antibacterial activities of methanol extracts of Alchornea cordifolia and four other Cameroonian plants against MDR phenotypes. **J Taibah Univ Med Sci** 2016;11, 121-127.

Mariac C, Luong V, Kapran I, Mamadou A, Sagnard F, Deu M, et al. Diversity of wild and cultivated pearl millet accessions (Pennisetum glaucum [L.] R. Br.) in Niger assessed by microsatellite markers. **Theor Appl Genet** 2006;114, 49-58.

Masuda Y, Misztal I, Tsuruta S, Legarra A, Aguilar I, Lourenco DAL, et al. Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. **J Dairy Sci** 2016;99, 1968-1974.

Mielenz JR. Ethanol production from biomass: technology and commercialization status. **Curr Opin Microbiol** 2001;4, 324-329.

Negawo AT, Jorge A, Hanson J, Teshome A, MuktaR MS, Azevedo ALS, et al. Molecular markers as a tool for germplasm acquisition to enhance the genetic diversity of a Napier grass (Cenchrus purpureus syn. Pennisetum purpureum) collection. **Trop. Grassl-Forrajes Trop.** 2018;6, 58-69.

Negawo AT, Teshome A, Kumar A, Hanson J, Jones CS. Opportunities for Napier grass (Pennisetum purpureum) improvement using molecular genetics. **Agronomy** 2017;7, 28.

Nguyen D, Rieu I, Mariani C, van Dam NM. How plants handle multiple stresses: hormonal interactions underlying responses to abiotic stress and insect herbivory. **Plant Mol Biol** 2016;91, 727-740.

Nicotra AB, et al. Plant phenotypic plasticity in a changing climate. **Trends Plant Sci.** 2010;15, 684-692.

Oakey H, Cullis B, Thompson R, Comadran J, Halpin C, Waugh R, et al. Genomic selection in multi-environment crop trials. G3 **Genes Genom Genet** 2016;6, 1313-1326.

Pauly M, Keegstra K. Biosynthesis of the plant cell wall matrix polysaccharide xyloglucan. **Annu Rev Plant Biol** 2016;67, 235-259 (2016).

Pereira AV, Lédo FJS, Morenz MJF, Leite JLB, Brighenti AM, Martins CE, et al. BRS Capiaçu: cultivar de capim-elefante de alto rendimento para produção de silagem. **Comunicado Técnico, Embrapa Gado de Leite**, 1-6. 2016.

Pereira AV, Auad AM, Lédo FJS, Barbosa S. **Plantas Forrageiras** (eds. Fonseca, D.M., Martuscello, J.A.) 197-219 (Editora UFV, 2010).

Pereira JF, Azevedo ALS, Pessoa Filho MACP, Romanel EAC, Pereira AV, Vigna BBZ, et al. Research priorities for next-generation breeding of tropical forages in Brazil. **Crop Breed. Appl Biotechnol.** 2018;18, 314-319.

Qin H, et al. Identification of associated SSR markers for yield component and fiber quality traits based on frame map and upland cotton collections. **PLoS One** 2015;10, e0118073.

R Development Core Team R: **A language and environment for statistical computing** (R Foundation for Statistical Computing, 2015).

Rajaram V, Nepolean T, Senthilvel S, Varshney RK, Vadez V, Srivastava RK, et al. Pearl millet [Pennisetum glaucum (L.) R. Br.] consensus linkage map constructed using four RIL mapping populations and newly developed EST-SSRs. **BMC Genomics** 2013;14, 159.

Reis GB, Mesquita AT, Torres GA, Andrade-Vieira LF, Pereira AV, Davide LC, et al. Genomic homeology between Pennisetum purpureum and Pennisetum glaucum (Poaceae). **Comp Cytogenet** 2014;8, 199-209.

Reiser L, Subramaniam S, Li D, Huala E. Using the Arabidopsis information resource (TAIR) to find information about Arabidopsis genes. **Curr Protoc Bioinformatics** 2017;60, 1.11.1-1.11.45.

Resende MDV, Silva FFE, Azevedo CF. **Estatística matemática, biométrica e computacional: Modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência** (eds. Resende MDV, Silva FFE, Azevedo CF) 627-768 (Suprema, 2014).

Ridzuan MJM, Majid MSA, Afendi M, Kanafiah SNA, Zahri JM, Gibson AG, et al. Characterisation of natural cellulosic fibre from Pennisetum purpureum stem as potential reinforcement of polymer composites. **Mater Des** 2016a;89, 839-847.

Ridzuan MJM, Abdul MSM, Afendi M, Mazlee MN, Gibson AG. Thermal behaviour and dynamic mechanical analysis of Pennisetum purpureum/glass-reinforced epoxy hybrid composites. **Compos Struct** 2016b;152, 850-859.

Rocha JRASC, Machado JC, Carneiro PCS, Carneiro JC, Resende MDV, Ledo FJS, Carneiro JES. Bioenergetic potential and genetic diversity of elephantgrass viamorpho-agronomic and biomass quality traits. **Ind Crops Prod.** 2017;95 :485-492.

Siraree A, Banerjee N, Kumar S, Khan MS, Singh PK, Kumar S, et al. Identification of marker-trait associations for morphological descriptors and yield component traits in sugarcane. **Physiol Mol Biol Plant** 2017;23, 185-196.

Takara D, Khanal SK. Green processing of tropical banagrass into biofuel and biobased products: An innovative biorefinery approach. **Bioresour Technol** 2011;102, 1587-1592.

Tilley JHA, Terry A. A two stage technique for in vitro digestion of forage crops. **J Br Grassl Soc** 1963;18, 104-111.

van Eeuwijk FA, Bustos-Korts DV, Malosetti M. What should students in plant breeding know about the statistical aspects of genotype x environment interactions? **Crop Sci** 2016;56, 2119-2140.

Viana AP, Resende MDV, Riaz S, Walke MA. Genome selection in fruit breeding: application to table grapes. **Sci agric** 2016;73, 142-149.

Wang C, Yan H, Li J, Zhou S, Liu T, Zhang X, et al. Genome survey sequencing of purple elephant grass (Pennisetum purpureum Schum 'Zise') and identification of its SSR markers. **Mol Breed** 2018;38, 94.

Wang H, Li K, Hu X, Liu Z, Wu Y, Huang C, Genome-wide association analysis of forage quality in maize mature stalk. **BMC Plant Biol** 2016;16, 227.

Wu G, Liu S, Zhao Y, Wang W, Kong Z, Tang D. Enhanced disease resistance4 associates with clathrin heavy chain2 and modulates plant immunity by regulating relocation of EDR1 in Arabidopsis. **Plant Cell** 2015;27, 857-872.

Wu Z, Zhang M, Wang L, Tu Y, Zhang J, Xie G, et al. Biomass digestibility is predominantly affected by three factors of wall polymer features distinctive in wheat accessions and rice mutants. **Biotechnol Biofuels** 2013;6, 183.

Zhou S, Wang C, Frazier TP, Yan H, Chen P, Chen Z, et al. The first Illumina-based de novo transcriptome analysis and molecular marker development in Napier grass (Pennisetum purpureum). **Mol Breed** 2018;38, 95.

Zhou X, Jacobs TB, Xue LJ, Harding SA, Tsai CJ. Exploiting SNP s for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and redundancy. **New Phytol** 2015;208, 298-301.

7. Supporting Information

Table S1. Genotypes' registrations names of the Active Elephantgrass Germplasm Bank (BAGCE) maintained by Embrapa Dairy Cattle Research Center and their respective code. †Code of accessions.

Code †	BAGCE registration	Code	BAGCE registration	Code	BAGCE registration	Code	BAGCE registration
1	Elefante da Colômbia	26	Mineiro	51	Guaco	76	12 AD IRI
2	BAGCE 2	27	Mole de Volta Grande	52	Cuba-115	77	07 AD IRI
3	Tres Rios	28	Porto Rico	53	Cuba-116	78	Pasto Panamá
4	Napier Volta Grande	29	Napier	54	Cuba-169	79	BAGCE 92
5	Mercker Santa Rita	30	Mercker Comum	55	King Grass	80	09 AD IRI
6	Pusa Napier Nº 2	31	Terezópolis	56	Roxo Botucatu	81	11 AD IRI
7	Gigante de Pinda	32	Taiwan A-26	57	Mineirão IPEACO	82	05 AD IRI
8	Napier Goiano	33	Duro de Volta Grande	58	Vruckwona Africano	83	06 AD IRI
9	Mercker S. E. A.	34	Mercker Comum Pinda	59	Cameroon	84	01 AD IRI
10	Taiwan A-148	35	Turrialba	60	BAGCE 69	85	04 AD IRI
11	Porto Rico 534-B	36	Taiwan A-146	61	Guaçu	86	13 AD IRI
12	Taiwan A-25	37	Cameroon - Piracicaba	62	Napierzinho	87	03 AD IRI
13	Albano	38	Taiwan A-121	63	IJ 7125	88	02 AD IRI
14	Híbrido Gigante da Colômbia	39	Vrukwona	64	IJ 7126	89	08 AD IRI
15	Pusa Gigante Napier	40	T241 Piracicaba	65	IJ 7127	90	Pioneiro
16	Elefante Híbrido 534-A	41	BAGCE 50	66	IJ 7136	91	Banhado
17	Costa Rica	42	BAGCE 51	67	IJ 7139	92	Roxo Farroupilha
18	Cubano de Pinda	43	Elefante Cachoeiro Itapemirim	68	IJ 7141	93	Roxo de Canguçu
19	Mercker Pinda	44	Sem Pelo	69	Goiano	94	Roxo do Itassú
20	Mercker 86 México	45	Capim Cana D'África	70	CAC 262	95	BRS Capiaçu
21	Taiwan A-144	46	Kizozi	71	Ibitinema	96	CNPGL 91-06-3
22	Napier S.E.A.	47	Gramafante	72	Australiano	97	CNPGL 96-25-3
23	Taiwan A-143	48	Roxo	73	BAGCE 82	98	BRS Canará
24	Pusa Napier Nº 1	49	Mott	74	13 AD	99	CNPGL 94-49-6
25	Elefante de Pinda	50	BAGCE 59	75	10 AD IRI	100	PCM 0701

GENERAL CONCLUSIONS

Elephant grass showed genetic variability for traits related to bioenergetic and animal feeding.

The random regression model was applied and gives genetic insights into elephantgrass clones' persistence and G x E interaction.

It was detected a gene pool that acts over the biomass yield (regardless of the harvest), as well as other gene pool, which show differences on genes expression (these genes are the major responsible for clones' persistence).

Genomic information increases the prediction accuracy for biomass quality traits on elephantgrass.

Two SSRs marker were associated with biomass digestibility.

The nucleotide sequence alignment revealed that the markers are close to candidate genes involved in biosynthesis of cell wall molecules, which support their association with biomass digestibility.

The genetic insights into persistence as well as the candidate genes found here are useful for breeders that need breed new cultivars for animal feed or bioenergetics uses.